Extending the Predictive Power and Scope of Electronic Structure Theory and Quantum Transport

Nicolas Poilvert

M.S., Quantum and Statistical Physics
École Normale Supérieure, 2006

Engineering Diploma, Mathematics and Physics
École Polytechnique, 2005

Submitted to the Department of Materials Science and Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MATERIALS SCIENCE AND
ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

Author ....
Department of Materials Science and Engineering
May 18, 2011

Certified by ....
Nicola Marzari
Associate Professor of Materials Science and Engineering
Thesis Supervisor

Accepted by ....
Christopher A. Schuh
Chair, Department Committee on Graduate Students

Extending the Predictive Power and Scope of Electronic Structure Theory and Quantum Transport

Nicolas Poilvert

Submitted to the Department of Materials Science and Engineering
on May 18, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Materials Science and Engineering

Abstract

The day 1998 Nobel Prize recipient Walter Kohn wrote his first article on Density Functional Theory, he could never have predicted its eventual impact on computational materials science. Almost 50 years after his original article, the field has seen tremendous improvement both in computer hardware and in software algorithms, and the resulting combination of an elegant theory and truly predictive power has enabled accurate, reliable simulation of relevant materials properties. But the story does not end here. Density Functional Theory still needs major improvements in at least two directions to really add the power of *ab-initio* quantum mechanics to the toolbox of materials engineers. The first direction aims at improving the accuracy of predicted materials properties, while the second aims at improving the scope of *first-principles* predictions. In this work, an attempt to push the field forward in each of the directions outlined above is set forth. A novel scheme that drastically reduces self-interaction errors in Density Functional Theory, and re-establish physical meaning in Kohn-Sham orbital energies is presented. The accuracy of the newly developed functional is shown to remedy a lot of the known deficiencies of local and semi-local functionals, while preserving their intrinsic qualities on established properties. A second key contribution from this work has been the development of a set of robust and efficient algorithms for large scale quantum transport calculations within Density Functional Theory combined with the Non-Equilibrium Green's Function formalism. Emphasis on user-friendliness was an underlying motivation throughout the implementation phase into the Wannier90 code. Systems with sizes up to two orders of magnitude larger than what DFT can currently deal with can now routinely be investigated. The automation of the whole process also opens up the possibility for high-throughput quantum conductance calculations with potential usefulness in the field of nanoelectronics.

Thesis Supervisor: Nicola Marzari
Title: Associate Professor of Materials Science and Engineering

À ma mère Évelyne, à mon père Pierre-Yves.

Acknowledgments

First of foremost, I would like to sincerely thank my advisor, Prof. Nicola Marzari for giving me the opportunity to join his research group at MIT. His very broad interest in all areas of Physics allowed me to explore different interesting ideas without any restrictions. As a consequence, I learned a great deal about autonomy in my research and the importance of establishing research directions. Science being so vast and complex, one can easily get lost and the support of a thesis committee is paramount. I wish to thank Prof. Ceder and Prof. Strano for their invaluable advices and their time during those years.

I was fortunate enough to join a group of remarkable people, composed of brilliant PhD students and outstanding post-doctoral fellows. These people patiently and kindly answered the mountains of questions that constantly came to my mind on a day-to-day basis, from the fundamentals of Density Functional Theory to the best way to program in parallel with MPI. I thank Ismaila, Young-Su, Arash, Patrick, Elise, Jivtesh, Sejoong, Nicephore, Boris, Nicholas, Francesca, Oswaldo, Xiaofeng, Dmitri, Nicola, Oliviero, Davide, Andrea, Samed, and everyone in the group that I shamefully forgot.

I wish to express my deepest gratitude to my close collaborator and dear friend Matthew Shelley from Imperial College London. My research really took off the day I started to collaborate with Matt and I wish him the very best for his future endeavors.

When the time came to take a break or simply release some frustration from research not moving forward, I could always count on my friends to help me out. My sincerest gratitude to Nicolas, Simon, Dustin, Herve, Matthieu, Benjamin, Ella, Tilke, Yann, Justin, Claudio, Takayo, Ole and Roman.

At last, how can one possibly live in a country so far away from home without the unconditional and loving support of a family ? I would never have gotten thus far

without them. A huge thank you to my mother, Évelyne and my father, Pierre-Yves. I also thank my sisters Morgane and Rozenn and the whole of the family.

Reading Guide and Contributions

The thesis has been written with a view to provide a somehow self-contained and pedagogical approach to modern day Density Functional Theory. The strong emphasis on establishing clean proofs for most of the major results was deliberate and the author hopes that it may help the reader to achieve a deeper level of understanding. The contributions of this thesis being rather theoretical and algorithmic in nature, the author feels that merging those smoothly alongside a presentation of the theory leads to a better account of the unity of the field. Hopefully, the reader will arrive at the same conclusion.

The work is organized in four parts. In Part I, an introduction to Density Functional Theory (DFT) from the point of view of functional analysis is presented. In the third chapter, the Kohn-Sham equations are derived as Euler-Lagrange equations to a Variation Principle, both for orbital-dependent functionals and more "classic" unitary invariant functionals like LDA or GGAs. Part II deals with the Bloch and Wannier representations in mean-field theories (DFT being one such mean-field theory). Although the content explicitly emphasizes the Kohn-Sham Hamiltonian, the results are generally true for *any* Hamiltonian modeling a translationally periodic system. In this part, a novel algorithm leading to a practical means to extract ordered real-space Hamiltonian matrices is set forth. Once an ordered Hamiltonian matrix is available, a large amount of physical properties can be derived, like band structures, Fermi surfaces, Fermi energies, transmission functions, etc... Part III constitutes one of the core contribution of this thesis. A new correction scheme to Density Functional Theory is found which, once implemented, leads to dramatic improvements in the prediction of a wealth of physical properties of quantum systems. This work was initiated by Dr. Ismaila Dabo and Profs. Matteo Cococcioni and Nicola Marzari, and led to a collaboration with Dr. Andrea Ferretti, Dr. Yanli Li and the author. A

comprehensive analysis of the scheme's accuracy is discussed in the second chapter of that part. Such diverse properties as spectroscopic data, equilibrium geometries, band structures of periodic solids, dissociation limits and chemical reaction barriers, as predicted by this newly developed theory, are systematically compared to either high-level accurate theoretical methods or experimental data. The last part of the manuscript, Part IV, covers the basics of quantum transport along with a thorough exposition of the author's contributions to the field. The contributions split into two parts. First, a novel set of algorithms developed explicitly to streamline and automate quantum conductance calculations was implemented into the Wannier90 code. This work is the result of a collaborative effort between Dr. Matthew Shelley[1], Dr. Arash Mostofi, Prof. Nicola Marzari and the author. Second, a new suite of computer codes, designed to be used as a post-processing to Wannier90, allows the "computational synthesis" of any quasi-one dimensional complex system from the fundamental Hamiltonian building blocks provided by Wannier90, in much the same way as one would assemble fundamental Lego® bricks to build large complex structures. This contribution is solely the work of the author and Prof. Nicola Marzari. We finally give some conclusions and prospects for future work.

Part III and IV of the thesis have been published in peer-reviewed journals in three different articles. Those are :

- "Koopmans' condition for density-functional theory", by Ismaila Dabo, Andrea Ferretti, Nicolas Poilvert, Yanli Li, Nicola Marzari and Matteo Cococcioni. *Physical Review B*, 82, 115121 (2010)

- "Switchable Conductance in Functionalized Carbon Nanotubes via Reversible Sidewall Bond Cleavage", by Elise Li, Nicolas Poilvert and Nicola Marzari. Accepted to *ACS Nano*, to appear in 2011.

- "Automated quantum conductance calculations using maximally-localized Wannier functions", by Nicolas Poilvert, Matthew Shelley, Arash Mostofi and Nicola Marzari. Accepted to *Computer Physics Communications*, to appear in 2011.

---

[1]Dr. Matthew Shelley former PhD student of Dr. Arash Mostofi at Imperial College, London

Contents

CONTENTS

Part I

Introduction to Density Functional Theory

1

Mathematical Preliminary

Density Functional Theory has its roots in the field of Functional Analysis. This mathematical domain is arguably one of the greatest achievement of 20th century mathematics, and is associated with such prestigious names as *Hilbert*, *Banach*, *Volterra* or *Hadamard* [1]. The central idea of functional analysis is to consider **functions** as **vectors** in a (functional) **vector space**, much like vectors in Euclidean spaces [2]. It is not an overstatement to say that this conceptual change has been critical in the development of much of modern day physics. Such diverse areas as Quantum Mechanics, Fluid Dynamics, and General Relativity, have been greatly impacted by those developments.

In this chapter, we will go over some of the most important notions in the fields of Functional Analysis and Variational Calculus. This, in turn, will give us the appropriate vocabulary in which to express Density Functional Theory. Since the purpose is to introduce only what is necessary, we will obviously not be comprehensive. Instead, we will try to introduce the notions in a simple and hopefully intuitive manner, using some mathematical and/or physical analogies.

1.1 Functional Spaces

1.1.1 General Definition

We are used to think of spaces like $\mathbb{R}$ (the set of real numbers) or $\mathbb{C}$ (the set of complex numbers). By *space*, we mean a collection or set of elements. In the case of $\mathbb{R}$, it is a set of points, each point corresponding to a real number. With these spaces in mind, we define in a similar way **functional spaces** as sets or collections of elements. The only difference is that those elements are **functions** in the usual sense.

Generally when one deals with functions, one likes to be able to multiply that function by a real or complex number, but also to add two functions together. In mathematics the above described processes are known respectively as multiplication by a scalar and addition. If we add those two properties in our functional space, we end up with what the mathematicians call a **vector functional space**. In the rest

---

[1]For a beautiful introduction to the field, the reader is directed to J. Dieudonné's "History of Functional Analysis"

[2]see http://en.wikipedia.org/wiki/Euclidean_space

of this thesis, whenever we will deal with a functional space, it will be understood that the functional space is actually a vector functional space. As a general rule, we will denote a vector functional space by the letter X.

1.1.2   Normed Functional Spaces

In order to be able to define a notion of *convergence* for a sequence of elements in a space, the mathematicians introduced the notion of a **norm**. A norm is simply a generalization of the well-know absolute value in the case of real numbers or modulus in the case of complex numbers. A norm is nothing else than a function that takes an element x of a space X and returns a *non-negative* real number. But on top of that, the norm has to satisfy the following properties

1. $||\alpha x|| = |\alpha|||x||$ where α is a real or complex number

2. $||x|| = 0$ implies that $x = 0$

3. $||x + y|| \leq ||x|| + ||y||$ for all x and y in X

A **Normed Functional Space** is nothing else than a vector functional space equipped with a norm.

1.1.3 L^p Spaces

There is an important special class of normed functional spaces that are constantly used in mathematics and physics. Those are called L^p spaces after the French mathematician *Henri Lebesgue*. Those spaces are better introduced by using the analogy with Euclidean spaces.

An element of an Euclidean space (a vector), can be written as a n-tuple of real numbers $x = (x_1, x_2, ..., x_n)$, for some integer n. We are used to define the *length* of such a vector by the Euclidean length

$$||x||_2 = \left(|x_1|^2 + |x_2|^2 + ... + |x_n|^2\right)^{1/2}$$

Where $||x||_2$ is a norm as defined above. But this is by no means the only way to define a *length* in $\mathbb{R}^n$. Indeed if x has a dimension of meters then the following expression also gives a number in meters

$$||x||_p = \left(|x_1|^p + |x_2|^p + ... + |x_n|^p\right)^{1/p}$$

and one can prove that $||...||_p$ is also a norm in the mathematical sense. p does not need to be an integer, and it can actually be any real number greater than 1. Now imagine that each of the coordinates x_i represents the coordinates of a function in a given normed functional space. To illustrate what happens in the case of a typical functional space, we will use the classic and intuitive example of the space of *continuous* functions over an interval $[a, b]$ of the real line. Then the "coordinates" of the function in that space are simply the values of the function at every point in $[a, b]$.

Then the idea of summing up all the absolute values of the coordinates to the power p is simply translated into taking the *integral* of the absolute value of the function to the power p over the interval $[a,b]$. Then for dimensionality reasons, we take the p-th root of that integral, and we end up with the following

$$||f||_p = \left(\int_a^b |f(x)|^p\,dx \right)^{1/p} \tag{1.1}$$

One can prove that the function $||...||_p$ defined above is indeed a norm in the mathematical sense. It is a norm on the vector functional space of continuous functions over $[a,b]$.

A vector functional space equipped with the above norm is called a L^p space. As an example, let us consider again the space $\mathscr{C}([a,b],\mathbb{R})$ (vector functional space of continuous functions over the interval $[a,b]$ into $\mathbb{R}$) equipped with $||...||_2$. Then from equation 1.1, we see that

$$||f||_2 = \left(\int_a^b |f(x)|^2\,dx \right)^{1/2}$$

which is the very well-known 2-norm extensively used in Quantum Mechanics.

1.2 Functionals

Now that we have introduced the notion of a functional space, we can introduce the notion of a **functional**. A functional, is nothing else than a mapping that takes a function as its input and delivers a real number [3]. This sounds complicated but it really isn't. Actually we are used to see and manipulate functionals all the time. An excellent example is a *norm* in an L^p space. Indeed such a norm (like $||...||_p$ defined in equation 1.1) can be thought of as a mapping that takes a continuous function $f(x)$ and delivers a (positive) real number $||f||_p$. Here are some examples of functionals over the functional space $\mathscr{C}([a,b],\mathbb{R})$

$$f \mapsto f(x_0) \text{ for some } x_0 \text{ in } [a,b] \tag{1.2}$$

$$f \mapsto \int_a^b f(x)dx$$

$$f \mapsto \int_a^b \left(\frac{df}{dx}(x) \right)^2 dx$$

$$f \mapsto \int_a^b G\left(x, f(x), \frac{df}{dx}(x) \right) dx$$

where $\frac{df}{dx}$ means the derivative of f with respect to x and $G(x,y,z)$ is for example a continuous function over $[a,b] \times \mathbb{R} \times \mathbb{R}$.

---

[3]In general, a functional can also map to a complex number or even a vector but in this thesis we will only consider real functionals

1.3 Dual Spaces

The reason to introduce the concept of a dual space is motivated by the fact that in Density Functional Theory we will see that the space of all possible admissible *external potentials* that a system can be subject to, is the dual space of the space of all *admissible densities*. Moreover, the notion of a dual space is central in many areas of mathematics so knowing the definition is quite important.

1.3.1   A definition

Let X be a normed functional space. We have seen in the previous section what a functional was. Let us now impose that the functional be *linear* and *continuous*. Then if we call F the functional and take f, g and α to be respectively two elements of X and a scalar, then F has the following property

$$F[\alpha f + g] = \alpha F[f] + F[g] \tag{1.3}$$

We then come to the definition of the **Dual Space**, which is the set of *all* linear and continuous functionals on X. We will generally write X^* for denoting the dual of X. Moreover, we will use the following symbols

$$\langle F, u \rangle = F(u) \text{ for } F \text{ in } X^* \text{ and } u \text{ in } X \tag{1.4}$$

1.3.2   Dual Space for L^p

An important result of functional analysis is to provide the dual space of L^p which is simply L^q with q such that

$$\frac{1}{p} + \frac{1}{q} = 1$$

Actually saying that the dual of L^p is L^q is technically not appropriate since L^q is a normed functional space while L^{p*} (the dual of L^p) is supposed to be a space of linear functionals over L^p. The reason for such a shortcut is that there exist an *isomorphism* between L^{p*} and L^q. An isomorphism being a one-to-one continuous mapping between two spaces, such that the inverse mapping exists and is also continuous and one-to-one. This means that for every element in L^{p*} there is one and only one element in L^q that can be brought into correspondence with it. Hence, because of that isomorphism, taking an element in L^{p*} or in L^q is really the same thing and so we say for short that L^q is the dual of L^p. Going back to equation 1.4, we see that using the result stated in this section we can write

$$\langle g, f \rangle = \int g(x) f(x) dx \text{ for } g \text{ in } L^{p*} \approx L^q \text{ and } f \text{ in } L^p \tag{1.5}$$

1.4 Convex Functions and Functionals

The notion of *convexity* of a function is quite natural and well-known. Let's describe what it means for a continuous function over the interval $[a,b]$. Convexity means that if one takes two points x_1 and x_2 in $[a,b]$ and the corresponding function values $f(x_1)$ and $f(x_2)$, then the segment linking the vectors of $\mathbb{R}^2$ $(x_1, f(x_1))$ and $(x_2, f(x_2))$ is lying "above" the graph of f. This is illustrated in figure 1-1. From this figure, we
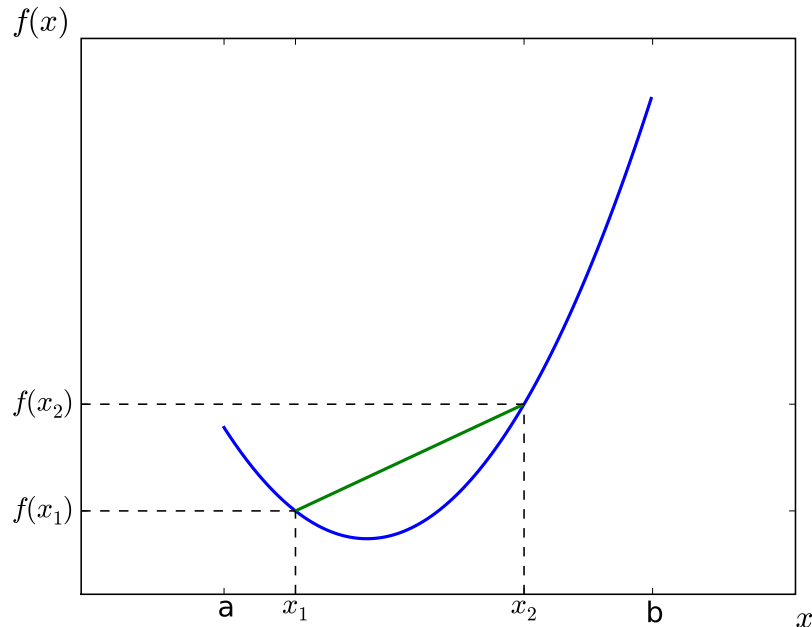
Figure 1-1: Illustration of the notion of convexity for a simple real function of a real argument. The segment linking $(x_1, f(x_1))$ and $(x_2, f(x_2))$ (in green) lies above the graph of $f(x)$ (in blue).

easily find out that the mathematical translation of *convexity* is

$$f(\alpha x_2 + (1-\alpha)x_1) \le \alpha f(x_2) + (1-\alpha)f(x_1) \text{ for all } \alpha \in [0,1] \qquad \boxed{1.6}$$

In substance, this equation says that if one picks up a point x between x_1 and x_2, then the function value at x, $f(x)$, is less or equal than the linear interpolation between $f(x_1)$ and $f(x_2)$ at x. For a twice differentiable function, one can easily check for convexity by simply computing the second derivative of f. If that second derivative is positive for all x in the function domain, then the function is convex over that domain.

In the case of functionals, we could draw a very similar picture than figure 1-1, but here x would actually be a function itself. The mathematical definition of *convexity* for functionals is still the same as in equation 1.6. In the case of a functional

F over the space $\mathscr{C}([a,b],\mathbb{R})$, convexity would read

$$F(\alpha g + (1-\alpha)f) \leq \alpha F(g) + (1-\alpha)F(f) \text{ for all } \alpha \in [0,1]$$

for any functions f and g in $\mathscr{C}([a,b],\mathbb{R})$.

1.5 Functional Derivatives

As soon as one comes up with the notion of a functional, one can start asking whether we can "differentiate" that functional. But of course here the question is : "how do you define the derivative of a functional?". In order to have a "feeling" for what that derivative might be, let us see how we would deal with a functional in practice.

On a computer, one cannot represent a function in its entirety because one cannot represent an entire interval, let say $[a,b]$, due to finite storage capabilities. So in order to represent a function $f(x)$ over $[a,b]$, we would actually discretize the interval $[a,b]$ by introducing some kind of subdivision of it and consider $f(x)$ to be the set of all function values on that subdivision. Let us call $\{x_i\}_{1 \leq i \leq N}$ that subdivision and $\{f_i = f(x_i)\}_{1 \leq i \leq N}$ the set of function values. Since f can take any value in $\mathbb{R}$ at every point x_i, we see that a discrete representation of a function in the space $\mathscr{C}([a,b],\mathbb{R})$ is akin to a vector $(f_1, f_2, ..., f_N)$ in $\mathbb{R}^N$. If we now consider a functional F over $\mathscr{C}([a,b],\mathbb{R})$, we realize that the discretized version of the functional is simply a function of many variables $\{f_i\}_{1 \leq i \leq N}$ which delivers a real number. At this point we know how to define what is called the *directional derivative* of F. Indeed if we fix a vector $\vec{f} = (f_1, f_2, ..., f_N)$ and consider another vector $\vec{g} = (g_1, g_2, ..., g_N)$, then the directional derivative of F at $\vec{f}$ in the direction of $\vec{g}$ is

$$F'_{\vec{g}}(\vec{f}) = \lim_{\epsilon \to 0^+} \frac{F(\vec{f} + \epsilon \vec{g}) - F(\vec{f})}{\epsilon} \tag{1.7}$$

Using some fundamental results from Calculus, we arrive at the following expression

$$F'_{\vec{g}}(\vec{f}) = \sum_{i=1}^{N} \frac{\partial F}{\partial f_i} g_i = \vec{\nabla} F . \vec{g} \tag{1.8}$$

Keeping those results in mind, we will now define what is called the **Gâteaux derivative** of a functional in the same way as equation 1.7. Let us choose two functions $f(x)$ and $g(x)$ and an $\epsilon > 0$ and define the Gâteaux derivative as

$$F'_g(f) = \lim_{\epsilon \to 0^+} \frac{F(f + \epsilon g) - F(f)}{\epsilon} \tag{1.9}$$

Then using the fact that when the subdivision becomes very dense we have $f_i \to f(x)$, $g_i \to g(x)$ and the sum in equation 1.8 becomes an integral over the interval $[a,b]$,

the "practical" definition of the Gâteaux derivative is

$$F'_g(f) = \int_a^b \frac{\delta F}{\delta f(x)} g(x) dx \qquad (1.10)$$

Equation 1.10 does not tell us what the derivative $\frac{\delta F}{\delta f(x)}$ is, but it *defines* it. It tells us that if we know the analytical expression for F, then we compute the ratio in equation 1.9 and let ϵ go to 0. If we can manage to recast the expression of that limit into something like 1.10, then the term $\frac{\delta F}{\delta f(x)}$ **is** by definition the **functional derivative** of F at $f(x)$.

Using the above definitions, let us compute some functional derivatives of great interest for the chapter on Density Functional Theory. We will be concerned here with the following real functionals

$$F[f] = \int_a^b \left(\frac{df}{dx}\right)^2 dx \qquad (1.11)$$

$$F[f] = \int_a^b f(x)V(x)dx$$

$$F[f] = \int_a^b \int_a^b \frac{f(x)f(y)}{|x-y|} dxdy$$

The first functional is actually a kinetic energy functional in disguise. The second functional will correspond to the energy associated with the external potential while the last functional will represent the Hartree electrostatic energy.

Focusing on the first functional in 1.11 and using the linearity of usual derivatives we have

$$F[f + \epsilon g] - F[f] = \int_a^b \left(\left(\frac{df}{dx} + \epsilon \frac{dg}{dx}\right)^2 - \left(\frac{df}{dx}\right)^2\right) dx$$

Expanding the first term and neglecting all the terms of order ϵ^2 or more into $\mathcal{O}(\epsilon^2)$ we find

$$F[f + \epsilon g] - F[f] = 2\epsilon \int_a^b \frac{df}{dx}\frac{dg}{dx} dx + \mathcal{O}(\epsilon^2)$$

Dividing by ϵ and taking the limit $\epsilon \to 0$ we end up with

$$F'_g[f] = 2 \int_a^b \frac{df}{dx}\frac{dg}{dx} dx$$

Let us now suppose that the functional space allows for second derivatives (such that $\frac{d^2f}{dx^2}$ exists). Then integrating by parts leads us to

$$F'_g[f] = 2\{[g(x)\frac{df}{dx}]_a^b - \int_a^b \frac{d^2f}{dx^2}g(x)\}$$

Finally introducing the δ functions $\delta(x-a)$ and $\delta(x-b)$, we can simplify the above

expression to

$$F'_g[f] = \int_a^b 2\left[-\frac{d^2f}{dx^2} + f'(b)\delta(x-b) - f'(a)\delta(x-a)\right]g(x)dx$$

The above integral form is exactly equivalent to equation 1.10. By identification we find

$$\frac{\delta F}{\delta f(x)} = -2\frac{d^2f}{dx^2} + 2f'(b)\delta(x-b) - 2f'(a)\delta(x-a) \tag{1.12}$$

In the special case where the functions f and g are periodic over $[a,b]$ (as will be the case in general in Density Functional Theory), the last two terms in equation 1.12 can be dropped and the functional derivative simplifies to

$$\frac{\delta F}{\delta f(x)} = -2\frac{d^2f}{dx^2}$$

We now focus on the second functional in 1.11. Computing $F[f+\epsilon g]-F[f]$ trivially gives us

$$F[f+\epsilon g] - F[f] = \epsilon \int_a^b V(x)g(x)dx$$

Dividing by ϵ and taking the limit $\epsilon \to 0$, we find the functional derivative by identification with formula 1.10

$$\frac{\delta F}{\delta f(x)} = V(x) \tag{1.13}$$

For the last functional in 1.11, let us expand the term $f(x)f(y)$ when f is replaced with $f+\epsilon g$. Up to first order in ϵ we have

$$(f(x)+\epsilon g(x))(f(y)+\epsilon g(y)) = f(x)f(y)+\epsilon(f(x)g(y)+f(y)g(x)) + \mathcal{O}(\epsilon^2)$$

Plugging the above expression back into $F[f]$ we end up with

$$F[f+\epsilon g] - F[f] = \epsilon \int_a^b \int_a^b \frac{f(x)g(y)+f(y)g(x)}{|x-y|}dxdy + \mathcal{O}(\epsilon^2)$$

Dividing by ϵ and taking the limit $\epsilon \to 0$ we end up with

$$F'_g[f] = \int_a^b \int_a^b \frac{f(x)g(y)+f(y)g(x)}{|x-y|}dxdy$$

Because of the symmetric nature of the integrand, the above expression can be simplified into

$$F'_g[f] = \int_a^b \left(2\int_a^b \frac{f(y)}{|x-y|}dy\right)g(x)dx$$

which by identification with equation 1.10 finally gives us

$$\frac{\delta F}{\delta f(x)} = 2\int_a^b \frac{f(y)}{|x-y|}dy \tag{1.14}$$

1.6 Legendre Transform of Convex Functions

The notion of Legendre transformation is very central in thermodynamics. It is underlying the whole of thermodynamic potential theory. But perhaps less known is the fact that Legendre transforms are also of paramount importance in Density Functional Theory. Once again we will first introduce Legendre transforms in the context of a simple function and then generalize to a functional.

Let us consider a *convex* function $f(x)$ and a number $p \in \mathbb{R}$. Then the function $-f(x)$ is a *concave* function. What's more the function $x \mapsto px$ is also concave (actually it is a linear function which has the special property to be both convex and concave). Since the sum of two concave functions is a concave function, then $x \mapsto px - f(x)$ is *concave*. A classic theorem about concave functions tells us that since the function $x \mapsto px - f(x)$ is bounded over $[a, b]$ (since it is continuous), then a **maximum** exists and is **unique**[4]. This allows us to define what we will call the **Legendre transform** of $x \mapsto px - f(x)$ at p

$$f^*(p) = \max_{x \in [a,b]} \{px - f(x)\}$$ (1.15)

We now want to visualize geometrically what the Legendre transform is. For this we draw a typical convex function between $[a, b]$ along with two representative elements from a family of straight lines with slope p on figure 1-2. Let us consider for simplicity that $f(x)$ is differentiable. Then a necessary condition for having a maximum of $x \mapsto px - f(x)$ is that the derivative of that function at the maximum (let us call it x_0) be zero

$$\frac{d}{dx}(px - f(x))|_{x=x_0} = 0 \text{ which gives } p = f'(x_0)$$

Since we know that this maximum exists and that it is unique, we realize that the maximum is obtained when the line $x \mapsto px + \beta$ is actually the tangent to the curve at $x = x_0$. But we also know the equation of that tangent. It is given by

$$x \mapsto f(x_0) + f'(x_0)(x - x_0)$$

The ordinate at the origin of that tangent is then $f(x_0) - f'(x_0)x_0$. But since $f'(x_0) = p$, that ordinate can be re-written as $f(x_0) - px_0$. Going back to the definition of the Legendre transform, we see that $f(x_0) - px_0 = -f^*(p)$. So, the Legendre transform is nothing else than the opposite of the ordinate at the origin of the tangent to the curve f with slope p. On figure 1-2 the ordinate at the origin of the tangent is at $(0,0)$. We see that for every other line parallel to the tangent crossing the curve of f, the ordinate at the origin is of higher value.

Now that we have a feel for what a Legendre transform actually is, let us simply extend the definition to a *convex functional*. Since the input space of a functional, X, is generally an unbounded infinite dimensional space, the "maximum" may actually not exist. Nevertheless the **supremum** does. We then define the **Legendre**

---

[4]strictly speaking the uniqueness is valid only for strictly concave functions
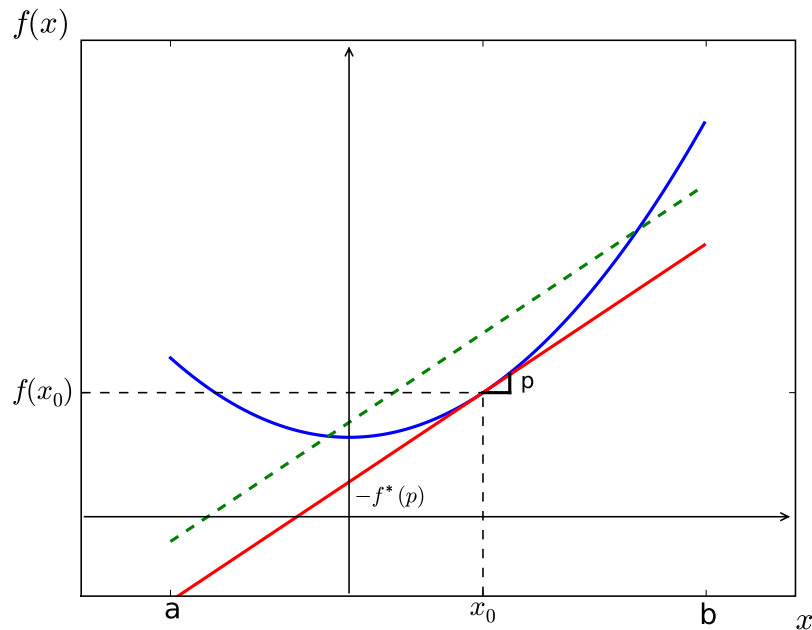
Figure 1-2: A convex function f (in blue) is shown along with 2 elements of a family of lines with given slope p. The tangent line to the curve (red line) at x_0 is the only line in the family that leads to the lowest ordinate at the origin (i.e. at $x = 0$). The second line of slope p (dashed green line) clearly displays an ordinate at the origin higher than the tangent line.

transform of a **functional** $F[f]$ as

$$F^*[g] = \sup_{f \in X}\{\langle g, f\rangle - F[f]\}$$

<div align="right">(1.16)</div>

What is remarkable about a Legendre transform is that it is a **convex** function (or functional) whatever the input function (or functional) $F[f]$ is. To see this let us take two element g_1 and g_2 in Y (actually the dual space to X, X^*) and $\alpha \in [0, 1]$. Then the following expression

$$\langle \alpha g_1 + (1 - \alpha)g_2, f\rangle - F[f]$$

can be re-written as (thanks to the linearity of $\langle,\rangle$ and the fact that $1 = \alpha + (1 - \alpha)$)

$$\alpha(\langle g_1, f\rangle - F[f]) + (1 - \alpha)(\langle g_2, f\rangle - F[f])$$

The last expression is obviously bounded from above by

$$\alpha F^*[g_1] + (1 - \alpha)F^*[g_2]$$

This comes from the very definition of the Legendre transform. We then deduce that

$\langle \alpha g_1 + (1-\alpha)g_2, f \rangle - F[f]$ is bounded from above by $\alpha F^*[g_2] + (1-\alpha)F^*[g_2]$ for all f in X. Since the supremum is the smallest of those bounds, we conclude that

$$F^*[\alpha g_1 + (1-\alpha)g_2] \le \alpha F^*[g_2] + (1-\alpha)F^*[g_2] \tag{1.17}$$

which proves the convexity of F^*.

We finish this review of Legendre transformation by stating an important theorem concerning the double Legendre transform of a *lower semi-continuous* [5] and *convex* functional.

Theorem 1.6.1. (Double Legendre Transform) *The Legendre transform of the Legendre transform (called the bi-conjugate) of a lower semi-continuous and convex functional F, is the functional itself, i.e. $F^{**} \equiv F$*

$$F^{**} \equiv F \Leftrightarrow F \text{ is convex lower semi-continuous} \tag{1.18}$$

1.7 The Technique of Lagrange Multipliers

The technique of Lagrange multipliers is a very elegant method for dealing with constraint optimization problems. Instead of minimizing the objective (or cost) function directly by ensuring that one stays inside the minimization space (which can be a complicated space because of the constraints), the Lagrange multiplier technique introduces an alternative objective function that can be minimized over a larger space without any constraints. In order to introduce Lagrange's technique, we will first get a feel for what the multipliers mean geometrically in the case of an objective function of 2 variables. Then we will generalize to a function of N variables. Eventually we will present the technique when the function is a functional.

1.7.1 A simple example

Let us consider the vector space $\mathbb{R}^2$ and a general objective function f of two variables (x, y). We wish to minimize the function f given that (x, y) are constraint to satisfy $g(x, y) = 0$. Figure 1-3 illustrate that idea. The black curve represents the constraint $g(x, y) = 0$ (which defines a 1 dimensional manifold in $\mathbb{R}^2$), while the objective function f is represented by a family of contour plots. We clearly see on that graph that f possesses a global minimum at (x_0, y_0). However, since the space of accessible points due to the constraint is reduced to the black curve, we see that we can never find the global minimum when minimizing f onto the constraint space. Now that we have a geometrical idea of the problem, let us try to understand how to solve that problem analytically. Let us suppose that we found the minimum of f (here at (x_1, y_1)) given that we can only take (x, y) onto the constraint space. Then if we move away from

---

[5] It is not so important in our discussion to know exactly what a semi-continuous functional is, but for the curious reader please refer to Zeidler [104]
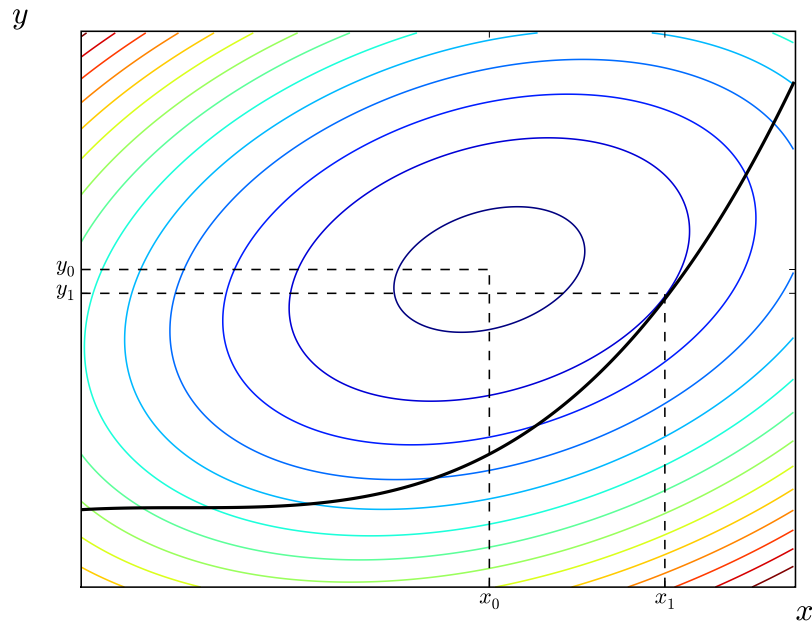
Figure 1-3: Illustration of a constraint optimization problem. The function to optimize, f, is displayed by a family of ellipses that represent some of its contours. On the other hand the minimization has to be carried out onto the black curve which specifies the constraint. Even though the global minimum of f is at (x_0, y_0) the constraint minimization gives a minimum at (x_1, y_1).

this minimum but still stay in the constraint space, we can only come up with higher values of f. This just means that the directional derivative of f in any direction in the constraint space has to be zero. We then conclude that the only direction in which the directional derivative of f can *possibly* not be zero, would be in an orthogonal direction to the constraint space. But we know that such a direction is given by the gradient of the function g that specifies the constraint space. We then deduce that there must be some λ such that

$$\vec{\nabla} f = \lambda \vec{\nabla} g \tag{1.19}$$

Looking back at figure 1-3, the above description can be interpreted by saying that the black curve (which represents the constraint space) has to be *tangent* to a contour plot of f (because a contour plot, by definition, is a locus of points for which f is constant, and so the derivative of f is 0 along the contour space). That λ is called a Lagrange multiplier. In order to find the constraint minimum, we could have consider the following function

$$h(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

and minimize that function for all (x, y, λ). Indeed finding the extremum points for that new function is easier because we can just find a set of necessary conditions by

equating the partial derivatives to 0

$$\frac{\partial h}{\partial x} = 0$$
$$\frac{\partial h}{\partial y} = 0$$
$$\frac{\partial h}{\partial \lambda} = 0$$

The first two equations lead back to equation 1.19 while the last gives back the constraint $g(x, y) = 0$. We then see that solving the original problem or the alternative one is equivalent. The alternative problem is called the **Lagrange multiplier** technique.

1.7.2 General Case in $\mathbb{R}^n$

Let us now widen the range of applicability of the Lagrange multipliers technique by considering an objective function f in $\mathbb{R}^n$. We take a set of m constraints ($m \leq n$) $g_i(x) = 0$ for $i \in 1, 2, ..., n$. We wish to minimize the following

$$\inf_{\substack{x \in \mathbb{R}^n \\ \forall i \in \{1, m\} g_i(x) = 0}} f(x) \tag{1.20}$$

As in the previous section, let us imagine that the infimum exists and is actually a minimum. Then there is a x_1 in $\mathbb{R}^n$ such that f can only increase when we move away from x_1. Just as in the 2 dimensional case, we find that the directional derivative of f has to be zero for all possible vectors in the constraint space at x_1. Equivalently, the only space in which the directional derivative of f could potentially not be 0, would be the complementary space of the constraint space. But this latter space is spanned by the vectors $\vec{\nabla} g_i$ which are vectors orthogonal to the constraint space. All in all, at the minimum, we should have a set of scalars λ_i such that the gradient of f can be written as

$$\vec{\nabla} f = \sum_{i=1}^{m} \lambda_i \vec{\nabla} g_i \tag{1.21}$$

Now if we define a new function h such that

$$h(x, \lambda_1, ..., \lambda_m) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x)$$

then we can easily show, just like in the last section, that the set of necessary conditions generated when one minimizes h without any constraints, lead to the previous set of equations. We are then solving the same problem as in 1.20.

1.7.3 The Case of Functionals

Finally, we "translate" the technique to the case of functionals. The space over which we minimize is now a vector functional space X. An element of that space is denoted f and a functional F. A constraint in this case is of the following form

$$G_i[f] = 0$$

where G_i is itself a functional. The technique of Lagrange multipliers then reads

$$\inf_{\substack{f \in X \\ \forall i \in \{1,m\} G_i[f]=0}} F[f]$$

is equivalent to

$$\inf_{\substack{f \in X \\ \forall i \in \{1,m\} \lambda_i \in \mathbb{R}}} \left[F[f] - \sum_{i=1}^{m} \lambda_i G_i[f] \right] \qquad (1.22)$$

The technique obviously holds when the λ's are complex and when we are dealing with any *extremalization* problem (i.e. not reduced to minimization problems).

2

Fundamental Density Functional Theory

In this chapter we will be concerned with establishing Density Functional Theory on clean mathematical grounds. The purpose being to explicit the existence of a Density Functional and of a Variational Principle that are exempt of mathematical uncertainties. This goal being quite demanding in terms of mathematical rigor, we simply wish to outline the necessary background and logic that underlies the construction of both the Density Functional and the Variational Principle.

For a thorough review of the foundations of Density Functional Theory, we direct to the excellent monograph by Eschrig [29]. It is interesting to note that until the 1980s, Density Functional Theory had not much mathematical rigor. The so-called Hohenberg-Kohn theorems were restricted to very limited sets of possible ground states and it took the genius of Elliott Lieb [62] to really establish Density Functional Theory as a mathematically sound theory.

Walter Kohn (1923-)

2.1 Description of a Quantum System

In this section, we wish to explore the fundamental assumptions that we make when we describe a system in Quantum Mechanics. We will see why we substitute the infinite space $\mathbb{R}^3$ for a bounded space $\mathbb{T}^3$ with periodic boundary conditions, and also how we describe a system especially in the grand canonical ensemble. Those assumptions are motivated both theoretically and practically as we will discover.

2.1.1 Replacing $\mathbb{R}^3$ with $\mathbb{T}^3$

Let us consider an isolated molecule X, with a given atomic number (the sum of the nuclei charges) Z. We know from experience that if one puts too many electrons in the system, then beyond a certain threshold, the extra electrons are not bound anymore to the molecule and escape to infinity. This means that for a given total number of electrons, say M, a quantum system may not have a ground state in the sense

that not all the particles are localized in space around the same region. In practice we would not be in a position to converge the calculation in such a situation simply because we cannot emulate an infinite space on a computer. Instead, in order to make sure that we have a ground state in every situation, we replace the whole of $\mathbb{R}^3$ space by a finite subspace $\mathbb{T}^3$ with periodic boundary conditions. The advantages of such a substitution go as follows. If the system is isolated and supports a ground state, then we will simply take a large enough but finite space. Since the charge density goes to zero exponentially when we move away from the system [48] the influence of the finiteness of $\mathbb{T}^3$ and of the periodic boundary conditions become negligible. If, however, the system does not bind some of the electrons we will see it in the calculation by observing a spreading of the charge over the whole space $\mathbb{T}^3$. In both cases, the calculation still converges because we are dealing with a *finite* space, but we can clearly distinguish the physical situations. At last when one deals with a system with some spatial periodicity, then the finiteness of space is actually *required*. In this case, periodicity introduces a reciprocal space characterized by a set of vectors $\vec{k}$, and what we need to do is make sure that we introduce enough $\vec{k}$ vectors in the calculation to converge the solution. As a consequence, we will always consider the finite space $\mathbb{T}^3$ in both theory and practice. For a further description of why we use $\mathbb{T}^3$ instead of $\mathbb{R}^3$, see Eschrig [29].

2.1.2 Wavefunction-based description

Basic quantum mechanics teaches us that an isolated system like an atom, a molecule, or even a nanoparticle, can be described by a *normalized* **ket** [12] $|\psi\rangle$ in a Hilbert space, or equivalently, by a normalized many-body **wavefunction** $\psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_M)$ where M is the total number of electrons in the system ($\vec{x}$ is a short-hand for both the position in space $\vec{r}$ and the spin σ). An additional property in the case of electrons (and fermions in general) is *antisymmetry*. If one exchanges 2 electrons, then the ket $|\psi\rangle$ (or the wavefunction) picks up a factor -1 in front. Given an isolated system, an Hamiltonian $\hat{H}$ can be defined. Calling $V(\vec{x})$ the external (electrostatic) potential exerted by the nuclei onto the electrons, and $w(|\vec{x}_i - \vec{x}_j|)$ the *electron-electron* interaction, then the real-space representation of that Hamiltonian is given by

$$\hat{H}[V, M] = -\frac{1}{2} \sum_{i=1}^{M} \vec{\nabla}_i^2 + \sum_{i=1}^{M} V(\vec{x}_i) + \frac{1}{2} \sum_{i \neq j}^{M} w(|\vec{x}_i - \vec{x}_j|) \qquad (2.1)$$

Given an observable $\hat{A}$ representing some physical quantity of interest like the kinetic energy, the particle number, or the Coulomb repulsion between the electrons, one can compute the expectation value of that observable in the quantum state ψ using

$$\langle \hat{A} \rangle = \langle \psi | \hat{A} | \psi \rangle = \int_{\mathbb{T}^{3M}} \psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_M) \hat{A} \psi^*(\vec{x}_1, \vec{x}_2, ..., \vec{x}_M) d\vec{x}_1 d\vec{x}_2 ... d\vec{x}_M \qquad (2.2)$$

In particular the expectation value of the external potential is

$$\langle \hat{V} \rangle = \int_{\mathbb{T}^{3M}} \psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_M) \sum_{i=1}^{M} V(\vec{x}_i) \psi^*(\vec{x}_1, \vec{x}_2, ..., \vec{x}_M) d\vec{x}_1 d\vec{x}_2 ... d\vec{x}_M$$

which, when using the antisymmetry of the many-body wavefunction, the indistinguishability of the electrons and the definition of the single electron density $n(\vec{x})$ simplifies to

$$\langle \hat{V} \rangle = \int_{\mathbb{T}^3} V(\vec{x}) n(\vec{x}) d\vec{x} \tag{2.3}$$

For the sake of completeness we also give the expression for the single electron density

$$
\begin{aligned}
n(\vec{x}) &= \int_{\mathbb{T}^{3M}} \psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_M) \sum_{i=1}^{M} \delta(\vec{x} - \vec{x}_i) \psi^*(\vec{x}_1, \vec{x}_2, ..., \vec{x}_M) d\vec{x}_1 d\vec{x}_2 ... d\vec{x}_M \\
&= M \int_{\mathbb{T}^{3(M-1)}} \psi(\vec{x}, \vec{x}_2, ..., \vec{x}_M) \psi^*(\vec{x}, \vec{x}_2, ..., \vec{x}_M) d\vec{x}_2 ... d\vec{x}_M
\end{aligned}
\tag{2.4}
$$

2.1.3 Density operator-based description

As convenient and powerful as the wavefunction description may be, it is, nevertheless, not the most general way to describe a system in Quantum Mechanics. Indeed, if one is interested in knowing what is the average number of electrons in a system in thermodynamic equilibrium with an electron reservoir, then a wavefunction description fails. The reason is that a wavefunction describes a system with a fixed number of electron and so this number cannot fluctuate. In order to extend the class of possible quantum states for a system, we introduce the **density operator** [72] $\hat{\Gamma}$. For a quantum system described by a ket $|\psi\rangle$ (with corresponding wavefunction $\psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_M)$), the density operator is simply

$$\hat{\Gamma} = |\psi\rangle\langle\psi| \tag{2.5}$$

If a quantum system is described by a single ket (i.e. the system has a well-defined many-body wavefunction), then one says that the system is in a **pure state**. If, however, the system of interest is a subsystem of a larger isolated system, or a macroscopic system in thermodynamic equilibrium with a particle and/or energy reservoir, then the number of electron can fluctuate and in general the system is in a **mixed state**. By mixed state, we mean that there is a probability distribution p_i over accessible pure states with well defined normalized kets $|\psi_i\rangle$, for which the density operator is written as

$$\hat{\Gamma} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \text{ with } \textbf{finitely} \text{ many } p_i \tag{2.6}$$

It is important to note that the kets $|\psi_i\rangle$ can correspond to different total number of electrons in the system. Since all the wavefunctions appearing in 2.6 are normalized,

and that we want the p_i to represent a probability distribution over those normalized accessible kets $|\psi_i\rangle$, we impose the following conditions onto the p_i's

$$p_i \geq 0 \text{ and } \sum_i p_i = 1 \tag{2.7}$$

In a quantum description based on density operators rather than kets, the expectation value for operator $\hat{A}$ is given by

$$\langle \hat{A} \rangle = \sum_i p_i \langle \psi_i | \hat{A} | \psi_i \rangle \tag{2.8}$$

In particular the expectation value of the external potential is given by

$$\langle \hat{V} \rangle = \int_{\mathbb{T}^3} V(\vec{x}) n(\vec{x}) d\vec{x}$$

just like equation 2.3 with the electron density given by

$$n(\vec{x}) = \sum_i p_i n_i(\vec{x}) \tag{2.9}$$

where n_i is the electron density associated with the normalized ket $|\psi_i\rangle$. In the case of density operators, the total number of electron is expressed as

$$N = \int_{\mathbb{T}^3} n(\vec{x}) d\vec{x} = \sum_i p_i N_i$$

where N_i is the number of electrons in a system described by the ket $|\psi_i\rangle$. Even though all the N_i's are **integer** numbers, N has no reason to be. It is in general a **non-negative real** number. This fundamental difference between a wavefunction-based description, where N is an integer, and a density operator-based description, where N is any non-negative real number, is of paramount importance for a proper description of a grand canonical state.

2.2 Functional Spaces for Densities and External Potentials

Density Functional Theory uses the electron density as its basic variable. Now that we know how to compute that density from the knowledge of the many-body wavefunction or the density operator, let us see what are the mathematical constraints on such an **admissible** density. We first consider a quantum system described by a many-body wavefunction ψ. We go back to equation 2.4 and write

$$n(\vec{x}) = M \int_{\mathbb{T}^{3(M-1)}} \psi(\vec{x}, \vec{x}_2, ..., \vec{x}_M) \psi^*(\vec{x}, \vec{x}_2, ..., \vec{x}_M) d\vec{x}_2 ... d\vec{x}_M$$

Every constraint that we can find on $n(\vec{x})$ will then necessarily come from constraints on ψ. What are those constraints ? Well we can *a priori* state two things

- ψ is **normalized** : $\int_{\mathbb{T}^{3M}} \psi(\vec{x}_1,...,\vec{x}_M)\psi^*(\vec{x}_1,...,\vec{x}_M)d\vec{x}_1...d\vec{x}_M = 1$

- ψ has **finite** kinetic energy : $\langle \hat{T} \rangle < +\infty$

The first constraint clearly leads us to

$$\int_{\mathbb{T}^3} n(\vec{x})d\vec{x} = M \qquad (2.10)$$

This simply tells us that $n(\vec{x}) \in L^1(\mathbb{T}^3)$. In order to exploit the other constraint, we first realize that equation 2.4 tells that for a fixed vector $\vec{x}$, the expression for the density is nothing but a scalar product of ψ with itself. Let us denote that scalar product by $\langle \psi | \psi \rangle'$ to emphasize that $\vec{x}$ is kept fixed. Then the density can be written as $n(\vec{x}) = M\langle \psi | \psi \rangle'$. taking the gradient of $n(\vec{x})$, the rule of differentiation of a product gives

$$\vec{\nabla} n(\vec{x}) = M\left(\langle \vec{\nabla}\psi | \psi \rangle' + \langle \psi | \vec{\nabla}\psi \rangle' \right) = 2M \operatorname{Re}(\langle \vec{\nabla}\psi | \psi \rangle')$$

From which we deduce the following inequality for the square (the square modulus of the real part of a complex number being less or equal than the square modulus of the complex number)

$$\left[\vec{\nabla} n(\vec{x}) \right]^2 \le 4M^2 |\langle \vec{\nabla}\psi | \psi \rangle'|^2$$

Using a theorem from mathematics, known as the Schwartz inequality, we find $|\langle \vec{\nabla}\psi | \psi \rangle'|^2 \le \langle \vec{\nabla}\psi | \vec{\nabla}\psi \rangle' \langle \psi | \psi \rangle'$ and so we arrive at

$$\left[\vec{\nabla} n(\vec{x}) \right]^2 \le 4M^2 \langle \vec{\nabla}\psi | \vec{\nabla}\psi \rangle' \langle \psi | \psi \rangle'$$

Inserting the definition for $n(\vec{x})$, we simplify the result to

$$\left[\vec{\nabla} n(\vec{x}) \right]^2 \le 4Mn(\vec{x})\langle \vec{\nabla}\psi | \vec{\nabla}\psi \rangle'$$

We are almost finished now. We just need to consider the following integral

$$\int_{\mathbb{T}^3} \left[\vec{\nabla}\sqrt{n(\vec{x})} \right]^2 d\vec{x} = \int_{\mathbb{T}^3} \frac{1}{4} \left[\vec{\nabla} n(\vec{x}) \right]^2 \frac{1}{n(\vec{x})} d\vec{x}$$

and use the previous inequality to arrive at

$$\int_{\mathbb{T}^3} \left[\vec{\nabla}\sqrt{n(\vec{x})} \right]^2 d\vec{x} \le \int_{\mathbb{T}^3} \langle \vec{\nabla}\psi | \vec{\nabla}\psi \rangle' d\vec{x}$$

This is where the constraint of finite kinetic energy allows us to conclude. The integral on the right-hand side in the previous inequality is the expectation value of the kinetic energy $\langle \hat{T} \rangle$. All in all

$$\int_{\mathbb{T}^3} \left[\vec{\nabla}\sqrt{n(\vec{x})} \right]^2 d\vec{x} \le \langle \hat{T} \rangle \qquad (2.11)$$

So if the kinetic energy is finite when the system is in state ψ, then the integral on the left-hand side of equation 2.11 is finite as well. Another general inequality due to Sobolev [62] (see Gagliardo-Nirenberg-Sobolev inequality) is

$$\left(\int_{\mathbb{T}^3} |u(\vec{x})|^6 d\vec{x}\right)^{\frac{1}{3}} \le C \int_{\mathbb{T}^3} |\vec{\nabla} u(\vec{x})|^2 d\vec{x}$$

with $C = \frac{1}{3}\left(\frac{2}{\pi}\right)^{\frac{4}{3}}$. Plugging $u(\vec{x}) = \sqrt{n(\vec{x})}$ into the above inequality we find

$$\left(\int_{\mathbb{T}^3} |n(\vec{x})|^3 d\vec{x}\right)^{\frac{1}{3}} \le C \int_{\mathbb{T}^3} |\vec{\nabla}\sqrt{n(\vec{x})}|^2 d\vec{x} \le C\langle\hat{T}\rangle$$

This last inequality tells us that $n(\vec{x}) \in L^3(\mathbb{T}^3)$. All in all, a proper set of *admissible* densities is

$$\mathscr{R}_N = \{n \,|\, n(\vec{x}) \ge 0, \int_{\mathbb{T}^3} n(\vec{x}) d\vec{x} = N, n \in L^3(\mathbb{T}^3)\} \tag{2.12}$$

Having gone this far a question arises immediately : what happens when we move from a wavefunction-based description to a density operator-based description ? The answer is nothing! The fundamental reason is that an admissible density coming from a density operator is just a convex combination of densities coming from wavefunctions (see equation 2.9). Then if for each wavefunction ψ_i, we have $n_i \in \mathscr{R}_{N_i}$, then one can show that $n \in \mathscr{R}_N$ with $N = \sum_i p_i N_i$. A last but important property of $\mathscr{R}_N$ is its **convexity**. Indeed if two admissible densities n_1 and n_2 are considered, then any convex combination $\alpha n_1 + (1-\alpha)n_2$ ($\alpha \in [0,1]$) is also an admissible density.

To define the space of admissible external potentials, let us go back to the only term in the expectation value of the Hamiltonian that involves that external potential, namely

$$\langle\hat{V}\rangle = \int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x}$$

Since we want that integral to be *finite* for all admissible densities n, we are led to consider the **dual** space of admissible densities as our appropriate space for admissible external potentials. But what is the dual space of $L^1(\mathbb{T}^3) \cap L^3(\mathbb{T}^3)$? A result of functional analysis gives us the answer

$$\mathcal{V} = \{V = V_1 + V_2 \,|\, V_1 \in L^{3/2}(\mathbb{T}^3) \text{ and } V_2 \in L^\infty(\mathbb{T}^3)\} \tag{2.13}$$

Admissible external potentials have to be written as a sum of two potentials, the first of which satisfies

$$\left(\int_{\mathbb{T}^3} |V_1(\vec{x})|^{3/2} d\vec{x}\right)^{2/3} < +\infty$$

and the second is bounded over $\mathbb{T}^3$. We mention at this point that similarly to the case of $\mathscr{R}_N$, the space of admissible potentials $\mathcal{V}$ is **convex**. Despite its unappealing nature, let us show that the Coulomb potential belongs to that space. To see this, just

write

$$\frac{1}{|\vec{x}|} = \frac{\theta(|\vec{x}|)}{|\vec{x}|} + \frac{1 - \theta(|\vec{x}|)}{|\vec{x}|}$$

with $\theta(|\vec{x}|) = 1$ for $|\vec{x}| \leq 1$ and $\theta(|\vec{x}|) = 0$ for $|\vec{x}| > 1$. It is then easy to verify, by switching to spherical coordinates, that the first term on the right-hand side is in $L^{3/2}(\mathbb{T}^3)$ and the second is a bounded potential.

To conclude, we indicate that thanks to our choice for the functional spaces of admissible densities and external potentials, for any ket $|\psi\rangle$ or any density operator $\hat{\Gamma}$, one can show that the expectation values for the kinetic energy, external potential energy and electron-electron repulsion energy are all *finite*. As a consequence, $\langle \hat{H} \rangle$ is always well-defined and finite.

2.3 The Density Functional and the Variational Principle

Having all the required ingredients, we can at last move to the heart of Density Functional Theory and establish two important theorems. The first will deal with the existence of a *universal* functional of the electron density. The second will introduce a variational principle allowing us to compute the ground state total energy of a quantum system from the knowledge of the universal functional.

2.3.1 The Universal Functional

The starting point of the whole business of Density Functional Theory is the Rayleigh-Ritz variation principle of Quantum Mechanics extended to the case of density operators. The variational principle reads

$$\begin{aligned} E_0[V,N] &= \inf_{\hat{\Gamma}} \left[\text{tr}(\hat{H}\hat{\Gamma}) = \langle \hat{H} \rangle | \text{tr}(\hat{N}\hat{\Gamma}) = \langle \hat{N} \rangle = N \right] \\ &= \inf_{\{p_i\}} \left[\sum_i p_i \langle \psi_i | \hat{H} | \psi_i \rangle | \sum_i p_i N_i = N \right] \end{aligned} \qquad \boxed{2.14}$$

This principle[1] tells in substance that in order to find the ground state energy of a quantum system with an average number of electrons equal to N, one needs to minimize the expectation value of the Hamiltonian over all possible density operators, given the constraint that the expectation value of the number of electrons is N. As one can see, we made explicit the dependence of the ground state energy with respect to the external potential V and the total number of electrons N. As a consequence, the ground state energy is a function of N and a functional of V. We will now prove two very important properties of the ground state energy.

---

[1]In theory the infimum may not exist for a general electron-electron interaction. Nevertheless, for **repulsive** interactions (like the Coulomb interaction), one can prove that the ground state energy is always bounded from below and so the infimum always exists.

- $E_0[V,N]$ is a **convex** function of N for a fixed potential V

- $E_0[V,N]$ is a **concave** functional of V for a fixed electron number N

Let us start with the convexity. We fix an external potential and consider two electron numbers N_1 and N_2, with a scalar $\alpha \in [0,1]$. $E_0[V,N_1]$ and $E_0[V,N_2]$ being energy infima over all density operators for respectively N_1 and N_2, there exist two sequences of density operators $\hat{\Gamma}_1^j$ and $\hat{\Gamma}_2^j$ such that $\text{tr}(\hat{N}\hat{\Gamma}_1^j) = N_1$ and $\text{tr}(\hat{N}\hat{\Gamma}_2^j) = N_2$ with $\lim_{j\to+\infty}\text{tr}(\hat{H}\hat{\Gamma}_1^j) = E_0[V,N_1]$ and $\lim_{j\to+\infty}\text{tr}(\hat{H}\hat{\Gamma}_2^j) = E_0[V,N_2]$. Let us consider the following sequence of density operator $\hat{\Gamma}^j = \alpha\hat{\Gamma}_1^j + (1-\alpha)\hat{\Gamma}_2^j$. By linearity of the trace, we have $\text{tr}(\hat{N}\hat{\Gamma}^j) = \alpha\text{tr}(\hat{N}\hat{\Gamma}_1^j) + (1-\alpha)\text{tr}(\hat{N}\hat{\Gamma}_2^j) = \alpha N_1 + (1-\alpha)N_2$. In the same spirit, we also have $\text{tr}(\hat{H}\hat{\Gamma}^j) = \alpha\text{tr}(\hat{H}\hat{\Gamma}_1^j) + (1-\alpha)\text{tr}(\hat{H}\hat{\Gamma}_2^j)$. Hence, that sequence is composed of admissible density operators for a total number of electrons equal to $\alpha N_1 + (1-\alpha)N_2$. By definition of the infimum we have for all j that $E_0[V,\alpha N_1 + (1-\alpha)N_2] \le \alpha\text{tr}(\hat{H}\hat{\Gamma}_1^j) + (1-\alpha)\text{tr}(\hat{H}\hat{\Gamma}_2^j)$. Letting j go to infinity we arrive at the desired convex inequality

$$E_0[V,\alpha N_1 + (1-\alpha)N_2] \le \alpha E_0[V,N_1] + (1-\alpha)E_0[V,N_2] \qquad (2.15)$$

Now we move on to prove the concavity. Let us fix an electron number N. We introduce two potentials V_1 and V_2 along with a scalar $\alpha \in [0,1]$. Given the linear dependence of the Hamiltonian with the external potential, it is easy to find that $\langle\hat{H}[\alpha V_1 + (1-\alpha)V_2]\rangle = \alpha\langle\hat{H}[V_1]\rangle + (1-\alpha)\langle\hat{H}[V_2]\rangle$. By definition of the infimum, this expression is bounded below by $\alpha E_0[V_1,N] + (1-\alpha)E_0[V_2,N]$. Finally, we remember that the infimum is also the *largest* lower bound and so in particular we have

$$E_0[\alpha V_1 + (1-\alpha)V_2,N] \ge \alpha E_0[V_1,N] + (1-\alpha)E_0[V_2,N] \qquad (2.16)$$

QED.

On top of the above properties, one can prove that $E_0[V,N]$ seen as a functional of the external potential V is *lower semi-continuous* [29]. We now come to the universal functional. We define the **Universal Functional** of the density as

$$H[n,N] = \sup_{V\in\mathcal{V}}\left[E_0[V,N] - \int_{\mathbb{T}^3}V(\vec{x})n(\vec{x})d\vec{x}\,\Big|\,\int_{\mathbb{T}^3}n(\vec{x})d\vec{x} = N\right] \qquad (2.17)$$

As one can readily observe, the above functional does not depend on the external potential since we compute a *supremum* over the space of external potentials. It is nothing else than the Legendre transform of the energy functional $-E_0[V,N]$ (which is a convex functional in V). The basic variable of the universal functional is the electron density n. There is also an N dependence to the functional through the constraint on n to integrate to the total number of electrons.

2.3.2 The Hohenberg-Kohn Variational Principle

Having proven the *convexity* of $-E_0[V,N]$ and mentioned its *lower semi-continuity*, we now use theorem 1.6.1 to prove the existence of the following **Variational Principle** (that we will call the Hohenberg-Kohn Variational Principle in honor of the pioneers of Density Functional Theory [41])

$$E_0[V,N] = \inf_{n \in \mathscr{R}_N} \left[H[n,N] + \int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x} \right] \qquad (2.18)$$

Pierre Hohenberg (1934-)

This completes our presentation of the foundations of Density Functional Theory. We succeeded in replacing the original Rayleigh-Ritz variational principle (in which the basic variable is either the many-body wavefunction or the density operator) with a variational principle whose basic variable is the electron density n.

Practical Density Functional Theory

The previous chapter was an attempt to present the field of Density Functional Theory in all its glory. The **conceptual** change introduced by DFT was fundamental and beautiful in its simplicity. It is a new way of looking at solving the Schrödinger equation that is, in theory, much easier to deal with than the original Rayleigh-Ritz variational principle. Nevertheless, the universal functional is elusive and its definition in equation 2.17 is hardly illuminating. We need a way to tackle the challenge of finding that functional. This is where the Kohn-Sham [52] approach really changed everything for practical DFT.

3.1 Total Energy for a Non-Interacting System

The breakthrough by Kohn and Sham came about from a simple observation. If the Hohenberg-Kohn universal functional exists and is exact for all possible *electron-electron* interactions, it also holds for *no interaction* at all. Let us then go back to the definition of the total energy in equation 2.14

$$E_0^{\mathrm{ni}}[V,N] = \inf_{\hat{\Gamma}}\left[\mathrm{tr}((\hat{T}+\hat{V})\hat{\Gamma})|\mathrm{tr}(\hat{N}\hat{\Gamma}) = \langle\hat{N}\rangle = N\right]$$

$$= \inf_{\{p_i\}}\left[\sum_i p_i\langle\psi_i|\hat{T}+\hat{V}|\psi_i\rangle|\sum_i p_i N_i = N\right]$$

where the superscript ni refers to the fact that the electron-electron interaction has been "switched-off". What is remarkable in this expression is the fact that one can compute all the terms in the expectation value of the Hamiltonian $\hat{H}_0 = \hat{T}+\hat{V}$ from the sole knowledge of the **one body reduced density matrix** $\gamma(\vec{x},\vec{x}')$ (1-RDM). What is this 1-RDM ? it is simply a generalization of the electron density

$$\gamma(\vec{x},\vec{x}') = \int_{\mathbb{T}^3}\psi(\vec{x},\vec{x}_2,...,\vec{x}_M)\psi^*(\vec{x}',\vec{x}_2,...,\vec{x}_M)d\vec{x}_2...d\vec{x}_M \qquad (3.1)$$

$$\text{or}$$

$$\gamma(\vec{x},\vec{x}') = \sum_i p_i\int_{\mathbb{T}^3}\psi_i(\vec{x},\vec{x}_2,...,\vec{x}_{N_i})\psi_i^*(\vec{x}',\vec{x}_2,...,\vec{x}_{N_i})d\vec{x}_2...d\vec{x}_{N_i}$$

depending on whether one decides to describe the system using wavefunctions or density operators. The "diagonal" of that "matrix" is nothing else than the electron

density $\gamma(\vec{x}, \vec{x}) = n(\vec{x})$. The mathematical expressions for the kinetic energy and the external potential in terms of the 1-RDM are

$$\langle \hat{T} \rangle = -\frac{1}{2} \int_{\mathbb{T}^3} \vec{\nabla}^2_{\vec{x}} \gamma(\vec{x}, \vec{x}')|_{\vec{x}'=\vec{x}} d\vec{x}$$

and

$$\langle \hat{V} \rangle = \int_{\mathbb{T}^3} V(\vec{x}) \gamma(\vec{x}, \vec{x}) d\vec{x}$$

The first expression concerning the kinetic energy can be read as follows : first take the second partial derivative of the function $\gamma(\vec{x}, \vec{x}')$ with respect to the variable $\vec{x}$. Second, in the final expression substitute all the $\vec{x}'$ for $\vec{x}$. What have we gained from all this ? After all, we just replaced a search over the space of density operators by a search over the space of 1-RDM coming from a density operator. In the next section we will see how we can exactly represent a 1-RDM in terms of orthonormal orbitals and occupations numbers. That representation is **always** exact, even when we deal with non zero electron-electron interactions. This will allow us to introduce a much more tractable but entirely equivalent variational principle.

3.2 The N-representability of 1-RDM

Having spent time reformulating the non-interacting problem in terms of the one-body reduced density matrix, we will now introduce a crucial representation result allowing us to translate the problem at hand from a search over the space of 1-RDM to a search over a set of much more tractable objects. Indeed a theorem due to Coleman [15] states how one can completely represent a one-body reduced density matrix (1-RDM) coming from a density operator, in terms of an infinite set of orthonormal single particle orbitals $\phi_i(\vec{x})$ and occupation numbers f_i. In short, this theorem states that an admissible 1-RDM can be written as

$$\gamma(\vec{x}, \vec{x}') = \sum_i f_i \phi_i(\vec{x}) \phi_i^*(\vec{x}') \tag{3.2}$$

with

$$0 \le f_i \le 1$$

$$\int_{\mathbb{T}^3} \phi_i(\vec{x}) \phi_j^*(\vec{x}) d\vec{x} = \delta_{ij}$$

Using theorem 3.2, we are now in a position to practically compute the total energy of a *non-interacting* system using the following optimization statement which refor-

mulates equation 2.14

$$E_0^{\mathrm{ni}}[V,N] = \inf_{\{\phi_i,f_i\}} \left[\frac{1}{2}\sum_i f_i \int_{\mathbb{T}^3} |\vec{\nabla}\phi_i(\vec{x})|^2 d\vec{x} + \int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x} \; | \right.$$
$$n(\vec{x}) = \sum_i f_i|\phi_i(\vec{x})|^2,$$
$$f_i \in [0,1],$$
$$\sum_i f_i = N,$$
$$\left. \int_{\mathbb{T}^3} \phi_i(\vec{x})\phi_j^*(\vec{x})d\vec{x} = \delta_{ij} \right]$$

The above optimization problem is particularly amenable to numerical calculations. So far of course we have been dealing with non-interacting systems, but we will see shortly how that knowledge can guide us in splitting the Universal Functional of equation 2.17 into terms that we know and that are all functionals of the electron density.

3.3 Kohn-Sham Density Functional Theory

The fundamental idea by Kohn and Sham [52] was to postulate that the kinetic energy of an *interacting* system is well approximated by the *non-interacting* one.

What's more, the classical Coulomb electrostatic energy is a fair guess at the electron-electron interaction energy and it has the advantage of being a natural functional of the electron density. Last but not least the external potential energy is exactly known as a functional of the electron density. The approach then consists in introducing the Legendre transform to the total energy of a non-interacting system $T_J[n,N]$, that we will call the Janak [31] kinetic energy functional

Lu Sham

$$T_J[n,N] = \sup_{V \in \mathscr{V}} \left[E_0^{\mathrm{ni}}[V,N] - \int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x} | \int_{\mathbb{T}^3} n(\vec{x})d\vec{x} = N \right]$$

Using theorem 3.2 one can prove that the previous expression is equivalent to the more tractable one

$$
T_J[n,N] = \inf_{\{\phi_i, f_i\}} \left[\frac{1}{2} \sum_i f_i \int_{\mathbb{T}^3} |\vec{\nabla}\phi_i(\vec{x})|^2 d\vec{x} \ | \right.
$$
$$
n(\vec{x}) = \sum_i f_i |\phi_i(\vec{x})|^2,
$$
$$
f_i \in [0,1],
$$
$$
\sum_i f_i = N,
$$
$$
\left. \int_{\mathbb{T}^3} \phi_i(\vec{x})\phi_j^*(\vec{x})d\vec{x} = \delta_{ij} \right]
$$

(3.3)

The classical Coulomb electrostatic energy associated with a charge density $\rho(\vec{x}) = -en(\vec{x})$ is

$$
J[n] = \frac{e^2}{2} \int_{\mathbb{T}^3 \times \mathbb{T}^3} \frac{n(\vec{x})n(\vec{x}')}{|\vec{x} - \vec{x}'|} d\vec{x}d\vec{x}'
$$

(3.4)

and the external potential energy is just

$$
\int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x}
$$

(3.5)

Kohn and Sham finally chose to **define** the **exchange and correlation** functional $E_{XC}[n]$ as

$$
H[n,N] = T_J[n,N] + J[n] + E_{XC}[n]
$$

(3.6)

This, in essence, constitutes what many call the "Kohn-Sham trick". But it is important to realize that it is absolutely exact in principle. It is merely a translation of our ignorance about the Universal Functional into that of the Exchange and Correlation Functional. Hopefully, the made explicit functionals are already accounting for a large part of the total energy and in fact they do. Now that we have our Universal Functional, we use theorem 3.2 and plug everything into the Hohenberg-Kohn Variational Principle in 2.18 to find

$$
E_0[V,N] = \inf_{\{\phi_i, f_i\}} \left[\frac{1}{2} \sum_i f_i \int_{\mathbb{T}^3} |\vec{\nabla}\phi_i(\vec{x})|^2 d\vec{x} + \int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x} \right.
$$
$$
+ \frac{e^2}{2} \int_{\mathbb{T}^3 \times \mathbb{T}^3} \frac{n(\vec{x})n(\vec{x}')}{|\vec{x} - \vec{x}'|} d\vec{x}d\vec{x}' + E_{XC}[n] \ |
$$
$$
n(\vec{x}) = \sum_i f_i |\phi_i(\vec{x})|^2,
$$
$$
f_i \in [0,1],
$$
$$
\sum_i f_i = N,
$$
$$
\left. \int_{\mathbb{T}^3} \phi_i(\vec{x})\phi_j^*(\vec{x})d\vec{x} = \delta_{ij} \right]
$$

(3.7)

We then introduce the following energy functional, whose basic variables are the real numbers f_i and complex-valued orbitals $\phi_i(\vec{x})$ and $\phi_i^*(\vec{x})$

$$
\begin{aligned}
E[\{f_i\}, \{\phi_i(\vec{x})\}, \{\phi_i^*(\vec{x})\}] \quad = \quad & \frac{1}{2}\sum_i f_i \int_{\mathbb{T}^3} |\vec{\nabla}\phi_i(\vec{x})|^2 d\vec{x} + \int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x} \qquad \boxed{3.8} \\
+ \quad & \frac{e^2}{2}\int_{\mathbb{T}^3 \times \mathbb{T}^3} \frac{n(\vec{x})n(\vec{x}')}{|\vec{x}-\vec{x}'|}d\vec{x}d\vec{x}' + E_{XC}[n] \\
= \quad & H[n,N] + \int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x}
\end{aligned}
$$

It is important not to confuse the total energy *function*, $E_0[V,N]$, which delivers a number which is the true total energy of the system for a fixed external potential $V(\vec{x})$ and fixed total electron number N, from the energy *functional*, $E[\{f_i\}, \{\phi_i(\vec{x})\}, \{\phi_i^*(\vec{x})\}]$, which is a mathematical function and not a physical observable.

3.4 The Kohn-Sham Equations

The last section left us with a workable variational principle in which the basic variables are single particle orthonormal orbitals ϕ_i and non-negative occupation numbers f_i. From the mathematical point of view, we are facing a non-linear optimization problem with equality constraints. In order to move forward we first need to ask ourselves if the Universal Functional is **differentiable**.That question has been answered in the affirmative by Englisch and Englisch [28]. Recently Eschrig [30] proved the existence of functional derivatives for the Universal Functional even in the case of finite temperature and non-collinear magnetism. Finally, in practice, all the approximations to the Exchange-Correlation functional are differentiable [31].

In order to solve the variational problem, we will use the method of Lagrange multipliers outlined in section 1.7. We introduce the following alternative Functional

$$
\begin{aligned}
F[n,N] = H[n,N] \quad + \quad & \int_{\mathbb{T}^3} V(\vec{x})n(\vec{x})d\vec{x} \qquad\qquad\qquad \boxed{3.9} \\
- \quad & \epsilon_F \left(\sum_i f_i - N \right) \\
- \quad & \sum_{ij} \lambda_{ij} \left(\int_{\mathbb{T}^3} \phi_i(\vec{x})\phi_j^*(\vec{x})d\vec{x} - \delta_{ij} \right)
\end{aligned}
$$

3.4.1 Extremality Conditions

Using the newly defined functional ($F[n,N]$), we will now write the Euler-Lagrange equations. Those equations will then give us a lot of information concerning the extrema of the functional. Since the variables over which we minimize the functional are f_i, ϕ_i and ϕ_i^* (here ϕ_i and ϕ_i^* have to be considered "independent" variables since the orbitals are complex functions), we will write 3 Euler-Lagrange equations. For this we use what we learned about functional derivatives in section 1.5.

The first Euler-Lagrange equation is $\frac{\delta F}{\delta \phi_i(\vec{x})} = 0$. At this point we will make an assumption about the functional dependence of the Exchange-Correlation part E_{XC}. We will only consider functionals that depend on the *single orbital densities* $\rho_i(\vec{x}) = f_i \phi_i(\vec{x}) \phi_i^*(\vec{x})$. In particular, a functional of the total electron density is one such functional because it depends only on $n(\vec{x}) = \sum_i \rho_i(\vec{x})$. More general functionals also fall into that category, and we will be more interested in the ones that naturally come out of Self-Interaction correction schemes. Given that hypothesis we now express the functional derivative of the Exchange-Correlation part as

For a functional of the electron density $n(\vec{x})$:
$$\frac{\delta E_{XC}}{\delta \phi_i(\vec{x})} = \int_{\mathbb{T}^3} \frac{\delta E_{XC}}{\delta n(\vec{x}')} \frac{\delta n(\vec{x}')}{\delta \phi_i(\vec{x})} d\vec{x}' \qquad \boxed{3.10}$$
For a functional of the orbital densities $\rho_i(\vec{x})$:
$$\frac{\delta E_{XC}}{\delta \phi_i(\vec{x})} = \int_{\mathbb{T}^3} \frac{\delta E_{XC}}{\delta \rho_i(\vec{x}')} \frac{\delta \rho_i(\vec{x}')}{\delta \phi_i(\vec{x})} d\vec{x}'$$

In the following we will give the results for the functional derivatives of the Exchange-Correlation functional for both cases at the same time in parenthesis.

Since $n(\vec{x}) = \sum_i \rho_i(\vec{x})$ and $\rho_i(\vec{x}) = f_i \phi_i(\vec{x}) \phi_i^*(\vec{x})$, we simply have

$$\frac{\delta n(\vec{x}')}{\delta \phi_i(\vec{x})} = \frac{\delta \rho_i(\vec{x}')}{\delta \phi_i(\vec{x})} = f_i \phi_i^*(\vec{x}) \delta(\vec{x} - \vec{x}')$$

Using the functional derivatives of section 1.5, the first Euler-Lagrange equation reads

$$f_i \left[-\frac{1}{2} \vec{\nabla}^2 \phi_i^*(\vec{x}) \quad + \quad V(\vec{x}) \phi_i^*(\vec{x}) + e^2 \int_{\mathbb{T}^3} \frac{n(\vec{x}')}{|\vec{x} - \vec{x}'|} d\vec{x}' \phi_i^*(\vec{x}) \right. \qquad \boxed{3.11}$$
$$\left. + \quad \left(\begin{array}{c} \frac{\delta E_{XC}}{\delta n(\vec{x})} \\ \frac{\delta E_{XC}}{\delta \rho_i(\vec{x})} \end{array} \right) \phi_i^*(\vec{x}) \right] = \sum_j \lambda_{ij} \phi_j^*(\vec{x})$$

The second Euler-Lagrange equation $\frac{\delta F}{\delta \phi_i^*(\vec{x})} = 0$ is basically almost equivalent to the preceding one. The reason being that the functionals only depend on the electron densities ρ_i which are symmetric functions in ϕ_i and ϕ_i^*. We then directly give the second Euler-Lagrange equation

$$f_i \left[-\frac{1}{2} \vec{\nabla}^2 \phi_i(\vec{x}) \quad + \quad V(\vec{x}) \phi_i(\vec{x}) + e^2 \int_{\mathbb{T}^3} \frac{n(\vec{x}')}{|\vec{x} - \vec{x}'|} d\vec{x}' \phi_i(\vec{x}) \right. \qquad \boxed{3.12}$$
$$\left. + \quad \left(\begin{array}{c} \frac{\delta E_{XC}}{\delta n(\vec{x})} \\ \frac{\delta E_{XC}}{\delta \rho_i(\vec{x})} \end{array} \right) \phi_i(\vec{x}) \right] = \sum_j \lambda_{ji} \phi_j(\vec{x})$$

Note that the right-hand side of equation 3.12 uses the transpose of the λ matrix as opposed to the right-hand side of equation 3.11 that uses the λ matrix directly.

The third Euler-Lagrange equation has to do with the variation with respect to f_i. Of course the f_i are constraint to be within 0 and 1. To write the variation with f_i in an unconstrained way we use the following trick. We will write $f_i = \sin^2(\theta_i)$ so that θ_i is free to vary in the whole of $\mathbb{R}$. The third Euler-Lagrange equation is then $\frac{\delta F}{\delta \theta_i} = 0$. Using the analytical expression for f_i in terms of θ_i, we obtain $\frac{\delta F}{\delta \theta_i} = \frac{\delta F}{\delta f_i}\frac{\partial f_i}{\partial \theta_i} = 0$. The Euler-Lagrange equation is then

$$\left[\frac{1}{2}\int_{\mathbb{T}^3}|\vec{\nabla}\phi_i(\vec{x})|^2 \quad + \quad \int_{\mathbb{T}^3}V(\vec{x})|\phi_i(\vec{x})|^2 + e^2\int_{\mathbb{T}^3\times\mathbb{T}^3}\frac{n(\vec{x}')}{|\vec{x}-\vec{x}'|}|\phi_i(\vec{x})|^2 d\vec{x}'d\vec{x}\right.$$
$$\left. + \quad \int_{\mathbb{T}^3}\begin{pmatrix}\frac{\delta E_{XC}}{\delta n(\vec{x})}\\\frac{\delta E_{XC}}{\delta \rho_i(\vec{x})}\end{pmatrix}|\phi_i(\vec{x})|^2 d\vec{x} - \epsilon_F\right]\sin(2\theta_i) = 0 \qquad (3.13)$$

3.4.2 Constraints on f_i and λ_{ij}

From equations 3.13, 3.11 and 3.12 we can deduce many interesting properties concerning the λ matrix. Indeed if one takes the complex conjugate of equation 3.12 and compares that equation to 3.11, one readily obtains

$$\sum_j \lambda_{ij}\phi_j(\vec{x}) = \sum_j \lambda_{ji}^*\phi_j(\vec{x}) \quad \forall\vec{x}$$

Using the orthonormality constraints, we can easily deduce the **hermiticity** of the λ matrix

$$\lambda_{ji}^* = \lambda_{ij} \qquad (3.14)$$

Moreover, multiplying equation 3.12 by ϕ_i^* and integrating over $\vec{x}$ leads us to the following expression for the *diagonal* elements of λ

$$f_i\left[\frac{1}{2}\int_{\mathbb{T}^3}|\vec{\nabla}\phi_i(\vec{x})|^2 \quad + \quad \int_{\mathbb{T}^3}V(\vec{x})|\phi_i(\vec{x})|^2 + e^2\int_{\mathbb{T}^3\times\mathbb{T}^3}\frac{n(\vec{x}')}{|\vec{x}-\vec{x}'|}|\phi_i(\vec{x})|^2 d\vec{x}'d\vec{x}\right.$$
$$\left. + \quad \int_{\mathbb{T}^3}\begin{pmatrix}\frac{\delta E_{XC}}{\delta n(\vec{x})}\\\frac{\delta E_{XC}}{\delta \rho_i(\vec{x})}\end{pmatrix}|\phi_i(\vec{x})|^2 d\vec{x}\right] = \lambda_{ii} \qquad (3.15)$$

Concerning f_i, equation 3.13 tells us that

- Either $\sin(2\theta_i) = 0$ and then $2\theta_i = 0\ [\pi]$ so $f_i = 0$ or $f_i = 1$

- Or $\sin(2\theta_i) \neq 0$ and then $f_i \in]0,1[$ and the second term is 0

In the second case (strict fractional occupation of orbital ϕ_i), using equation 3.15 simplifies the constraint to

$$\frac{\lambda_{ii}}{f_i} = \epsilon_F \longrightarrow \lambda_{ii} = f_i \epsilon_F \tag{3.16}$$

In the case where $f_i = 0$, equation 3.11 reduces to

$$\sum_j \lambda_{ij} \phi_j^*(\vec{x}) = 0 \quad \forall \vec{x}$$

which, when using the orthonormality constraints, gives

$$\lambda_{ij} = 0 \quad \forall j \text{ when } f_i = 0$$

The hermiticity of λ makes us conclude that **only occupied** orbitals (full occupation 1, or fractional occupation $f_i \in]0,1[$) couple to each other.

3.4.3 Expression for the Total Energy

We give here an explicit expression for the total energy of a quantum system. For this we go back to the definition of the total energy in terms of the single particle orbitals and occupation numbers in equation 3.7 and use equation 3.15 to find

If E_{XC} only depends on $n(\vec{x})$ (3.17)

$$E_0[V,N] = \sum_i \lambda_{ii} \quad - \quad \frac{e^2}{2} \int_{\mathbb{T}^3 \times \mathbb{T}^3} \frac{n(\vec{x}')n(\vec{x})}{|\vec{x} - \vec{x}'|} d\vec{x}' d\vec{x}$$
$$+ \quad E_{XC}[n] - \int_{\mathbb{T}^3} \frac{\delta E_{XC}}{\delta n(\vec{x})} n(\vec{x}) d\vec{x}$$

If E_{XC} depends on each $\rho_i(\vec{x})$ (3.18)

$$E_0[V,N] = \sum_i \lambda_{ii} \quad - \quad \frac{e^2}{2} \int_{\mathbb{T}^3 \times \mathbb{T}^3} \frac{n(\vec{x}')n(\vec{x})}{|\vec{x} - \vec{x}'|} d\vec{x}' d\vec{x}$$
$$+ \quad E_{XC}[n] - \sum_i \int_{\mathbb{T}^3} \frac{\delta E_{XC}}{\delta \rho_i(\vec{x})} f_i |\phi_i(\vec{x})|^2 d\vec{x}$$

3.4.4 Further Constraints Due to Minimality

So far, we have not used the fact that the extremum is actually a *minimum*. If it is the case, then an **Aufbau Principle** can be established. Indeed, let us start out from the minimum with given occupation numbers f_i and orbitals ϕ_i. Let us imagine that we slightly perturb our system by depleting a fully occupied orbital $f_i = 1$ and at the same time filling a partially occupied or empty orbital $0 \leq f_j < 1$. Since the perturbation must be number conserving (indeed $\sum_k f_k$ always equals N), we must have $\delta f_i = -\delta f_j = -\delta f$ with $\delta f > 0$. The change in total energy must then be **positive**

because we started from the minimum state. The energy change can be written as

$$\delta E_0 = \frac{\partial E}{\partial f_i}\delta f_i + \frac{\partial E}{\partial f_j}\delta f_j = \left(\frac{\partial E}{\partial f_j} - \frac{\partial E}{\partial f_i}\right)\delta f$$

Using equations 3.13 and 3.15 we can simplify the above to

$$\delta E_0 = \left(\langle\phi_j|\hat{H}_j\phi_j\rangle - \langle\phi_i|\hat{H}_i\phi_i\rangle\right)\delta f > 0$$

This result is very important. It states that fully occupied orbitals have an orbital energy $\langle\phi_i|\hat{H}_i\phi_i\rangle$ less or equal than the orbital energies of fractionally occupied states and empty states. Using a similar reasoning one can also prove that fractionally occupied states (for which the orbital energy is equal to the Fermi energy because of equation 3.13) have their orbital energies below the ones of the empty states. It is worth noting that if the frontier orbitals have full occupation then their orbital energy is not necessarily equal to the Fermi energy. Only if fractional occupations exist, do we have the Fermi energy equal to one of the orbital energy at least. We also see that, unless some states are fractionally occupied (and at the top of the "Fermi sea"), then all occupations are 1.

As a consequence, we can arrange the occupied orbitals in ascending value of $\langle\phi_i|\hat{H}_i\phi_i\rangle$, which means that we have an Aufbau Principle. We will see later on that this principle is of paramount importance to give a proper physical meaning to the Fermi energy ϵ_F.

3.5 Defining the Kohn-Sham "Hamiltonian"

Let us define the following family of operators

$$\hat{H}_i = -\frac{1}{2}\vec{\nabla}^2 + V(\vec{x}) + e^2\int_{\mathbb{T}^3}\frac{n(\vec{x}')}{|\vec{x} - \vec{x}'|}d\vec{x}' + \left(\begin{array}{c}\frac{\delta E_{XC}}{\delta n(\vec{x})}\\[6pt]\frac{\delta E_{XC}}{\delta\rho_i(\vec{x})}\end{array}\right) \tag{3.19}$$

If we go back to equation 3.11, we see that when one multiplies to the left by $\phi_j^*(\vec{x})$ and integrate over $\vec{x}$, then using the orthonormality condition, one can easily arrive at

$$\lambda_{ji} = f_i\langle\phi_j|\hat{H}_i\phi_i\rangle \tag{3.20}$$

Now using 3.14, we find

$$f_i\langle\phi_j|\hat{H}_i\phi_i\rangle = \left(f_j\langle\phi_i|\hat{H}_j\phi_j\rangle\right)^* \tag{3.21}$$

In all generality, it is the λ matrix which is hermitian at the extremum of the functional. In the case where the functional only depends on the electron density $n(\vec{x})$, we will introduce an hamiltonian $\hat{h}_{KS}$ that will not depend directly on occupation numbers. When the functional is orbital dependent, the result will be similar only

when all occupation numbers are unity.

3.5.1 Case where E_{XC} only depends on $n(\vec{x})$

Let us suppose that the Exchange and Correlation functional does not depend on each electron densities $\rho_i(\vec{x}) = f_i |\phi_i(\vec{x})|^2$ but rather on the total electron density alone $n(\vec{x})$. This situation is by far the most common in practice. Then equation 3.19 demonstrates that $\hat{H}_i$ is the same for all the orbitals. What's more, this operator (seen as an operator that takes a wavefunction $\phi_i(\vec{x})$ and returns another wavefunction $\hat{H}\phi_i(\vec{x})$) is **self-adjoint**. This implies the following identity

$$\langle \phi_j | \hat{H}\phi_i \rangle = \langle \phi_i | \hat{H}\phi_j \rangle^*$$

which, coupled to equation 3.21, leads to

$$(f_i - f_j)\langle \phi_j | \hat{H} | \phi_i \rangle = 0 \quad \text{for all } i \text{ and } j \tag{3.22}$$

Equation 3.22, coupled with the Aufbau principle, teaches us that if some states are fractionally occupied, then those are *decoupled* from the fully occupied states. More generally, if we group the orbitals in subsets with identical occupation numbers, then those subsets are *decoupled* to each other whenever the occupations are distinct. In all cases, we can *define* the Kohn-Sham hamiltonian to be

$$\hat{h}_{KS} = \sum_{ij} |\phi_i\rangle h_{ij} \langle \phi_j| \quad \text{with } h_{ij} = \langle \phi_i | \hat{H} | \phi_j \rangle \tag{3.23}$$

An example of a Kohn-Sham hamiltonian for a system with 5 fully occupied orbitals, 1 orbital with fractional occupation f and 2 orbitals with equal fractional occupation $f' \neq f$ is shown below

$$\hat{h}_{KS} = \begin{pmatrix} * & * & * & * & * & 0 & 0 & 0 \\ * & * & * & * & * & 0 & 0 & 0 \\ * & * & * & * & * & 0 & 0 & 0 \\ * & * & * & * & * & 0 & 0 & 0 \\ * & * & * & * & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * \end{pmatrix}$$

Let us introduce a unitary transformation that brings any subset of equally occupied minimizing orbitals ϕ_i to another set of orbitals $\psi_k = \sum_l U_{kl}\phi_l$ that spans the same vector space. We can prove that both the *charge density* and the *total energy* are unaffected[1] by this change of basis. Indeed, the contribution of any subset of orbitals

---

[1]By "unaffected" we mean that the mathematical expressions for the density and the total energy are *covariant*.

to the charge density can be written as

$$n'(\vec{x}) = \sum_i |\phi_i(\vec{x})|^2 = \sum_i \phi_i^*(\vec{x})\phi_i(\vec{x})$$

for i in the prescribed subset. Using the inverse relation $\phi_i = \sum_k U_{ik}^{-1}\psi_k$, the above equality becomes

$$n'(\vec{x}) = \sum_i \left(\sum_k [U_{ik}^{-1}]^*\psi_k^*(\vec{x})\right)\left(\sum_l U_{il}^{-1}\psi_l(\vec{x})\right) = \sum_{k,l} \psi_k^*(\vec{x})\left(\sum_i [U_{ik}^{-1}]^* U_{il}^{-1}\right)\psi_l(\vec{x})$$

now since the U matrix is *unitary* we have $U^{-1} = U^\dagger = [U^T]^*$. So $[U_{ik}^{-1}]^* = U_{ki}$. Going back to the expression for the charge density we arrive at

$$n'(\vec{x}) = \sum_{k,l} \psi_k^*(\vec{x})\left(\sum_i [U_{ik}^{-1}]^* U_{il}^{-1}\right)\psi_l(\vec{x}) = \sum_{k,l} \psi_k^*(\vec{x})\delta_{kl}\psi_l(\vec{x})$$

which simplifies into

$$n'(\vec{x}) = \sum_k |\psi_k(\vec{x})|^2$$

We observe that inside any subset of orbitals with equal occupations, we can change basis at will, without modifying the contribution to the charge density. Let us now look at the total energy in equation 3.17. Its expression only depends on the total charge density and on the **trace** of the lambda matrix. We already realized that the charge density will be unchanged by rotating the basis inside every subset, so is there some invariance of the trace of the lambda matrix with an arbitrary rotation of the subsets ? The answer is of course yes. To see this, let us express the trace of the lambda matrix

$$\text{tr}(\lambda) = \sum_i \lambda_{ii} = \sum_{\text{subset } G} f_G \sum_{j \in G} \langle \phi_j | \hat{H} | \phi_j \rangle$$

where G is a generic subset of orbitals ϕ_j with equal occupation numbers $f_j = f_G \ \forall j$. We can of course replace $\sum_{j \in G} \langle \phi_j | \hat{H} | \phi_j \rangle$ with the trace over the subset G. But since the trace is invariant by a change of basis we see that in the end the trace of λ is also unchanged by a change of basis in each subset.

All in all, we see that from the point of view of mathematics, *any* set that is equivalent in each subset (up to a unitary rotation) to the set of minimizing orbitals, can be used as a valid description of the Kohn-Sham system. In particular, we may choose to take some unitary rotations that diagonalize the sub-blocks of $\hat{h}_{KS}$. In this case we end up with a set of optimality conditions (equations 3.11 and 3.12) in which the right-hand side reduces to $\epsilon_i\phi_i(\vec{x})$ and $\epsilon_i\phi_i^*(\vec{x})$ respectively. Those equations are generally called **canonical** Kohn-Sham equations. In that basis the Kohn-Sham hamiltonian, as we defined it, is diagonal and does not explicitly depend on the occupation numbers. With this choice we also observe that the orbital energies $\langle \phi_i | \hat{H} | \phi_i \rangle$ correspond to the eigenvalues of the Kohn-Sham hamiltonian.

3.5.2 Case where E_{XC} is orbital-dependent

In the case where the functional is orbital dependent, we cannot duplicate the results of the previous section. If at the minimum all the occupation numbers are unity, then the matrix

$$\langle \phi_i | \hat{H}_j | \phi_j \rangle$$

is hermitian and we can define the Kohn-Sham hamiltonian to correspond to that matrix. If some of the orbitals are fractionally occupied, then for those, the diagonal element of the above matrix is equal to the Fermi energy

$$\langle \phi_i | \hat{H}_i | \phi_i \rangle = \epsilon_F$$

Part II

The basic tools in Electronic Structure Theory

4

Translation Symmetry and its Consequences

4.1 Introduction

If one tries to encapsulate the whole idea of Density Functional Theory in just a few words, one could argue that the most important consequence of the Hohenberg and Kohn theorems and the Kohn-Sham method has been to transform the original many-body Schrödinger equation into an *effective one-body* equation. From the Kohn-Sham equations, we have seen previously that we can extract a Kohn-Sham hamiltonian in its matrix representation. What's more, we have seen that what we really care about at the end of the calculation, is to find the eigenvalues and eigenvectors of that matrix.

In this part, we will assume that the Kohn-Sham hamiltonian matrix is a given. We will also assume some translational symmetries for the system to further simplify the search for the eigenvalues and eigenvectors of that matrix. In doing so we will introduce powerful tools used in electronic structure theory that are of very wide applicability. This part can be considered as a short review of Band Theory [9].

Whenever we will need an operational form for the Kohn-Sham hamiltonian, we will use the following real-space representation

$$\hat{h}_{KS} = -\frac{1}{2}\vec{\nabla}^2 + V(\vec{x}) \text{ with } V(\vec{x}) \text{ periodic in space} \qquad \boxed{4.1}$$

4.2 The Fundamental Translation Symmetry

There are two main reasons why we would want to study systems with inherent translational symmetries. The first reason has to do with the fact that we are, in practice, very interested in studying crystalline solids like Silicon, diamond or carbon nanotubes. The second reason has to do with the very fabric of most Density Functional Theory computer codes that embed systems in a box and assume periodic boundary conditions. This seemingly simple property, the one of translational symmetry, is of tremendous consequences for the Quantum Mechanics of a single particle. We will start by describing the notions of direct lattice and reciprocal lattice.

Then we will move on to the main result, called **Bloch's Theorem** and discuss its consequences in terms of labeling of the eigenstates of the system.

4.2.1 Direct and Reciprocal Lattices

A periodic system is characterized by a fundamental unit, called the *elementary unit cell* or unit cell for short, that is repeated identically in one, two, or three dimensions depending on the dimensionality of the system. We show examples of 1D periodic structures on figure 4-1. One can entirely characterize a periodic system if one knows
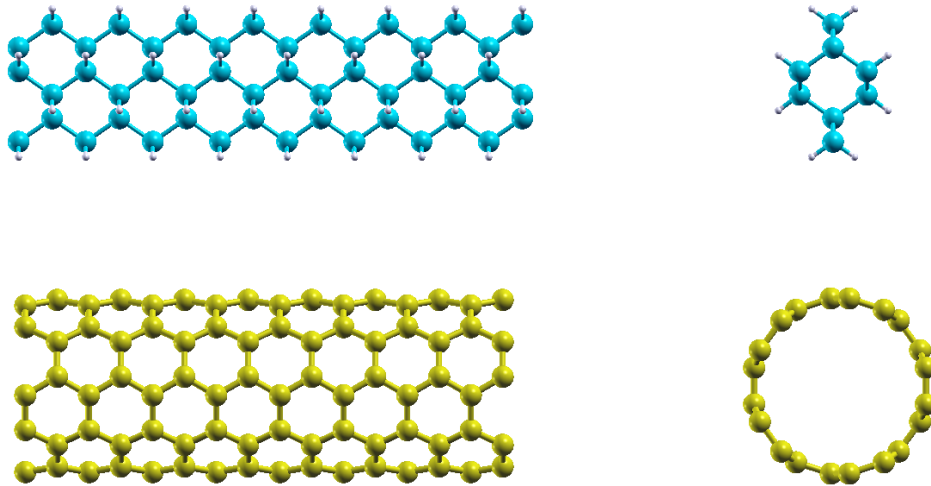
Figure 4-1: Some examples of periodic systems. Here a Silicon Nanowire (top) and a carbon nanotube (bottom) are displayed in both side view and cross sectional view. It is easy to identify the fundamental unit of translation (called the elementary unit cell) that repeats itself *ad infinitum*.

2 fundamental objects [46] : the *elementary unit cell* and the *Bravais lattice*. The Bravais lattice is just the collection of points in space that are images of a seed point by all possible translations compatible with the system's symmetries. The set of all such compatible translations is characterized by 1, 2, or 3 "basis vectors" depending on the dimensionality of the system. We will generically use the notation $\vec{a}_1$, $\vec{a}_2$ and $\vec{a}_3$ to specify those basis vectors and the notation $\vec{R}$ to denote a generic vector representing one of the compatible translation. The fundamental property of those translation vectors $\vec{R}$ is to have a unique decomposition in the basis $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$ with **integer** coordinates

$$\vec{R} = r_1 \vec{a}_1 + r_2 \vec{a}_2 + r_3 \vec{a}_3 \text{ with } (r_1, r_2, r_3) \in \mathbb{Z}^3 \qquad (4.2)$$

When referring to basis vectors or the Bravais lattice or translation vectors, we will generically use the word **direct space** to describe the space in which all those elements "live".

We now introduce another Bravais lattice, called the *reciprocal lattice* that is defined in such a way that its "basis vectors" (the analogous to $\vec{a}_1$, $\vec{a}_2$, and $\vec{a}_3$) verify

$$\vec{b}_i.\vec{a}_j = 2\pi\delta_{ij} \qquad (4.3)$$

where δ_{ij} is the Kronecker symbol (which equals 1 if $i = j$ and 0 otherwise). From equation 4.3, one can easily arrive at the following formulas for the reciprocal basis vectors expressed in the $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$ basis

For the 1D case: $\qquad (4.4)$

$$\vec{b}_1 \;=\; \frac{2\pi}{||\vec{a}_1||^2}\vec{a}_1$$

For the 2D case:

$$\vec{b}_1 \;=\; \frac{2\pi}{||\vec{a}_1||^2||\vec{a}_2||^2 - (\vec{a}_1.\vec{a}_2)^2}\left(||\vec{a}_2||^2\vec{a}_1 - (\vec{a}_1.\vec{a}_2)\,\vec{a}_2\right)$$

$$\vec{b}_2 \;=\; \frac{2\pi}{||\vec{a}_1||^2||\vec{a}_2||^2 - (\vec{a}_1.\vec{a}_2)^2}\left(-(\vec{a}_1.\vec{a}_2)\,\vec{a}_1 + ||\vec{a}_1||^2\vec{a}_2\right)$$

For the 3D case:

$$\vec{b}_1 \;=\; 2\pi\frac{\vec{a}_2 \times \vec{a}_3}{\vec{a}_1.(\vec{a}_2 \times \vec{a}_3)}$$

$$\vec{b}_2 \;=\; 2\pi\frac{\vec{a}_3 \times \vec{a}_1}{\vec{a}_1.(\vec{a}_2 \times \vec{a}_3)}$$

$$\vec{b}_3 \;=\; 2\pi\frac{\vec{a}_1 \times \vec{a}_2}{\vec{a}_1.(\vec{a}_2 \times \vec{a}_3)}$$

where the dot . means scalar product and the cross × means vectorial product. Any point in the reciprocal Bravais lattice can be generated from a seed point and one of the reciprocal vectors $\vec{G}$ whose coordinates in the $(\vec{b}_1, \vec{b}_2, \vec{b}_3)$ basis are **integer** numbers

$$\vec{G} = g_1\vec{b}_1 + g_2\vec{b}_2 + g_3\vec{b}_3 \text{ with } (g_1, g_2, g_3) \in \mathbb{Z}^3 \qquad (4.5)$$

In a similar manner as we introduced the word *direct space*, we will introduce the word **reciprocal space** to describe the space in which the reciprocal Bravais lattice, the reciprocal basis vectors and the reciprocal vectors "live".

To motivate the reason for introducing a reciprocal space, let us look at a representative function that will be periodic when the system is periodic, namely the electron density $n(\vec{x})$. If a function is such that

$$n(\vec{x} + \vec{R}) = n(\vec{x}) \text{ for all } \vec{R} \qquad (4.6)$$

(since the $\vec{R}$ vectors are naturally expressed in the $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$ basis, we will do the same for $\vec{x}$. The coordinates x, y, z of $\vec{x}$ will refer to the $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$ basis), then we know from the Theory of Fourier Series [96], that we can expand $n(\vec{x})$ in a series of

plane waves

For the 1D case: $\qquad\qquad\qquad\qquad\qquad\qquad$ (4.7)
$$n(x) \;=\; \sum_{\lambda} n_{\lambda} e^{2\pi i (x\lambda)}$$

For the 2D case:
$$n(x,y) \;=\; \sum_{\lambda\mu} n_{\lambda\mu} e^{2\pi i (x\lambda + y\mu)}$$

For the 3D case:
$$n(x,y,z) \;=\; \sum_{\lambda\mu\nu} n_{\lambda\mu\nu} e^{2\pi i (x\lambda + y\mu + z\nu)}$$

where $(\lambda,\mu,\nu) \in \mathbb{Z}^3$. Because of equation 4.3, we see that we can introduce the reciprocal vectors $\vec{G} = \lambda \vec{b}_1 + \mu \vec{b}_2 + \nu \vec{b}_3$ (with equivalent expressions in 1D and 2D) and simplify the expression of the argument in the complex exponential

$$2\pi i (x\lambda + y\mu + z\nu) = i\vec{G}\vec{x}$$

The above equation is the reason for introducing the concept of reciprocal space. It is the natural space in which to express the Fourier components of a periodic function in the direct space. With such a simplification, we can introduce a generic notation for the Fourier Series of a periodic function that will replace equation 4.7

$$n(\vec{x}) = \sum_{\vec{G}} n_{\vec{G}} e^{i\vec{G}\vec{x}} \qquad (4.8)$$

Integrating equation 4.8 over the *elementary unit cell* (characterized by the collection of all the points $\vec{x}$ for which the coordinates in the $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$ basis are between 0 and 1), we find the expression for the Fourier coefficients

$$n_{\vec{G}} = \frac{1}{\Omega} \int_{\Omega} n(\vec{x}) e^{-i\vec{G}\vec{x}} d\vec{x} \qquad (4.9)$$

where Ω is the volume of the elementary unit cell which is nothing else than $\Omega = \vec{a}_1 . (\vec{a}_2 \times \vec{a}_3)$. So far, we have been focusing on functions (like the electron density) that are periodic in *direct space*. In full analogy, any function $f(\vec{k})$ which will be periodic in *reciprocal space*, will have a Fourier Series expansion analogous to equation 4.8. In this case the series reads
$$f(\vec{k}) = \sum_{\vec{R}} f_{\vec{R}} e^{i\vec{k}\vec{R}} \qquad (4.10)$$

and the equivalent of equation 4.9 will be

$$f_{\vec{R}} = \frac{1}{\Omega_B} \int_{\Omega_B} f(\vec{k}) e^{-i\vec{k}\vec{R}} d\vec{k} \qquad (4.11)$$

where Ω_B is the volume of the elementary unit cell in reciprocal space called the **Brillouin Zone**. The volume can be worked out from the expression for reciprocal

basis vectors (equation 4.4) and give

$$\Omega_B = \frac{(2\pi)^d}{\Omega}, \text{ d is the dimensionality} \tag{4.12}$$

4.2.2 Translation Operators $T_{\vec{R}}$

Let us introduce the linear **translation operators**, $T_{\vec{R}}$, whose action on a function of space $f(\vec{x})$ is simply

$$[T_{\vec{R}}f](\vec{x}) = f(\vec{x} + \vec{R}) \text{ where } \vec{R} \text{ is a translation vector} \tag{4.13}$$

Straight from the definition, it is easy to establish that every operator $T_{\vec{R}}$ has an inverse operator, which is simply given by

$$T_{\vec{R}}^{-1} = T_{-\vec{R}} \tag{4.14}$$

We also have the following property

$$T_{\vec{R}_1}T_{\vec{R}_2} = T_{\vec{R}_1 + \vec{R}_2} \tag{4.15}$$

What's more, the translation operators commute with each other

$$T_{\vec{R}_1}T_{\vec{R}_2} = T_{\vec{R}_2}T_{\vec{R}_1} \tag{4.16}$$

In order to see this we just apply the definition

$$
\begin{aligned}
T_{\vec{R}_1}T_{\vec{R}_2}f(\vec{x}) = T_{\vec{R}_1}(f(\vec{x} + \vec{R}_2)) &= f(\vec{x} + \vec{R}_2 + \vec{R}_1) \\
&= f(\vec{x} + \vec{R}_1 + \vec{R}_2) = T_{\vec{R}_2}(f(\vec{x} + \vec{R}_1)) = T_{\vec{R}_2}T_{\vec{R}_1}f(\vec{x})
\end{aligned}
$$

We can also prove that the translation operators commute with the hamiltonian in equation 4.1. Indeed let us look first at the potential term

$$T_{\vec{R}}(V(\vec{x})f(\vec{x})) = V(\vec{x} + \vec{R})f(\vec{x} + \vec{R}) = V(\vec{x})f(\vec{x} + \vec{R}) = V(\vec{x})(T_{\vec{R}}f(\vec{x}))$$

in the above equation, we used the fact that the potential is periodic $V(\vec{x} + \vec{R}) = V(\vec{x})$. Now we move on to prove that translation operators commute with the following derivative operator $p_x = \frac{\partial}{\partial x}$. Indeed we see that

$$T_{\vec{R}}(p_x f(\vec{x})) = T_{\vec{R}}(\frac{\partial f}{\partial x}(\vec{x})) = \frac{\partial f}{\partial x}(\vec{x} + \vec{R})$$

On the other hand, the reverse reads

$$p_x(T_{\vec{R}}f(\vec{x})) = \frac{\partial}{\partial x}(f(\vec{x} + \vec{R})) = \frac{\partial}{\partial x}(f(h(\vec{x})))$$

where we defined h as the function $h : \vec{x} \mapsto \vec{x} + \vec{R}$. Now we conclude by using the chain rule formula

$$p_x(T_{\vec{R}} f(\vec{x})) = \sum_i \frac{\partial f}{\partial x_i}(h(\vec{x})).\frac{\partial h_i}{\partial x}$$

where x_i is a generic name for x, y and z. Of course here only $\frac{\partial h_1}{\partial x} = 1$ and all the other terms are zero such that the final expression is

$$p_x(T_{\vec{R}} f(\vec{x})) = \frac{\partial f}{\partial x}(h(\vec{x})) = \frac{\partial f}{\partial x}(\vec{x} + \vec{R})$$

The above expression is identical to the one we obtained previously. This proves the commutativity of p_x and $T_{\vec{R}}$. The same line of arguments can be used to prove the commutativity of p_y and p_z with $T_{\vec{R}}$. Since the Laplacian operator in equation 4.1 is simply

$$-\frac{1}{2}\vec{\nabla}^2 = -\frac{1}{2}(p_x^2 + p_y^2 + p_z^2)$$

it is a simple matter of verification to prove the commutativity of that latter operator with the translation operator $T_{\vec{R}}$. Gathering all the results we finally arrived at

$$T_{\vec{R}}(\hat{h}_{KS} f(\vec{x})) = T_{\vec{R}}\left(\left[-\frac{1}{2}\vec{\nabla}^2 + V(\vec{x})\right] f(\vec{x})\right) = [-\frac{1}{2}\vec{\nabla}^2 + V(\vec{x})](T_{\vec{R}} f(\vec{x}))$$

which is the desired result.

4.2.3 Bloch's Theorem

So far, we have proved that any translation operator $T_{\vec{R}}$ commutes with the Kohn-Sham hamiltonian and commutes with any other translation operator. We will now introduce two *fundamental Theorems* that will allow us to label all the distinct **invariant subspaces** of $\hat{h}_{KS}$. What does this mean ?

In short, it says that, once we have a way to label the distinct eigenvalues of all the translation operators (we will use $\vec{k}$ to label those), we will be able to use the same label to separate all the invariant subspaces of $\hat{h}_{KS}$ (an invariant subspace is a vector space such that if ψ belongs to that subspace, then so does $\hat{h}_{KS}\psi$). A very detailed yet accessible presentation of all the following results can be found in the excellent book by Altmann [6].

Felix Bloch (1905-1983)

Theorem 4.2.1. (Irreducible representations of the Space Group) *The set of all the eigenvalues ($\lambda_{\vec{k}}$) of the translation operators ($T_{\vec{R}}$) allows for the complete labeling of **all** the invariant subspaces of the Kohn-Sham hamiltonian $\hat{h}_{KS}$.*

Theorem 4.2.2. (Simultaneous Diagonalization of Commuting Operators) *We can choose the eigenfunctions of the Kohn-Sham hamiltonian ($\hat{h}_{KS}$) to be simultaneously eigenfunctions of the translation operators ($T_{\vec{R}}$)*

Thanks to theorems 4.2.1 and 4.2.2, we can tackle the problem of finding the eigenvalues and eigenfunctions of the Kohn-Sham hamiltonian by looking at the eigenvalues and eigenvectors of the (much easier to deal with) translation operators. For this we will introduce the **Born Von-Karman** periodic boundary conditions.

The Born Von-Karman periodic boundary conditions allow in practice to treat an infinite system (and most importantly crystals) by imposing that any wavefunction be periodic over a supercell. This condition is reminiscent of Part 1 and the replacement of the whole space $\mathbb{R}^3$ with a periodic "large box" $\mathbb{T}^3$. The Born Von-Karman conditions are exactly doing that. Let us set the stage and choose the supercell to be made of N_i elementary unit cells in direction $\vec{a}_i$. In total, the supercell will contain $N = \prod_i N_i$ elementary unit cells. The total volume of the supercell will be $V = N\Omega$, where we recall that Ω is the volume of the elementary unit cell. Let us now take any wavefunction $\psi(\vec{x})$. The Born Von-Karman periodic boundary conditions can then be written as

$$\forall i, \quad \psi(\vec{x} + N_i \vec{a}_i) = \psi(\vec{x}) \tag{4.17}$$

Let us suppose that $\psi(\vec{x})$ is now an eigenfunction of all the $T_{\vec{R}}$'s. We can find such an eigenfunction because of theorem 4.2.2. Then there must be some number $\lambda_{\vec{R}}$ such that

$$T_{\vec{R}}\psi(\vec{x}) = \lambda_{\vec{R}}\psi(\vec{x}), \quad \forall \vec{R} \tag{4.18}$$

Using equation 4.15 we have

$$T_{\vec{R}_1}T_{-\vec{R}_1}\psi(\vec{x}) = \lambda_{\vec{R}_1}\lambda_{-\vec{R}_1}\psi(\vec{x}) = \psi(\vec{x})$$

So we find that

$$\lambda_{-\vec{R}_1} = \frac{1}{\lambda_{\vec{R}_1}} \tag{4.19}$$

Another use of equation 4.15 leads to the following result

$$\lambda_{\vec{R}_1+\vec{R}_2} = \lambda_{\vec{R}_1}\lambda_{\vec{R}_2} \tag{4.20}$$

Introducing the following definitions

$$\lambda_1 = \lambda_{\vec{a}_1}, \quad \lambda_2 = \lambda_{\vec{a}_2}, \quad \lambda_3 = \lambda_{\vec{a}_3} \tag{4.21}$$

it is easy to combine those definitions with equations 4.19, 4.20 and 4.2 to prove that

$$\lambda_{\vec{R}} = \lambda_1^{r_1}\lambda_2^{r_2}\lambda_3^{r_3} \tag{4.22}$$

The numbers introduced in equation 4.21 are *a priori* complex numbers. We can choose to use an exponential representation for those numbers and write them in the following way

$$\lambda_j = e^{i2\pi\alpha_j}, \quad \text{with} \quad \alpha_j \in \mathbb{C}$$

Introducing that new representation into equation 4.22, we arrive at

$$\lambda_{\vec{R}} = e^{i2\pi(\alpha_1 r_1 + \alpha_2 r_2 + \alpha_3 r_3)}$$

The final trick consists in defining the following complex vector

$$\vec{k} = \alpha_1 \vec{b}_1 + \alpha_2 \vec{b}_2 + \alpha_3 \vec{b}_3$$

With that final definition we can rewrite equation 4.22 into a more suggestive one

$$\lambda_{\vec{R}} = e^{i\vec{k}\vec{R}}$$

<div align="right">(4.23)</div>

Of course all of the above steps are not answering the question of what the α's are. This is the point at which the Born Von-Karman periodic boundary conditions come in handy in order to address that issue.

Indeed equations 4.17, 4.18 and 4.23 lead us to

$$e^{i2\pi\alpha_j N_j} = \left(e^{i2\pi\alpha_j}\right)^{N_j} = 1, \quad \forall\, j$$

$z_j = e^{i2\pi\alpha_j}$ is then a N_j-th root of unity. It is a simple result of mathematics to find all the N_j-th root of unity. Those are simply

$$z_j = e^{i2\pi\frac{p}{N_j}} \quad \text{for} \quad p \in [0,...,N_j-1] \iff \alpha_j = \frac{p}{N_j} \quad \text{for} \quad p \in [0,...,N_j-1]$$

Going back to the expression for $\vec{k}$, we see that $\vec{k}$ is now a real vector. What's more, all the allowed values for $\vec{k}$ span the entire elementary unit cell of reciprocal space. That elementary cell is called the **Brillouin Zone**. Retrospectively, we observe that the introduction of the $\vec{k}$ parametrization in labeling the eigenvalues of the translation operators was the right thing to do.

We just made explicit all the eigenvalues of the translation operators. Let us now focus our attention on the analytical structure of their eigenfunctions. For this we use equation 4.23 along with equation 4.18

$$T_{\vec{R}}\psi(\vec{x}) = e^{i\vec{k}\vec{R}}\psi(\vec{x})$$

Since for different values of $\vec{k}$ we end up with a different eigenvalues, we see that the eigenfunction $\psi(\vec{x})$ needs to also depend on $\vec{k}$. We then introduce the notation $\psi_{\vec{k}}(\vec{x})$ to denote one such eigenfunction. Because of the above equation, the following function

$$u_{\vec{k}}(\vec{x}) = e^{-i\vec{k}\vec{x}}\psi_{\vec{k}}(\vec{x})$$

can be shown to be periodic. Indeed we have

$$u_{\vec{k}}(\vec{x}+\vec{R}) = e^{-i\vec{k}(\vec{x}+\vec{R})}\psi_{\vec{k}}(\vec{x}+\vec{R}) = e^{-i\vec{k}(\vec{x}+\vec{R})}e^{i\vec{k}\vec{R}}\psi_{\vec{k}}(\vec{x}) = u_{\vec{k}}(\vec{x})$$

So inverting the relationship between $\psi_{\vec{k}}(\vec{x})$ and $u_{\vec{k}}(\vec{x})$, we arrive at

$$\psi_{\vec{k}}(\vec{x}) = e^{i\vec{k}\vec{x}}u_{\vec{x}}(\vec{x}) \quad \text{with } u_{\vec{k}}(\vec{x}+\vec{R}) = u_{\vec{k}}(\vec{x})$$

<div align="right">(4.24)</div>

The above equation is often called **Bloch's Theorem**. And Theorem 4.2.2 proves that the eigenfunctions of the Kohn-Sham hamiltonian can be taken to satisfy Bloch's Theorem.

4.3 Eigenfunctions and Eigenvalues of the Kohn-Sham hamiltonian

We have been working quite hard to find a way to label the eigenvalues and eigenvectors of the translation operators. But all this work was not in vain. Indeed we now have huge insight into the analytical structure of the eigenfunctions inside every invariant subspace of the Kohn-Sham hamiltonian through Bloch's Theorem. But knowing about the invariant subspaces is not knowing the whole spectrum. We will now move on and introduce another label, the band index, that will help us find all the eigenvalues and eigenfunctions of $\hat{h}_{KS}$.

4.3.1 Existence of bands

To proceed we use 4.24 and inject the analytical expression inside the differential equation for $\psi_{\vec{k}}(\vec{x})$

$$\left[-\frac{1}{2}\vec{\nabla}^2 + V(\vec{x}) \right] \psi_{\vec{k}}(\vec{x}) = E(\vec{k})\psi_{\vec{k}}(\vec{x})$$

to find

$$\left[-\frac{1}{2}\left(\vec{\nabla} + i\vec{k}\right)^2 + V(\vec{x}) \right] u_{\vec{k}}(\vec{x}) = E(\vec{k})u_{\vec{k}}(\vec{x}) \tag{4.25}$$

Equation 4.25 is a linear partial differential equation for the eigenvalues and eigenvectors of the $\vec{k}$ dependent Kohn-Sham hamiltonian $\hat{h}_{KS}(\vec{k}) = -\frac{1}{2}\left(\vec{\nabla} + i\vec{k}\right)^2 + V(\vec{x})$. Since $u_{\vec{k}}(\vec{x})$ is periodic over the elementary unit cell, it is sufficient to solve that differential equation over the "home" unit cell (for which $\vec{R} = \vec{0}$). But that region of space is bounded, so a result of functional analysis [43] states the following

Theorem 4.3.1. (Spectral Theorem for $\hat{h}_{KS}(\vec{k})$) *The spectrum of all the Bloch Hamiltonians $\hat{h}_{KS}(\vec{k})$ consist in discrete eigenvalues $E_n(\vec{k})$, indexed by an integer number n, with a finite multiplicity (only a finite number of eigenvectors for each $E_n(\vec{k})$)*

Theorem 4.3.1 proves the existence of "energy bands" for any periodic system. Since we have a new index for the eigenvalues, we must also introduce a new index for the eigenfunctions. We will finally use the following notation for the eigenfunctions and eigenvalues of the Kohn-Sham hamiltonian

$$\text{Eigenfunctions and Eigenvalues of } \hat{h}_{KS} : \tag{4.26}$$

$$\psi_{n\vec{k}}(\vec{x}) = e^{i\vec{k}\vec{x}} u_{n\vec{k}}(\vec{x}) \quad \text{and} \quad E_n(\vec{k})$$

In the next section, we will explicit a most useful property of the energy bands.

4.3.2 A Useful result concerning the Energy Bands

Let us take a closer look at equation 4.25. If we take the complex conjugate of that equation we end up with

$$\left[-\frac{1}{2}\left(\vec{\nabla} - i\vec{k}\right)^2 + V(\vec{x}) \right] v_{\vec{k}}(\vec{x}) = E(\vec{k}) v_{\vec{k}}(\vec{x})$$

where $v_{\vec{k}}(\vec{x}) = u_{\vec{k}}^*(\vec{x})$. But the above equation is nothing else than the equation for the wavevector $-\vec{k}$. So we see that once we know the eigenfunctions and eigenvalues for $\vec{k}$, we also know them for $-\vec{k}$. We can then decide to take $u_{-\vec{k}}(\vec{x}) = u_{\vec{k}}^*(\vec{x})$ and so we end up with the following property for $E_n(\vec{k})$

$$E_n(-\vec{k}) = E_n(\vec{k}) \tag{4.27}$$

4.3.3 Electronic Structure in the Eigenfunction basis

We would like to finish this review by collecting a list of very important and useful formulas concerning the set of eigenfunctions and the Kohn-Sham hamiltonian. We will make an effort to differentiate the formulas in the discrete case (i.e. when considering a finite supercell) from the continuous case. That latter case does not actually correspond to a "mathematically" infinite solid (which is something that does not exist in nature!), but rather a solid of *macroscopic* volume. In that case the sums over the $\vec{k}$ vectors can be turned into Riemann *integrals*. Indeed, using a discrete approximation to a Riemann integral we have

$$\sum_{\vec{k} \in BZ} f(\vec{k}) \Delta\vec{k} \approx \int_{BZ} d\vec{k} f(\vec{k})$$

where BZ stands for the Brillouin Zone and $f(\vec{k})$ is an arbitrary function. The elementary volume in $\vec{k}$-space, $\Delta\vec{k}$, corresponds to the volume taken up by *one $\vec{k}$* vector. We saw earlier that the Born Von-Karman periodic boundary conditions were introducing a *uniform* grid a $\vec{k}$ points over the Brillouin Zone, with exactly N points in it (N being the number of elementary unit cells in the supercell). Since the volume of the Brillouin Zone is given by equation 4.12, we conclude that

$$\Delta\vec{k} = \frac{\Omega_B}{N} = \frac{(2\pi)^d}{V}$$

where V is the total volume of the supercell. All in all, we observe that, whenever we want to go for the continuous limit, we will have to use the following substitution

$$\sum_{\vec{k} \in BZ} = \frac{V}{(2\pi)^d} \int_{BZ} d\vec{k} \tag{4.28}$$

The first formula is concerned with the **completeness** relation

In the discrete case $\qquad\qquad$ (4.29)

$$\sum_n \sum_{\vec{k} \in BZ} |\psi_{n\vec{k}}\rangle\langle\psi_{n\vec{k}}| = \mathbb{1}$$

In the continuous case

$$\sum_n \frac{V}{(2\pi)^d} \int_{BZ} d\vec{k} |\psi_{n\vec{k}}\rangle\langle\psi_{n\vec{k}}| = \mathbb{1}$$

Here we use the Dirac notation for "bra"s and "ket"s [12] which is a most useful notation in Quantum Mechanics. $|\psi_{n\vec{k}}\rangle\langle\psi_{n\vec{k}}|$ is geometrically interpreted as being the projector operator onto the subspace corresponding to the eigenfunction $\psi_{n\vec{k}}(\vec{x})$. $\mathbb{1}$ is simply the identity operator.

The second formula has to do with the **orthonormality** of the eigenfunction basis. Indeed we have

$$\langle\psi_{n\vec{k}}|\psi_{n'\vec{k}'}\rangle = \int_V \psi^*_{n\vec{k}}(\vec{x})\psi_{n'\vec{k}'}(\vec{x})d\vec{x} = \sum_{\vec{R}} \int_\Omega \psi^*_{n\vec{k}}(\vec{x}+\vec{R})\psi_{n'\vec{k}'}(\vec{x}+\vec{R})d\vec{x}$$

where $\vec{R}$ runs over all the elementary unit cells in the supercell and Ω corresponds to the "home" unit cell (i.e. $\vec{R} = \vec{0}$). Now we use Bloch's Theorem (equation 4.24) to simplify that integral

$$\langle\psi_{n\vec{k}}|\psi_{n'\vec{k}'}\rangle = \sum_{\vec{R}} e^{i(\vec{k}'-\vec{k})\vec{R}} \int_\Omega \psi^*_{n\vec{k}}(\vec{x})\psi_{n'\vec{k}'}(\vec{x})d\vec{x}$$

we then realize that the integral on the right-hand side is independant of $\vec{R}$ and we move it to the left

$$\langle\psi_{n\vec{k}}|\psi_{n'\vec{k}'}\rangle = \left(\int_\Omega \psi^*_{n\vec{k}}(\vec{x})\psi_{n'\vec{k}'}(\vec{x})d\vec{x}\right)\sum_{\vec{R}} e^{i(\vec{k}'-\vec{k})\vec{R}}$$

now we compute the sum on the far right by remembering that $\vec{k}' - \vec{k}$ has to be a vector inside the Brillouin Zone. First the allowed $\vec{R}$ vectors are going to be given by $\vec{R} = \sum_j r_j \vec{a}_j$ with r_j in $[0,...,N_j - 1]$. What's more, if we call $\vec{q} = \vec{k}' - \vec{k}$, then given the allowed values for $\vec{q}$, that latter vector can be written as $\vec{q} = \sum_l \frac{q_l}{N_l}\vec{b}_l$. Those results lead to

$$\sum_{\vec{R}} e^{i(\vec{k}'-\vec{k})\vec{R}} = \prod_l \sum_{j=0}^{N_l-1} \left(e^{i2\pi\frac{q_l}{N_l}}\right)^j$$

The above sums are simple geometrical sums. We can easily compute them and arrive at

$$\sum_{\vec{R}} e^{i(\vec{k}'-\vec{k})\vec{R}} = \prod_l \frac{e^{i2\pi q_l} - 1}{e^{i2\pi\frac{q_l}{N_l}} - 1}$$

q_l being an integer number we see that, unless q_l is zero for all l, the above will

vanish. The case $\vec{q} = \vec{0}$, which amounts to saying that $\vec{k} = \vec{k}'$, leads to $\sum_{\vec{R}} e^{i(\vec{k}'-\vec{k})\vec{R}} = N$ (where N is the number of elementary unit cells in the supercell). All in all, the orthonormality condition, using the convention that the eigenfunctions are **normalized over the supercell**, reads

$$\langle \psi_{n\vec{k}} | \psi_{n'\vec{k}'} \rangle = \int_V \psi^*_{n\vec{k}}(\vec{x}) \psi_{n'\vec{k}'}(\vec{x}) d\vec{x} = \delta_{nn'} \delta_{\vec{k}\vec{k}'} \tag{4.30}$$

The case $n \neq n'$ when $\vec{k} = \vec{k}'$ corresponds to two eigenfunctions from the same invariant subspace (same $\vec{k}$) with different band indices. In that latter case, one can easily prove that the eigenfunctions are orthogonal.

The final formula gives a representation of the Kohn-Sham hamiltonian in terms of the projectors defined in equation 4.29

In the discrete case $\hspace{4cm}$ (4.31)

$$\hat{h}_{KS} = \sum_n \sum_{\vec{k} \in BZ} E_n(\vec{k}) | \psi_{n\vec{k}} \rangle \langle \psi_{n\vec{k}} |$$

In the continuous case

$$\hat{h}_{KS} = \sum_n \frac{V}{(2\pi)^d} \int_{BZ} d\vec{k} E_n(\vec{k}) | \psi_{n\vec{k}} \rangle \langle \psi_{n\vec{k}} |$$

<div style="text-align: right; font-size: 4em;">**5**</div>

Wannier Functions

5.1 Motivation for introducing Wannier Functions

In 1937, Gregory Wannier introduced a set of orthonormal orbitals that spans the same Hilbert space as the eigenfunctions of the Kohn-Sham hamiltonian. The motivation for such an alternative set was to be able to describe the electronic structure of a solid in terms of localized orthonormal orbitals much like what happens in Quantum Chemistry when one describes the molecular states in terms of atomic orbitals.

In order to convince oneself that such a set exists, we follow Wannier [102] and consider an extremely simple system, namely a free electron, in a periodic solid. Using Bloch's Theorem we see that the eigenfunctions $\psi_{n\vec{k}}(\vec{x})$ can be written as

$$\psi_{n\vec{k}}(\vec{x}) = a_n e^{i\vec{k}\vec{x}}, \quad a_n \in \mathbb{C}$$

Gregory Wannier (1911-1983)

Let us focus on a single electronic band n. The intuition for constructing a set of "localized" functions draws on our knowledge of Fourier analysis. Indeed we "know" that when a function is really delocalized in space (like a Bloch function), then its Fourier transform generally exhibits some strong degree of localization. So we naturally define a set of functions, which depend on $\vec{R}$, as

$$w_{n\vec{R}} = \frac{1}{\sqrt{N}} \sum_{\vec{k}\in BZ} \psi_{n\vec{k}}(\vec{x}) e^{-i\vec{k}\vec{R}}$$

We then insert the analytical expression for $\psi_{n\vec{k}}(\vec{x})$ and after some easy calculations we find

$$w_{n\vec{R}}(\vec{x}) = \frac{a_n}{\sqrt{N}} \prod_{j=1}^{d} \left[e^{i(1-\frac{1}{N})\vec{b}_j(\vec{x}-\vec{R})} \frac{\sin\left(\frac{1}{2}\vec{b}_j(\vec{x}-\vec{R})\right)}{\sin\left(\frac{1}{2N}\vec{b}_j(\vec{x}-\vec{R})\right)} \right] \quad \boxed{5.1}$$

Despite its not-so-appealing look, equation 5.1 demonstrates the "localized" character of the Wannier functions, as one can see on figure 5-1.

Though the idea of using a Fourier transform to generate a set of localized Wannier functions from the set of delocalized Bloch functions is straightforward, there remains some arbitrariness in the Bloch function phases which leads to a non-uniqueness
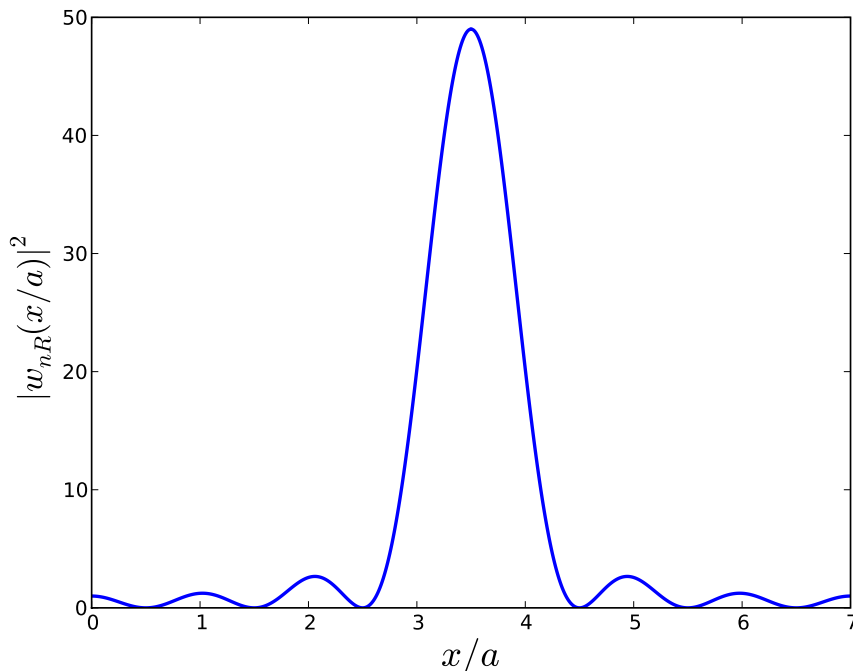
Figure 5-1: Illustration of the "localized" character of Wannier functions. The above plot shows the behavior of the square modulus of a normalized 1D version of a Wannier Function given by equation 5.1. The "crystal" consists in 7 elementary cells in the Born Von-Karman supercell, and the Wannier function is centered around the middle unit cell.

in the Wannier functions. Indeed if one performs the following transformation

$$\psi_{n\vec{k}}(\vec{x}) \longmapsto e^{i\alpha_{n\vec{k}}}\psi_{n\vec{k}}(\vec{x}), \quad \alpha_{n\vec{k}} \in \mathbb{R}$$

onto the set of Bloch functions, then the Kohn-Sham hamiltonian is left unchanged (see 4.31 where $|\psi_{n\vec{k}}\rangle$ is transformed into $e^{i\alpha_{n\vec{k}}}|\psi_{n\vec{k}}\rangle$ and $\langle\psi_{n\vec{k}}|$ into $\langle\psi_{n\vec{k}}|e^{-i\alpha_{n\vec{k}}}\rangle$). As a consequence, the physics of the system is equally well described by this new set of Bloch functions. However, if one multiplies each Bloch function by a different phase factor, the set of Wannier functions is obviously changed. This indeterminacy has been a major draw back against a wide adoption of Wannier functions in solid state physics until very recently [65]. We will see in the next section that this non-uniqueness problem can be circumvented by actually using the phase indeterminacy of the Bloch functions to our advantage.

5.2 Maximally Localized Wannier Functions

Because of Bloch functions' phase indeterminacy (and more generally that the energy functional is mathematically invariant under a general unitary transformation of the occupied Bloch functions[1]), let us introduce the following most general transformation

For the discrete case $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (5.2)

$$|w_{n\vec{R}}\rangle = \frac{1}{\sqrt{N}} \sum_{\vec{k} \in BZ} \left[\sum_m U_{mn}^{\vec{k}} |\psi_{m\vec{k}}\rangle \right] e^{-i\vec{k}\vec{R}}$$

For the continuous case

$$|w_{n\vec{R}}\rangle = \frac{1}{\sqrt{N}} \frac{V}{(2\pi)^d} \int_{BZ} d\vec{k} \left[\sum_m U_{mn}^{\vec{k}} |\psi_{m\vec{k}}\rangle \right] e^{-i\vec{k}\vec{R}}$$

If the Bloch functions are normalized over the whole supercell (volume V), then one can easily prove from the orthonormality of the Bloch basis (equation 4.30) and the unitarity of the $\vec{k}$-dependent matrices $U_{mn}^{\vec{k}}$ that the Wannier functions so defined are orthonormal

$$\langle w_{n\vec{R}} | w_{n'\vec{R}'} \rangle = \int_V d\vec{x}\, w_{n\vec{R}}^*(\vec{x})\, w_{n'\vec{R}'}(\vec{x}) = \delta_{nn'}\delta_{\vec{R}\vec{R}'} \qquad (5.3)$$

The inverse relation is given by

$$|\psi_{n\vec{k}}\rangle = \frac{1}{\sqrt{N}} \sum_{\vec{R}} \left[\sum_m \left(U_{mn}^{\vec{k}} \right)^{-1} |w_{m\vec{R}}\rangle \right] e^{i\vec{k}\vec{R}} \qquad (5.4)$$

where $\vec{R}$ runs over all the elementary unit cells inside the supercell. Equation 5.2 can be understood as a *two-step process*. In the first step, we define a new set of Bloch functions through

$$|\phi_{n\vec{k}}\rangle = \sum_m U_{mn}^{\vec{k}} |\psi_{m\vec{k}}\rangle \qquad (5.5)$$

that we Fourier transform in a second step

$$|w_{n\vec{R}}\rangle = \frac{1}{\sqrt{N}} \sum_{\vec{k} \in BZ} |\phi_{n\vec{k}}\rangle e^{-i\vec{k}\vec{R}}$$

The $\vec{k}$-dependent matrices can of course be chosen completely arbitrarily, so that for each given set of matrices we obtain a different set of Wannier functions. In order to define a unique set of Wannier functions, we will follow Marzari and Vanderbilt

---

[1]This statement is obviously not true if the functional is orbital-dependent. However, in that latter case, since we are only interested in the eigenvalues of the Kohn-Sham hamiltonian, we can go from the Bloch basis to any other basis. The reason comes from basic linear algebra. Indeed if A is a matrix, the eigenvalues can formally be computed by solving the secular equation $\det(A - \lambda\mathbb{I}) = 0$. Now if we perform a change of basis, then A become $B = U^{-1}AU$ where U is any invertible matrix. From there we have $\det(B - \lambda\mathbb{I}) = \det(U^{-1}AU - \lambda U^{-1}U)$, which simplifies to $\det(B - \lambda\mathbb{I}) = \det(U^{-1})\det(A - \lambda\mathbb{I})\det(U) = \det(A - \lambda\mathbb{I})$. This proves that the eigenvalues of B are the same as the ones of A

[65] and find the special set of $\vec{k}$-dependent matrices for which the Wannier functions are maximally localized. By maximal locality we mean that the set of Wannier functions will minimize the sum of their mean square spread

$$\Omega = \sum_n \left[\langle \vec{r}^2 \rangle_n - \langle \vec{r} \rangle_n^2 \right] = \sum_n \left[\langle w_{n\vec{0}} | r^2 | w_{n\vec{0}} \rangle - \langle w_{n\vec{0}} | \vec{r} | w_{n\vec{0}} \rangle^2 \right] \qquad \boxed{5.6}$$

In the above, we defined Ω as the *spread functional* and one should not be confused with Ω the elementary unit cell volume. For a given transformation (i.e. given $U_{mn}^{\vec{k}}$ matrices), the required matrix elements of the position operator in the Wannier function basis are given in terms of the periodic part of the Bloch functions in equation 5.5 [9][2]

$$\langle \vec{r} \rangle_n = i \sum_{\vec{k} \in BZ} \langle u_{n\vec{k}} | \vec{\nabla}_{\vec{k}} | u_{n\vec{k}} \rangle = i \frac{V}{(2\pi)^d} \int_{BZ} d\vec{k} \langle u_{n\vec{k}} | \vec{\nabla}_{\vec{k}} | u_{n\vec{k}} \rangle \qquad \boxed{5.7}$$

$$\langle \vec{r}^2 \rangle_n = - \sum_{\vec{k} \in BZ} \langle u_{n\vec{k}} | \vec{\nabla}_{\vec{k}}^2 | u_{n\vec{k}} \rangle = - \frac{V}{(2\pi)^d} \int_{BZ} d\vec{k} \langle u_{n\vec{k}} | \vec{\nabla}_{\vec{k}}^2 | u_{n\vec{k}} \rangle$$

The Wannier functions that minimize the spread functional of equation 5.6 are called **Maximally Localized Wannier functions** (MLWF). The algorithm to compute the MLWFs generally proceeds in two steps. The first step, called the **disentanglement procedure**, takes a certain amount of initial Bloch functions ($M \times N$ if M is the number of bands computed at each $\vec{k}$ vector in the Brillouin Zone and N is the number of $\vec{k}$ such vectors) and extracts an optimally smooth sub-manifold of $M' \times N$ Bloch functions (where $M' \leq M$). The disentanglement procedure has been designed to be able to deal with metallic systems for which no energy gap separates the occupied manifold of Bloch functions from the unoccupied manifold. The second step, generally called the **wannierisation procedure**, starts from the optimal sub-manifold of Bloch states and determines the set of $\vec{k}$-dependent matrices that minimizes the spread functional Ω. The procedure is iterative in nature and uses both steepest-descent and conjugate-gradient techniques [4]. A detailed description of both procedures (disentanglement and wannierisation) and more can be found in reference [70]. We end up this short introduction to MLWFs by stating some of the most important properties of those functions that triggered their wide adoption in the solid state physics community

- The set of MLWF is unique

---

[2]We mention here in passing that the expressions for the matrix elements of the position operator (equation 5.7) are very non-trivial. Indeed the position operator $\hat{x}$ (or $\hat{y}$ or $\hat{z}$) seen simply as a multiplicative operator (i.e. an operator that takes a wavefunction $\psi(\vec{x})$ and returns $x\psi(\vec{x})$) *cannot* be a proper operator because it is *incompatible* with periodic boundary conditions. Defining proper expectation values for the position operator in periodic boundary conditions is very difficult. The interested reader is invited to look at R. Resta's paper on the subject [83]. It is quite captivating to discover that the recent developments in our understanding of the phenomenon of polarization in periodic systems is very much related to those issues concerning the position operator. Despite the beautiful and mind-boggling mathematical and physical aspects of polarization, we will not venture into that field in this thesis.

- MLWFs are orthonormal orbitals

- In insulators MLWFs can be shown to be exponentially localized [10]

- If the spread functional Ω has a unique minimum, the MLWFs are real at that minimum [10]

- MLWFs can be interpreted in terms of bonding and anti-bonding local orbitals

- MLWFs are central in the modern theory of polarization [99]

- MLWFs are central to building model Hamiltonians (more on this later) [61]

- MLWFs are the fundamental basis functions in many $\mathcal{O}(N)$ computer codes

5.3 Electronic Structure in a MLWF basis

It is now time to look at how to represent the Kohn-Sham hamiltonian in the MLWF basis. For this, we simply use equation 4.31 and inject the expression for $|\psi_{n\vec{k}}\rangle$ of equation 5.4 to find

$$\hat{h}_{KS} = \sum_{\vec{R},\vec{R}'} \sum_{m,m'} H_{mm'}(\vec{R} - \vec{R}')|w_{m\vec{R}}\rangle\langle w_{m'\vec{R}'}| \tag{5.8}$$

with

In the discrete case $\qquad\qquad$ (5.9)

$$H_{mm'}(\vec{R}) = \frac{1}{N} \sum_{\vec{k}\in BZ} H_{mm'}(\vec{k})e^{i\vec{k}\vec{R}}$$

In the continuous case

$$H_{mm'}(\vec{R}) = \frac{\Omega}{(2\pi)^d} \int_{BZ} H_{mm'}(\vec{k})e^{i\vec{k}\vec{R}}$$

and

$$H_{mm'}(\vec{k}) = \sum_{n} \left(U_{mn}^{\vec{k}}\right)^{-1} E_n(\vec{k})U_{nm'}^{\vec{k}} \tag{5.10}$$

As a remainder, N is the total number of elementary unit cells inside the Born Von-Karman supercell (which equals the number of $\vec{k}$ points in the Brillouin Zone) and Ω is the volume of the elementary unit cell. Equation 5.10 is nothing else than a change of basis from the basis of the eigenfunctions of $\hat{h}_{KS}$ to the basis of the quasi-Bloch states of equation 5.5 applied to the following diagonal matrix

$$\begin{pmatrix} E_1(\vec{k}) & 0 & \ddots & \\ 0 & E_2(\vec{k}) & 0 & \ddots \\ \ddots & 0 & E_3(\vec{k}) & 0 \\ & \ddots & 0 & \ddots \end{pmatrix}$$

5.3.1 Band Structure from the Real-Space Matrices

Equations 5.9 and 5.10 allows us to devise a way of computing the band structure of a periodic solid whenever we know the real-space representation of the Kohn-Sham hamiltonian, i.e. we know all the $H(\vec{R})$ matrices. Indeed, we just have to do two things

1. For a given $\vec{k}$ in the Brillouin Zone, we compute the $\vec{k}$-dependent matrix $H(\vec{k})$

2. then we diagonalize that matrix to obtain the band energies for that $\vec{k}$

Using equation 5.9 we can easily prove the following inversion formula for $H(\vec{k})$

$$H_{mm'}(\vec{k}) = \sum_{\vec{R}} H_{mm'}(\vec{R})e^{-i\vec{k}\vec{R}} \qquad \boxed{5.11}$$

So we see that all we have to do in order to compute the band structure is to sum up all the $H(\vec{R})$ matrices, modulating the sum by complex exponential factors $e^{-i\vec{k}\vec{R}}$, and diagonalize the so-obtained matrix.

5.3.2 A Practical Algorithm

In the last section, we realized that computing the band structure from the real-space representation of the Kohn-Sham hamiltonian is pretty straightforward. But some subtleties are implied when one wants to implement the algorithm in practice.

The first subtlety has to do with the fact that one needs to be able to properly extract the block $H(\vec{R})$ corresponding to the proper unit cell $\vec{R}$. This requires an algorithm capable of telling which are the "first nearest neighbor" cells, which are the "second nearest neighbor" cells and so forth.

The second subtlety has to do with the order of Wannier functions in each elementary unit cell. In theory, the Wannier functions belonging to any unit cell can be ordered in any random fashion. But here we choose a consistent order among unit cells. We will see, in the transport part of this thesis, that a consistent order is of paramount importance in practice and leads to a greater transferability of hamiltonian matrices.

The very last practical subtlety has to do with the phase uncertainty of Wannier functions. Just like Bloch functions, Wannier functions can be multiplied by an arbitrary phase. As a consequence, Wannier functions tend to arise with no predefined relative phases. When the functions are real, there is no predefined relatives signs. In practice though, we impose a given sign pattern.[3]

Using all of the above, we implemented a general-purpose algorithm into *Wannier90*[4], a computer code implementing the procedure of Marzari et *al.* for computing

---

[3]The detailed description of how one can impose a relative sign pattern and how one can sort Wannier functions will be explained in the transport part of the thesis.

[4]http://www.wannier.org

MLWFs. The code starts from a set of Bloch functions computed onto a Born Von-Karman supercell and proceeds by computing the MLWFs of that supercell. From the MLWF centers, our algorithm goes through the following steps

---

Real-Space Matrices Extraction Algorithm

1. Use the first Wannier function as seed and build an elementary unit cell around that function. Then find all the Wannier functions belonging to that unit cell.

2. Compute the maximum distance L between Wannier function centers from the size of the supercell. That distance is half the *smallest* distance one can go in any of the supercell's directions before the periodicity of the supercell gets us back where we started.

3. Compute the maximum number of nearest neighbors that we can extract from the supercell. For this use L and the elementary cell basis vectors ($\vec{a}_1$ and/or $\vec{a}_2$ and/or $\vec{a}_3$).

4. Extract the "home" elementary cell matrix $H(\vec{0})$.

5. Use the Wannier function basis in the "home" cell as a reference for both the order and signs of the Wannier functions so as to extract consistent nearest neighbor matrices $H(\vec{R})$.

6. Store the nearest neighbor matrices in file.

Required from the user are

- The elementary unit cell basis vectors $\vec{a}_1$ and/or $\vec{a}_2$ and/or $\vec{a}_3$

- The number of Wannier functions per elementary unit cell

- Three logical values telling the code which of the basis vector to use internally (those logical values basically sets the dimensionality of the underlying periodic system, hence allowing to deal with strictly 1D, 2D or fully 3D systems.)

From the elementary cell matrices $H(\vec{R})$ we can directly use equation 5.11 and compute $H(\vec{k})$ for a given set of $\vec{k}$ vectors inside the Brillouin Zone and diagonalize each matrix to find the band energies. This technique is an alternative to the classic method for computing band structures from the Bloch eigenfunctions, as implemented in most *ab-initio* computer codes like Quantum-Espresso[5], VASP[6] or SIESTA[7]. In terms of efficiency, this technique has the obvious drawback of requiring a calculation for a supercell instead of a single unit cell which, given the cubic scaling of most *ab-initio* codes with the number of electrons, precludes its use on large 3D systems.

---

[5] http://www.quantum-espresso.org/
[6] http://cms.mpi.univie.ac.at/vasp/
[7] http://www.icmab.es/siesta/

However, if we are interested in computing band structures using *Self-Interaction functionals*, we will later see that this technique is the only one available in practice.

We illustrate the agreement between the *supercell technique* and the *Bloch function technique* by computing the band structure of graphene using the Local Density Approximation (LDA) with a kinetic energy cutoff of 45 Rydbergs and a charge density cutoff of 180 Rydbergs. We used a Martins-Trouiller Norm-Conserving pseudopotential obtained from an all-electron calculation with the Perdew-Wang 1992 parametrization of the LDA functional. The bands obtained for the Bloch function technique were converged using a 20×20 $\vec{k}$ mesh in the Brillouin Zone. By contrast, we only used a 6×6 supercell of graphene to compute the bands in the supercell technique case. Figure 5-2 basically validates our supercell technique when computing band structures of periodic systems.
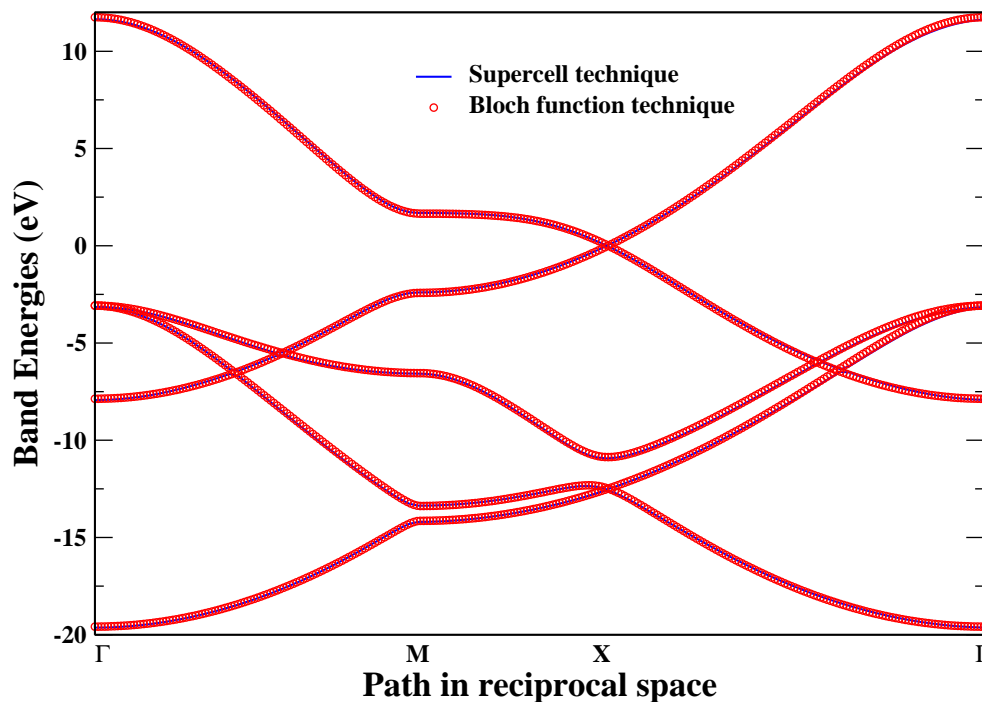
Figure 5-2: Graphene bands as computed from the supercell technique (blue solid line) and the Bloch function technique (red circles). We basically see an excellent agreement all throughout the Brillouin Zone.

Part III

Non-Koopman correction to Density Functional Theory

6

Fundamental Issues of Approximate Density Functionals

In this introduction, we would like to outline some of the most urging issues associated with common local and semi-local Approximate Density Functionals (ADF). In doing so we will present both qualitative failures in predicting physical properties of quantum systems and what the community considers to be the reasons for those failures. In the next chapter, we will propose a methodology for addressing some of these issues.

6.1 Local and Semi-Local Density Functionals

Before to say anything about the failures of commonly used Density Functionals, it is necessary to first show what those Functionals are in some details. Roughly speaking there are two types of "simple" functionals used in practice[1]. The simplest functional, called the Local (Spin) Density Approximation - L(S)DA, was already suggested by Kohn and Sham in their fundamental paper [52]. The LDA, as we will see, is actually uniquely defined. This is not the case, however, for the second class of functionals, namely the Generalized Gradient Approximations - GGA.

6.1.1 The Local Density Approximation

The main idea underlying the construction of the LDA functional is very simple. In a system, the electron density, $n(\vec{x})$, is a well-defined positive function of space. In an infinitesimal volume of size $d\vec{x}$, the number of electrons in that volume is $n(\vec{x})d\vec{x}$. The local density approximation proceeds by computing the total exchange-correlation energy of a system by summing up all the infinitesimal volumes containing $n(\vec{x})d\vec{x}$ electrons for which the contribution to the energy is $n(\vec{x})\epsilon_{xc}^{\mathrm{LDA}}(n(\vec{x}))d\vec{x}$. $\epsilon_{xc}^{\mathrm{LDA}}(n(\vec{x}))$ is the exchange-correlation energy per electron of an *interacting homogeneous electron*

---

[1]In this chapter we will be focusing mostly on Local and Semi-local functionals. In particular we will not say anything concerning so-called "hybrid" functionals. The reason is that although some of the deficiencies of local and semi-local functionals are remedied by hybrids, those also fail on many of the important issues mentioned in this chapter. Their mathematical expressions are on the other hand more complex than local and semi-local functionals. As a consequence, we will simplify the discussion by focusing solely on those latter functionals.

gas of constant density $n_0 = n(\vec{x})$ equal to that of the system at point $\vec{x}$. By using that approximation, the total Exchange and Correlation energy $E_{XC}[n]$ can be written as

$$E_{XC}^{\text{LDA}}[n] = \int_{\mathbb{T}^3} n(\vec{x}) \epsilon_{xc}^{\text{LDA}}(n(\vec{x})) d\vec{x} \tag{6.1}$$

In appendix A, we give the analytical expression for $\epsilon_{xc}^{\text{LDA}}(n)$ as parametrized by Perdew and Wang in 1992 [80]. By construction, the LDA is exact for a system of constant electron density, namely the **homogeneous electron gas**. As a consequence, the exchange-correlation energy per electron, $\epsilon_{xc}^{\text{LDA}}(n)$, corresponds to an actual physical (although probably not experimentally realizable) and *unique* system. This functional is then also unique.

6.1.2 Generalized Gradient Approximations

In order to improve upon the LDA, we first present a general formal expression for the exchange-correlation energy and introduce the **exchange-correlation hole** $h_{xc}(\vec{x}, \vec{x}')$. Indeed, one can justify [40] the mathematical expression in equation 6.1

$$E_{XC}[n] = \int_{\mathbb{T}^3} n(\vec{x}) \epsilon_{xc}(n(\vec{x})) d\vec{x} \tag{6.2}$$

where the exact exchange-correlation energy per electron is

$$\epsilon_{xc}(\vec{x}) = \frac{1}{2} \int_{\mathbb{T}^3} \frac{h_{xc}(\vec{x}, \vec{x}')}{|\vec{x} - \vec{x}'|} d\vec{x}' \tag{6.3}$$

Simply said, the exchange-correlation hole $h_{xc}(\vec{x}, \vec{x}')$ is the probability to find a "hole" (lack of an electron) at $\vec{x}'$ knowing that an electron is at $\vec{x}$. The hole can be split into an *exchange-only* part, $h_x(\vec{x}, \vec{x}')$, and a *correlation-only* part, $h_c(\vec{x}, \vec{x}')$. Then we have the further following exact properties

$$\int_{\mathbb{T}^3} h_x(\vec{x}, \vec{x}') d\vec{x}' = -1, \quad \forall \, \vec{x} \tag{6.4}$$

and

$$\int_{\mathbb{T}^3} h_c(\vec{x}, \vec{x}') d\vec{x}' = 0, \quad \forall \, \vec{x} \tag{6.5}$$

In building Generalized Gradient Approximations, the idea is to expand the exchange-correlation hole in the limit of slowly varying densities. So we would write

$$\epsilon_{xc}^{\text{exact}}(n) \approx \epsilon_{xc}^{\text{LDA}}(n) + f_1(n)|\vec{\nabla} n| + \dots \tag{6.6}$$

If one uses such an expansion, the quality of the exchange-correlation functional is actually worsened compared to LDA. The fundamental reason behind this is the fact that the sum rules 6.4 and 6.5 for the exchange and correlation holes are not satisfied anymore, along with other known limits [76]. As a result, one needs to introduce a

"generalized" notion of gradient expansion. One then writes

$$E_{XC}^{\mathrm{GGA}}[n] = \int_{\mathbb{T}^3} n(\vec{x}) f(n(\vec{x}), \vec{\nabla} n(\vec{x})) d\vec{x} \qquad (6.7)$$

The goal is then to recover the exact slowly varying expansion in 6.6 while still satisfying as much sum rules and known limits of the exchange-correlation hole as possible. The detail of the function $f(n(\vec{x}), \vec{\nabla} n(\vec{x}))$ is given in appendix A for a popular GGA functional by Perdew, Burke and Ernzerhof.

6.2 Important Results in Exact DFT

In this section, we would like to review and prove some very fundamental results of exact DFT. After doing that, we will be in a unique position to assess the accuracy of common local and semi-local functionals by directly comparing their predictions to expected results or trends from exact DFT.

6.2.1 The One Electron Limit

A first obvious but nevertheless important limit, is the one of a system with a single electron. Indeed in this latter case, Hartree, exchange and correlations are by definition absent. This leads us to conclude that the Hartree, exchange and correlation terms for *both* the energy functional and the Kohn-sham equation should cancel each other exactly. That this should be so is simple to realize, since in the case of a single electron, the Hohenberg-Kohn functional should reduce to the Rayleigh-Ritz functional and the Kohn-Sham equation should reduce to the Schrödinger equation. An example of mean-field theory that get this limit right is provided by Hartree-Fock theory. Indeed in the limit of a single electron, the Slater determinant ansatz of Hartree-Fock theory for the wavefunction of the system is exact. As a consequence, the Hartree-Fock functional and the Hartree-Fock equation reduce respectively to the Rayleigh-Ritz functional and Schrödinger equation. Let us summarize those findings in equations

$$\text{for one electron systems with density } \; n(\vec{x}) \qquad (6.8)$$

$$\begin{aligned} E_H[n(\vec{x})] \;&+\; E_{XC}[n(\vec{x})] = 0 \\ V_H[n](\vec{x}) \;&+\; V_{XC}[n](\vec{x}) = 0 \end{aligned}$$

6.2.2 Linearity of Ground State Energy with Electron Number

In 1982, Perdew, Parr, Levy and Balduz [79] proved a most important property of the ground state energy, seen as a function of the total electron number N. In short this theorem states that the exact ground state energy of a system with a fractional number of electrons N, between integer values $M-1$ and M, is nothing else than a **linear interpolation** of the exact energies $E_0(M-1)$ and $E_0(M)$. Mathematically, the result reads

Theorem 6.2.1. (Linearity of Total Energy with N) *If N is a fractional number of electrons between integer numbers $M-1$ and M, i.e. $N = (1-\omega)(M-1) + \omega M$ (for $\omega \in]0,1[$), then the exact total energy of the system is given by $E_0(N) = (1-\omega)E_0(M-1) + \omega E_0(M)$*

We will prove the above theorem by using the full formalism of part I. We start from the very definition of what the ground state energy of a quantum system with N electrons is. According to equation 2.14, the ground state energy is obtained when one searches over all the admissible density operators $\hat{\Gamma} = \sum_i p_i \hat{\Gamma}_i$ for the lowest value of average energy $\text{tr}(\hat{H}\hat{\Gamma})$. Let us then consider one such admissible density operator $\hat{\Gamma}$. This operator will be admissible if and only if the average number of electron is given by N. But we can easily compute this average number since we have

$$\langle \hat{N} \rangle = \text{tr}(\hat{N}\hat{\Gamma}) = \sum_i p_i \text{tr}(\hat{N}\hat{\Gamma}_i) = \sum_i p_i N_i$$

where N_i is the (integer) number of electrons associated with the density operator $\hat{\Gamma}_i$. Any convex combination of operators $\hat{\Gamma}_i$ with integer number of electrons N_i, for which the average $\sum_i p_i N_i$ is N is then allowed. In particular, the following combination is valid

$$\hat{\Gamma} = (1-\omega)\hat{\Gamma}_{M-1} + \omega\hat{\Gamma}_M$$

where $\hat{\Gamma}_{M-1}$ and $\hat{\Gamma}_M$ are respectively the *minimizing* density operator for the system with $M-1$ and M electrons. Using the linearity of the Trace, we end up with

$$\text{tr}(\hat{H}\hat{\Gamma}) = (1-\omega)\text{tr}(\hat{H}\hat{\Gamma}_{M-1}) + \omega\text{tr}(\hat{H}\hat{\Gamma}_M)$$

Of course the right-hand side of the above equation is nothing else than $(1-\omega)E_0(M-1) + \omega E_0(M)$ because of our choice of $\hat{\Gamma}_{M-1}$ and $\hat{\Gamma}_M$. Since $(1-\omega)E_0(M-1) + \omega E_0(M)$ is independant of $\hat{\Gamma}$, we see that the above right-hand side is a *lower bound* to the energy $E_0(N)$, which by definition is the highest lower bound to $\text{tr}(\hat{H}\hat{\Gamma})$. We deduce the first inequality

$$E_0(N) \geq (1-\omega)E_0(M-1) + \omega E_0(M)$$

With the above inequality, it is then easy to conclude the proof. Indeed, we proved in part I that $E_0(N)$ was a *convex* function of N (see equation 2.15). So we have

$$E_0(N) = E_0((1-\omega)(M-1) + \omega M) \leq (1-\omega)E_0(M-1) + \omega E_0(M)$$

We now conclude from the two inequalities that

$$E_0(N) = (1-\omega)E_0(M-1) + \omega E_0(M) \tag{6.9}$$

QED.

6.2.3 Derivative of Ground State Energy with Electron Number

In the previous section, we established a very general property of the ground state energy. That property is generally true, whether we place ourselves in the framework of DFT or not. In this section we would like to establish another fundamental property of the total energy and related it to some key quantity appearing in the framework of DFT. More precisely, we will establish a connection between the *derivative* of the ground state energy with respect to the total number of electron N, with the *Fermi energy*, ϵ_F, that we introduced as a Lagrange multiplier associated with the conservation of the number of electrons in the system.

In order to avoid confusion, let us stress again that the ground state energy for a given number of electrons is $E_0(N)$, and the energy functional is E. From equation 3.7, we see that there exists some set of orbitals ϕ_i, ϕ_i^* and occupation numbers f_i, for which the ground state energy is calculated

$$E_0(N) = E(\phi_i(N), \phi_i^*(N), f_i(N))$$

Assuming some smoothness of the optimal parameters (ϕ_i, ϕ_i^* and f_i) with N, we can write down the derivative of the ground state energy with N using the chain rule (the derivative is computed from the optimal point corresponding to N electrons in the system)

$$\frac{dE_0}{dN} = \sum_i \int_{\mathbb{T}^3} \left[\frac{\delta E}{\delta \phi_i(\vec{x})} \frac{\partial \phi_i}{\partial N} + \frac{\delta E}{\delta \phi_i^*(\vec{x})} \frac{\partial \phi_i^*}{\partial N} \right] + \sum_i \frac{\partial E}{\partial f_i} \frac{\partial f_i}{\partial N}$$

To move forward, we now invoke the three optimality conditions, equations 3.11, 3.12 and 3.13 and simplify the above equation to

$$\frac{dE_0}{dN} = \sum_i \int_{\mathbb{T}^3} \left[\sum_j \lambda_{ij} \phi_j^* \frac{\partial \phi_i}{\partial N} + \sum_j \lambda_{ji} \phi_j \frac{\partial \phi_i^*}{\partial N} \right] + \sum_i \frac{\partial E}{\partial f_i} \frac{\partial f_i}{\partial N}$$

We remember that the orthonormality condition is always satisfied for the orbitals. As a consequence we have

$$\frac{\partial}{\partial N} \left[\int_{\mathbb{T}^3} \phi_i^*(\vec{x}) \phi_j(\vec{x}) d\vec{x} \right] = \frac{\partial}{\partial N} \left[\delta_{ij} \right] = 0$$

The above equation then leads to

$$\int_{\mathbb{T}^3} \phi_j \frac{\partial \phi_i^*}{\partial N} d\vec{x} = - \int_{\mathbb{T}^3} \frac{\partial \phi_j}{\partial N} \phi_i^* d\vec{x}$$

Using this result, we arrive at

$$\frac{dE_0}{dN} = \sum_{ij} \int_{\mathbb{T}^3} \left[\lambda_{ij} \phi_j^* \frac{\partial \phi_i}{\partial N} \right] - \sum_{ji} \int_{\mathbb{T}^3} \left[\lambda_{ji} \phi_i^* \frac{\partial \phi_j}{\partial N} \right] + \sum_i \frac{\partial E}{\partial f_i} \frac{\partial f_i}{\partial N}$$

The two terms on the left of the right-hand side naturally cancel each other, which simplifies the whole expression to

$$\frac{dE_0}{dN} = \sum_i \frac{\partial E}{\partial f_i} \frac{\partial f_i}{\partial N} \tag{6.10}$$

The last step consists in invoking the Aufbau principle of section 3.4.4. We know indeed, that for the optimal values of f_i (the ones leading to a minimum of the energy functional), all the occupations of the orbitals below the Fermi level are 1 and only the orbitals the closest to the Fermi level can have fractional occupation. As a consequence of this, we realize that **only the orbitals with fractional occupations** can lead to a variation of f_i with N as we change the total number of electrons. The direct corollary of that is the fact that only the orbitals with fractional occupation numbers enter the sum in equation 6.10. But if that is the case, then equations 3.13 and 3.16 lead to

$$\frac{\partial E}{\partial f_i} = \epsilon_F, \ \forall i \text{ such that } f_i \in]0,1[$$

That final result lets us conclude that

$$\frac{dE_0}{dN} = \epsilon_F \tag{6.11}$$

since $\sum_i \frac{\partial E}{\partial f_i} \frac{\partial f_i}{\partial N} = \epsilon_F \sum_i \frac{\partial f_i}{\partial N} = \epsilon_F \frac{\partial}{\partial N}(\sum_i f_i)$. This result is quite interesting because it tells us that we can compute the true derivative of the total energy of the system upon removal of a fraction of electron. And that derivative is nothing else than the Fermi level, which computationally speaking is the Lagrange multiplier to the constraint of electron number conservation.

6.2.4 Exact Ionization of the System in DFT

The two previous section allow us to establish a truly remarkable theorem in exact DFT. Indeed, if we couple the linearity of the ground state energy (equation 6.9) with the expression for its derivative (equation 6.11), we see that for a fractional number of electron N between integer values $M-1$ and M, the Fermi energy is independant of N

$$\epsilon_F(N) \text{ is constant for } N \text{ between } M-1 \text{ and } M$$

Now we imagine a process by which we go from $N = M$ to $N = M-1$. Then the total energy change in the ground state is given by

$$E_0(M-1) - E_0(M) = \int_M^{M-1} \frac{\partial E_0}{\partial N} dN = \int_M^{M-1} \epsilon_F dN = -\epsilon_F$$

But the above energy is nothing else than the **exact ionization** of the system. So the exact ionization of the M-electron system is given by the opposite to the Fermi energy

$$I_M = -\epsilon_F \tag{6.12}$$

As we can see, in exact DFT, the Fermi energy ϵ_F is endowed with a clear physical meaning. This result has been known in DFT ever since the work of Perdew *et al.* [79], but the above proof is also true if the functional is explicitly orbital-dependent as long as it is the exact one. It is important to note that Almbladh and Von Barth [5] also linked the exact ionization of the system to the energy of the highest occupied orbital in the case of Kohn-Sham theory for a unitary invariant functional.

6.2.5 Asymptotics of the Exact Exchange-Correlation Potential

In exact DFT, the exchange and correlation potential V_{XC} of a finite system (or away from a surface for a solid) can be shown to have the following asymptotics [5]

$$V_{XC}(\vec{x}) \sim -\frac{1}{|\vec{x}|} \qquad (6.13)$$

More precisely there exists a split in the exchange-correlation potential between an exchange only term and correlation only term. And we know what the asymptotics for each term is [97]. It reads

$$V_X(\vec{x}) \sim -\frac{1}{|\vec{x}|} \qquad (6.14)$$

and

$$V_C(\vec{x}) \text{ decays faster than a Coulombic potential} \qquad (6.15)$$

Let us motivate equation 6.13 with some simple physical argument. When a single electron ventures away from the system (at large distance from the nucleus or nuclei and all the other electrons), a net positive charge is left behind. As a consequence, at large distance, this electron should experience a Coulombic interaction potential $-\frac{1}{|\vec{x}|}$. Now, because of the form of the Kohn-sham potential, this Coulomb interaction is the net effect of three terms, namely the ionic Coulomb interaction of the nuclei, the Hartree potential (which depends upon the *whole* charge density $n(\vec{x})$), and the exchange-correlation potential. The asymptotics of the first two terms is given by Gauss's theorem

$$V_{\text{ion}} \sim -\frac{Z}{|\vec{x}|}, \quad V_H \sim \frac{Z}{|\vec{x}|}$$

We then see that the ionic and Hartree terms cancel each other out. So the net result is that the exchange-correlation potential should be decaying as $-\frac{1}{|\vec{x}|}$.

6.3 Qualitative Failures of ADF

We are now ready to explicit some of the most important *qualitative* failures of approximate density functionals (ADF) used in practice. Here we will focus on the so-called "orbital self-interaction", the "lack of binding" and the "band gap" problems. We will try to find both, the origin of the problem, and the consequences of it in terms of predictions.

6.3.1 Orbital Self-Interaction Effects

Going back to the one-electron limit mentioned earlier (equation 6.8), we would like to assess whether for a one-electron density $n(\vec{x})$ the Hartree and exchange-correlation energies cancel each other out. A simple way of visualizing this on a restricted set of single electron densities is to look at a family of Gaussian densities for which we have the freedom to change the "size" of the density cloud through a change of its Gaussian spread σ. This simplified view, although not representative of every possible single electron densities, should nevertheless teach us a great deal about the inaccuracies of common density functional approximations. For each width σ, the single electron density $n_\sigma(\vec{x})$ is given by

$$n_\sigma(\vec{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^3 e^{-\frac{|\vec{x}|^2}{2\sigma^2}} \tag{6.16}$$

Using that family of single-electron densities, Körzdörfer [53] computed the sum of the Hartree and exchange-correlation energies given by the Local Density Approximation. The result of that calculation is shown on figure 6-1. Since for a single electron
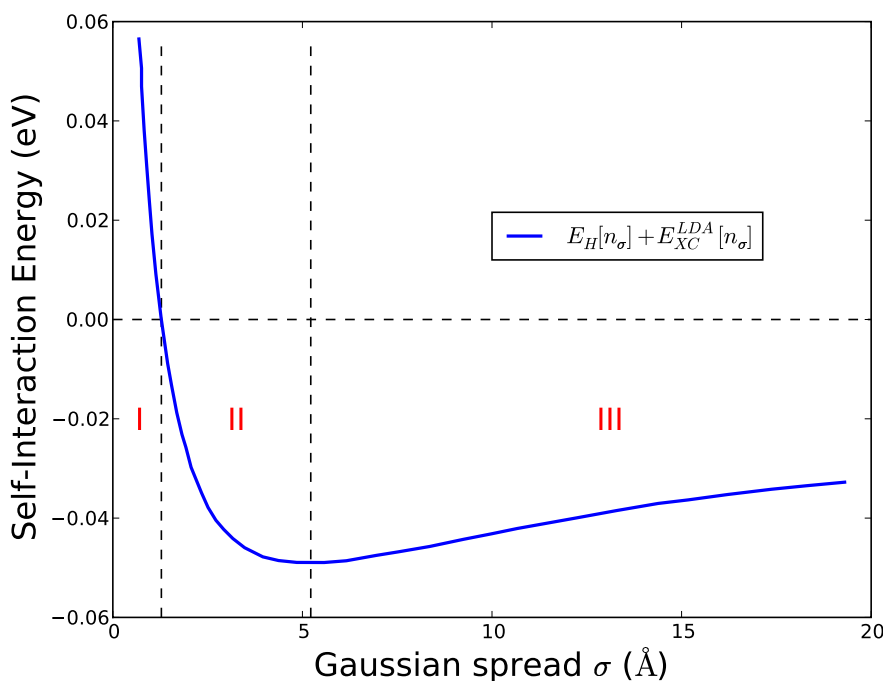
Figure 6-1: Visualization of the relationship between the degree of localization of an orbital and the strength of the orbital self-interaction as given by the sum of the Hartree and exchange-correlation energies. Adapted from Körzdörfer [53].

system the density is given by $n(\vec{x}) = f|\phi(\vec{x})|^2$, we see that the degree of localization of the orbital is directly related to the degree of localization of the density. In fact if

σ_ϕ is the "typical spread" of the orbital, then

$$\sigma_\phi = \sqrt{2}\sigma \qquad (6.17)$$

Now we can start analyzing figure 6-1. We see that for **localized** orbitals, with typical spread of less than $\sqrt{2} \times 1.3 \approx 1.85$Å (in region I on figure 6-1), the orbital self-energy in LDA is *positive*. The consequence is that those localized orbitals will tend to be energetically unfavored by LDA, leading to too small orbital energies. On the contrary, we see that for larger spreads (regions II and III of figure 6-1), the orbital self-interaction makes the total energy *decrease*. However, the difference in behavior is quite drastic between localized and delocalized orbitals. Indeed, when one expects localized orbitals to be strongly unfavored, we observe that over a very large span of spreads (roughly from 2.5Å to more than 20Å), the (negative) orbital self-interaction is pretty constant. This will, in turn, induce a quite homogeneous rigid shift of the orbital energies of delocalized orbitals making a comparison with experiment possible. Localized orbitals on the other hand, will see their orbital energies shifted up in energy. A perfect illustration of that phenomena is provided by the work of Dori *et al.* [21]. The focus of the work was to compare theory and experiment on a prototypical molecule used in the synthesis of organic molecular semiconductors : 3,4,9,10-perylene tetracarboxylic acid dianhydride or PTCDA (shown on figure 6-2). When compared to experiment, the eigenvalue spectrum given by LDA
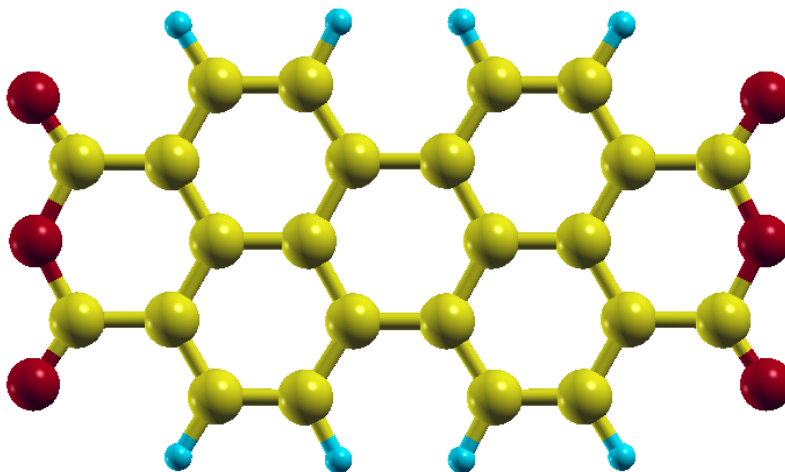
Figure 6-2: top view of 3,4,9,10-perylene tetracarboxylic acid dianhydride or PTCDA as studied by Dori *et al.* The extremal anhydride groups are strong electron attractors which leads to some localized molecular orbitals like A, A', B and B' (see [21]) and table 6.1)

and GGA functionals seem completely off. But a careful analysis of the situation on an orbital-by-orbital basis, confirms the trends discussed above. Table 6.1 gives the orbital energies of the computed six highest states of PTCDA compared to the

GW spectrum and experiment. What's obvious from this table is the large energy

State	Experiment	GW	LDA	GGA
HOMO	0	0	0	0
A	~ -1.5	-1.54	-0.59	-0.58
A'	~ -1.5	-1.54	-0.59	-0.58
B	~ -2.2	-2.24	-1.21	-1.20
B'	~ -2.2	-2.24	-1.21	-1.20
C	~ -1.5	-1.43	-1.34	-1.34
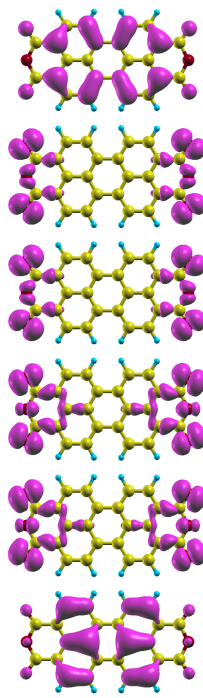
Table 6.1: Comparing orbital energies (in eV) in theory and experiment for the PTCDA molecule. The probability density corresponding to each molecular orbital is shown on the right-hand side. The difference in localization of the orbitals is then obvious. Adapted from Dori *et al.* [21]

renormalization of the four localized orbitals (A, A', B and B') compared to the two delocalized ones (HOMO and C). This is a very prototypical illustration of the fact that the "orbital self-interaction" in LDA and GGA affect much more localized orbitals than delocalized ones. As a consequence, whenever the physics of the system is driven by some interaction between localized and delocalized orbitals (like the interaction of d bands with s bands in transition metal compounds), then LDA and GGA might be *qualitatively* wrong in their predictions.

6.3.2 Lack of "Binding" in LDA and GGA

By "lack of binding", it is generally implied that LDA and GGA tend to predict negative ions to be unstable in contradiction with experiment [1]. The fundamental reason for this qualitatively wrong prediction has been linked to the lack of proper asymptotic behavior of the exchange-correlation potential. Remembering equation 6.13, we know that this potential should fall off as $-\frac{1}{|\vec{x}|}$ far away from any finite system (in a region where the density is exponentially suppressed). Looking at the analytical

expression of the exchange-correlation energy in appendix A, we can prove that the
exchange-correlation potential in LDA and GGA has the following asymptotic beha-
vior

$$V_{XC}^{\text{LDA}} \sim e^{-\alpha|\vec{x}|}, \ \alpha > 0$$
$$\text{and}$$
$$V_{XC}^{\text{GGA}} \sim e^{-\alpha|\vec{x}|}, \ \alpha > 0 \ \text{ or } \ -\frac{1}{|\vec{x}|^2}$$

For the PBE functional, the asymptotics is exponential, while some GGA functionals
by Becke [7] achieve a slower decay in $-\frac{1}{|\vec{x}|^2}$. Despite a slower decay of the exchange-
correlation potential for some GGA's, it has been shown [27] that a potential in $-\frac{1}{|\vec{x}|}$
is actually *necessary* for a Rydberg series[2] to exist, which itself is a requirement for
a system to bind an extra electron and hence lead to stable anions. At this point,
we would like to mention that the asymptotic behavior of the exchange-correlation
potential displays Coulomb-like behavior only for *neutral* systems, for which the net
charge is zero. In conclusion, we observe that LDA and GGA's cannot bind extra
electrons and hence lead to stable anions.

6.3.3 The "Band Gap" Problem and Derivative Discontinuities

Fundamentally, the "band gap" (or "charge gap" or "quasi-particle gap") is a measure
of how different a system reacts to an ionization process or the addition of an extra
electron. In fact, the definition of the gap is straightforward. It is exactly the sum of
the energy it takes to add an electron to the system and the energy it takes to remove
an electron from the system. If I_N denotes the ionization energy of the N-electron
system and A_N is the electron affinity of the N-electron system, then the gap E_g is
given by

$$E_g = (E_0(N-1) - E_0(N)) + (E_0(N+1) - E_0(N)) = I_N - A_N \qquad \boxed{6.18}$$

where $E_0(N)$ is the ground state energy of the N-electron system.

In DFT, the ionization theorem of equation 6.12 allows us to express the gap as a
difference in Fermi energies for respectively the $N+1$-electron system and N-electron
system

$$E_g = \epsilon_F(N+1) - \epsilon_F(N)$$

where $\epsilon_F(N)$ is the Fermi energy of the N-electron system. In order to relate that
gap to the Kohn-Sham band structure gap, we just have to be careful of the precise
definition of the latter. Indeed, when one computes the "band structure gap" in DFT,
one looks at the energy difference of the lowest unoccupied state of the N-electron
system (the LUMO) with the highest occupied state of the same N-electron system.
That latter state, the HOMO, is nothing else than the Fermi level. So inserting

---

[2]A Rydberg Series is a collection of electronic exited states in finite systems like atoms of molecules
that get closer and closer in energy as the states get closer to an ionization limit. This is reminiscent
to the Rydberg series of the hydrogen atom for which the energy difference between states goes as $\frac{1}{n^2}$.

$\epsilon_{\text{LUMO}}(N)$ into the previous equation allows us to re-write the gap as

$$E_g = \epsilon_F(N+1) - \epsilon_{\text{LUMO}}(N) + \epsilon_{\text{LUMO}}(N) - \epsilon_F(N) = \Delta + E_g^{\text{DFT}} \qquad (6.19)$$

We observe that the true gap, E_g, is never equal to the DFT gap, E_g^{DFT}, unless the constant $\Delta = \epsilon_F(N+1) - \epsilon_{\text{LUMO}}(N) = \epsilon_{\text{HOMO}}(N+1) - \epsilon_{\text{LUMO}}(N)$ is zero. Of course what matters in practice is how much does the two terms in 6.19 contribute to the true gap. If the term Δ is negligible, then we are confident that we can estimate accurately the true gap from the DFT gap. If not, then even with the exact functional, we would still need to add Δ to the DFT gap to get the true gap. The problem being that we don't really know how to compute Δ other than through a combination of equations 6.19 and 6.18. But these equations require us to compute total energies of systems with $N-1$, N and $N+1$ electrons which, for solids, is practically impossible.

The relative importance of Δ and E_g^{DFT} has been thoroughly investigated by Sham and Schluter [89] and Sham, Schluter and Godby [36]. Their conclusion is that both term have very similar contributions to the true gap. Table 6.2 shows that relative importance for some common semiconductors. From the table, it is quite clear that

Material	Δ	E_g^{DFT}	$E_g^{\text{DFT/LDA}}$	E_g (exp.)
Si	0.58	0.66	0.52	1.17
GaAs	0.67	0.91	0.67	1.63
AlAs	0.63	1.55	1.37	2.32
C (diamond)	1.12	4.21	3.90	5.48

Table 6.2: Values in eV of Δ, E_g^{DFT}, $E_g^{\text{DFT/LDA}}$ and E_g for some common semiconductors. As one can readily see, the Δ term is not at all negligible in all cases. What's more, the LDA band gap is found to be quite close to the exact DFT band gap. Adapted from [36].

the LDA band gap is far from the true band gap, but the question is why? For this, we have to go back to the fundamental equations of DFT (equations 2.18 and 3.6). With the splitting of the universal functional in "non-interacting kinetic", "Hartree" and "exchange-correlation" terms, one can apply the variational principle directly using the method of Lagrange multipliers and arrive at the general equation

$$\frac{\delta T_J}{\delta n(\vec{x})} + V(\vec{x}) + V_H(\vec{x}) + \frac{\delta E_{XC}}{\delta n(\vec{x})} = \epsilon_F$$

where T_J is the kinetic energy functional, $V(\vec{x})$ is the external potential, $V_H(\vec{x})$ is the Hartree potential, $\frac{\delta E_{XC}}{\delta n(\vec{x})}$ the exchange-correlation potential, and ϵ_F the Fermi energy (which is the Lagrange multiplier associated to the conservation of the electron number). Because of the continuity of the electron density, the Hartree potential and the external potential, when one goes from $N - \delta$ electrons to $N + \delta$ electrons in the

system ($0 < \delta \ll 1$), one can prove that the true gap can be expressed as [75]

$$E_g = \frac{\delta T_J}{\delta n(\vec{x})}\bigg|_{N-\delta}^{N+\delta} + \frac{\delta E_{XC}}{\delta n(\vec{x})}\bigg|_{N-\delta}^{N+\delta} \qquad (6.20)$$

In the limit of a large system, the discontinuity in the kinetic energy term can be directly related to the DFT band structure gap E_g^{DFT} [89]. This, in turn, proves that the Δ term that was mentioned earlier (see equation 6.19), was nothing else than the *derivative discontinuity* of the exchange-correlation potential.

Now we can understand why the LDA and GGA band gaps are far from experiment. Indeed we observe in equation 6.20 that the exchange-correlation potential should be discontinuous when one goes through an integer number of electrons in the system. But since the electron density is continuous when one adds a fraction of electron to the system, and that the analytical expressions of the LDA and GGA exchange-correlation energies are themselves continuous in $n(\vec{x})$, we see that there can be **no discontinuity** in the exchange-correlation potential. This is a serious issue, and in order to improve upon the LDA and GGA's, one needs to include a "natural" discontinuity of the exchange-correlation energy functional with the number of electrons.

7

Non-Koopman Correction to ADF

Time has come to present in detail what our new correction to approximate density functionals looks like. For this we will first digress and present the most influential correction, the one of Perdew and Zunger [77]. We will outline the idea behind it and show how that scheme performs on atomic and molecular systems. We will then build upon that scheme and introduce a substantial improvement over it, the one of Non-Koopman corrections [16]. The presentation will first introduce the scheme with its physical motivations and mathematical details, and then move on to show extensive applications for some important atomic, molecular and solid properties. In particular, we will present some new fully self-consistent band structure calculations of periodic solids using our real-space technique of section 5.3.

7.1 The Perdew-Zunger correction scheme

In a landmark paper [77], Perdew and Zunger introduced a correction scheme to approximate density functionals with a view to correct the spurious residual self-interaction. For this they chose to correct any functional on an orbital-to-orbital basis, building upon an insight given by the one-electron limit seen earlier in section 6.2.1. In this limit, we know that the sum of the Hartree and exchange-correlation energies should vanish.

$$E_H[n(\vec{x})] + E_{XC}[n(\vec{x})] = 0$$

Since the charge density formally looks like the addition of one-electron densities

$$n(\vec{x}) = \sum_i \rho_i(\vec{x}) = \sum_i f_i |\phi_i(\vec{x})|^2$$

the idea is that one can try to correct the energy functional by simply removing the "self-interaction" terms corresponding to every single electron densities ρ_i on an orbital-by-orbital basis. If $E^{\mathrm{ADF}}[n(\vec{x})]$ is the energy functional of an approximate density functional (ADF), then the Perdew-Zunger corrected functional is simply

$$E^{\mathrm{PZ}}[\{\rho_i(\vec{x})\}] = E^{\mathrm{ADF}}[n(\vec{x})] - \sum_i E_{HXC}^{\mathrm{ADF}}[\rho_i(\vec{x})] \qquad \boxed{7.1}$$

where E_{HXC}^{ADF} is the Hartree and exchange-correlation part of the approximate functional E^{ADF}. Two important observations can be drawn from the Perdew-Zunger correction scheme

- The Perdew-Zunger scheme is *general*, in the sense that any approximate density functional can be corrected (it is not bound to one particular form of approximate density functional)

- The Perdew-Zunger scheme is *orbital dependent*. The corrected Perdew-Zunger functional is now a functional of all the ρ_i, not just $n(\vec{x})$.

A question then arises : is the Perdew-Zunger functional a density functional ? This question is fundamental indeed, because if the Perdew-Zunger scheme is not a density functional anymore, then we are pushed away from the DFT formalism and all the results known in that area are lost. Fortunately, since the basic ingredients of DFT are the single particle orbitals ϕ_i and occupation numbers f_i, the scheme still falls into the realm of the Hohenberg-Kohn theorem and hence is a density functional [77]. However, that scheme does *not* belong to the Kohn-Sham paradigm [75].

7.1.1   PZ hamiltonian and implementation

Given the Perdew-Zunger functional of equation 7.1, one can right away derive the orbital-dependent hamiltonian similar to the one of equation 3.12 from the general formalism of chapter 3

$$\hat{h}_i^{PZ} = \hat{h}^{ADF} - V_H[\rho_i] - V_{XC}^{ADF}[\rho_i] \qquad (7.2)$$

where $\hat{h}^{ADF}$ is the Kohn-Sham hamiltonian corresponding to the Approximate Density Functional (ADF), $V_H[\rho_i]$ is the Hartree potential calculated at an electron density equal to ρ_i, and $V_{XC}^{ADF}[\rho_i]$ is the approximate exchange-correlation potential of the ADF calculated for an electron density equal to ρ_i.

When a functional is *not* orbital-dependent, we have seen in section 3.5.1 that one can directly work with a diagonal matrix of Lagrange multipliers ($\{\lambda_{ii}\}$), and solve for the orbitals ϕ_i and orbital energies $\epsilon_i = \lambda_{ii}$ directly by iteratively diagonalizing the orbital-*in*dependent Kohn-Sham hamiltonian $\hat{h}^{ADF}$. Unfortunately, when the functional *does* depend on the orbitals, such a diagonalization scheme is impossible, just because there are no hamiltonian to diagonalize. In implementing such schemes (like Perdew-Zunger and later Non-Koopman), one needs to introduce a matrix of Lagrange multipliers λ_{ij} to make sure that the single electron orbitals ϕ_i stay orthogonal to each other. What's more, one can only use a *direct* minimization of the functional. The Car-Parrinello method [13] is one such technique that directly attacks the problem of minimizing the functional without resorting to diagonalization techniques. As a consequence, the Perdew-Zunger scheme (and subsequently the Non-Koopman scheme) was introduced into the Car-Parrinello code of the *Quantum Espresso*[1] distribution.

---

[1] visit the homepage at http://www.quantum-espresso.org

7.1.2 General trends of PZ

A comprehensive study of the predictive power of the Perdew-Zunger scheme on atomic and molecular systems has been performed by Vydrov [101]. We will review some general trends of the Perdew-Zunger scheme for some physical and chemical properties and then move on to the next section on Non-Koopman corrections.

We first take a look at atomic ionization potentials and electron affinities using a ΔSCF technique, which means that those quantities are computed as total energy differences

$$\text{IP} = E(X^+) - E(X) \quad \text{and} \quad \text{EA} = E(X) - E(X^-) \tag{7.3}$$

where X is any atom or molecule, X^+ is the singly charged cation and X^- the singly charged anion. Results are given in table 7.1. We can observe an improvement in

Functional	IPs (MAD)	EAs (MAD)
LDA (LSDA)	0.30	0.29
PBE	0.15	0.13
PZ-LDA	0.24	0.18
PZ-PBE	0.39	0.57

Table 7.1: Computed Mean Absolute Deviation (MAD) from experiment of atomic ionization potentials (IP) and electron affinities (EA) for all atoms of the periodic table from H to Ar. IPs and EAs have been computed from ΔSCF calculations using the 6-311++G* basis set. Energies are in eV. Adapted from Vydrov [101].

the calculated IPs and EAs when going from LDA to PZ-LDA (i.e. Perdew-Zunger correction on top of LDA). But we can also see that a typical GGA (like PBE) performs even better than PZ-LDA. What's more, when the Perdew-Zunger correction is applied to functionals other than LDA (Vydrov explicitly tried PBE, BLYP, PBE0 and even TPSS) the results are *worsened*.

Moving on to molecular systems, the picture is very similar, except that this time, the PZ corrected functionals are worse in their predictions than the original approximate functionals as can be seen from table 7.2. As we can see, the Perdew-Zunger

Functional	IPs (MAD)	EAs (MAD)
LDA (LSDA)	0.20	0.21
PBE	0.20	0.10
PZ-LDA	0.25	0.21
PZ-PBE	0.42	0.60

Table 7.2: Computed Mean Absolute Deviation (MAD) from experiment of 44 molecular ionization potentials (IP) and 32 electron affinities (EA). IPs and EAs have been computed from ΔSCF calculations using the 6-311+G(3df,2p) basis set. Energies are in eV. Adapted from Vydrov [101].

correction, when it comes to predict total energy differences, does not improve predictions. In Vydrov's own words : "It seems that PZ-SIC overcorrects many-electron systems and overcorrection intensifies with increase in system size". This conclusion is in sharp contrast with the original Perdew-Zunger paper, where the focus was mostly on atomic systems and some approximations (for example the sphericalization of the charge density and the use of only diagonal lagrange multipliers) were used. Of course, since the Perdew-Zunger correction is exact for one electron systems, there must be some properties that are predicted correctly. We give here a rough summary of the general trends of the Perdew-Zunger correction scheme

- PZ improves the description of transition states in chemical reactions and also the dissociation curves of odd-electron systems [45, 73, 38]

- PZ does not improve reaction energies [37]

- PZ predicts too short bonds in molecular systems [37, 45]

7.2 The Non-Koopman correction scheme

We first present a fundamental observation from exact DFT to motivate the introduction of the Non-Koopman scheme. We then move on to the mathematics and implementation of that scheme in both "unrelaxed" and "relaxed" cases.

7.2.1 Origin and Motivation for Koopman's condition

We will motivate the introduction of the Non-Koopman correction scheme by first looking at what happens in exact DFT to the highest occupied orbital energy, namely the Fermi energy. As we saw in sections 6.2.3 and 6.2.4, for any of the highest occupied states, for which the frontier orbital is generically written $|\phi_f\rangle$, we have

$$\frac{dE_0}{dN} = \frac{dE_0}{dx_f} = \langle\phi_f|\hat{H}_f\phi_f\rangle = \epsilon_F \qquad (7.4)$$

The above equation is strictly true only when we have fractional occupations x_f for those frontier orbitals. Let us imagine to go from M electrons in the system to $M-1$ by depleting the frontier orbital $|\phi_f\rangle$[2] from full occupancy $x_f = 1$ to zero occupancy $x_f = 0$. When we do so, we allow all the orbitals to relax so as to keep the energy equal to the ground state energy for a total number of electron equal to $N \in]M-1,M[$ or equivalently a fractional occupation of the frontier orbital between 0 and 1, $x_f \in]0,1[$. Using the linearity of the total energy established in section 6.2.2, we observe that in exact DFT, the orbital energy $\langle\phi_f|\hat{H}_f\phi_f\rangle$ is *independent* of the orbital's occupation x_f

$$\langle\phi_f|\hat{H}_f\phi_f\rangle(x_f) = \epsilon_F \quad \forall x_f \in]0,1[\qquad (7.5)$$

---

[2]we consider the Fermi level to be non-degenerate here for clarity

The consequence being that we can interpret that orbital energy as a true ionization of the system

$$\langle \phi_f | \hat{H}_f \phi_f \rangle = \epsilon_F = E_0(x_f = 1) - E_0(x_f = 0)$$

The Non-Koopman scheme consists in imposing the same property to not only frontier orbitals but *all occupied orbitals* in the system. The fundamental key equation in the formalism is what we call the Koopman's condition

$$\frac{\partial}{\partial f_i} \left[\langle \phi_i | \hat{H}_i \phi_i \rangle \right] = 0 \quad \forall i \text{ and } \forall f_i \in]0, 1[\tag{7.6}$$

Let's clarify the exact meaning of the above equation. The interpretation goes as follow. If one depletes orbital ϕ_i from $f_i = f_i^0$ (its original occupation at the exact minimum of the functional) to $f_i = 0$, while keeping *all other occupations intact* but allowing for a re-optimization of all the orbitals, then the orbital energy $\langle \phi_i | \hat{H}_i \phi_i \rangle$ should not change at all during that process. Equation 7.6 is the mathematical translation of it. Imposing that condition allows for an interpretation of the orbital energies, from now on denoted by $\epsilon_i = \langle \phi_i | \hat{H}_i \phi_i \rangle$, in terms of electron *removal energies*. To see this, we define the following constraint energy function

$$E_0^i = \inf_{\{\phi_j, \phi_j^*\}} E[f_j, \{\phi_j\}, \{\phi_j^*\}] \tag{7.7}$$

In the above function, all the occupations f_j, except the one for orbital i, are unchanged with respect to the configuration minimizing the total energy functional for $N = M$ electrons. Only f_i is changed (as a parameter), and for every chosen value of f_i, the energy functional is minimized for the orbital degrees of freedom. Using a technique similar to the one we used in section 6.2.3, we can prove that the expression for the derivative of the constraint energy E_0^i with the occupation f_i is given by

$$\frac{dE_0^i}{df_i} = \langle \phi_i | \hat{H}_i \phi_i \rangle = \epsilon_i(f_i) \tag{7.8}$$

This result is sometimes known as the *Hellmann-Feynman* theorem, and it is notable to see that it derives straightforwardly from optimization theory. Using equation 7.8 and the fundamental theorem of integration we arrive at

$$E_0^i(f_i = 0) - E_0^i(f_i = f_i^0) = \int_{f_i^0}^{0} \frac{dE_0^i}{df_i} df_i = \int_{f_i^0}^{0} \epsilon_i(f_i) df_i \tag{7.9}$$

Imposing the Koopman's condition of equation 7.6, we end up with

$$E_0^i(f_i = 0) - E_0^i(f_i = f_i^0) = -\epsilon_i$$

The above equation tells us that $-\epsilon_i$ is exactly the electron removal energy in a relaxed vertical ionization process. Indeed we start from the exact ground state energy with $N = M$ electrons, $E_0(M) = E_0^i(f_i = f_i^0)$, and deplete only orbital i (f_i goes from f_i^0 to 0) letting the orbitals re-optimize along the way.

7.2.2 Non-Koopman Correction Scheme

Now that we made Koopman's condition explicit through equation 7.6, let us use it to devise a way to correct ADF so as to reintroduce physical meaning into the orbital energies. In general, orbital energies obtained from an ADF vary with the orbital's occupation

$$\epsilon_i(f_i) \text{ function of } f_i$$

As a consequence, for a given ADF, the electron removal energy reads

$$E_0^i(f_i = 0) - E_0^i(f_i = f_i^0) = \int_{f_i^0}^0 \frac{dE_0^i}{df_i} df_i = \int_{f_i^0}^0 \epsilon_i(f_i) df_i$$

If the functional would satisfy equation 7.6, then the above equation would reduce to

$$E_0^i(f_i = 0) - E_0^i(f_i = f_i^0) = \int_{f_i^0}^0 \frac{dE_0^i}{df_i} df_i = -f_i^0 \epsilon_i(f_i^{\text{ref}})$$

for some reference occupation f_i^{ref} between 0 and f_i^0. We are then led to the definition of "Non-Koopman" terms as the difference of what we have to what we would want to have, namely

$$\Pi_i(f_i^{\text{ref}}, f_i^0) = E_0^i(f_i = 0) - E_0^i(f_i = f_i^0) - (-f_i^0 \epsilon_i(f_i^{\text{ref}})) = \int_0^{f_i^0} \left(\epsilon_i(f_i^{\text{ref}}) - \epsilon_i(f) \right) df \quad \boxed{7.10}$$

In order to move forward and start correcting ADFs, we place ourselves in a constraint situation, where it is easier to find the appropriate way to correct the functional. This restricted situation corresponds to the *unrelaxed* case, for which the orbitals and occupation numbers are left unchanged and only the occupation of orbital i is allowed to change. In this case, we introduce some functions $\delta_i(f)$ such that the corrected functional looks like

$$E^{\text{ADF/K}} = E^{\text{ADF}} + \sum_i \delta_i(f_i)$$

The superscript in $E^{\text{ADF/K}}$ acknowledges the fact that we want the corrected functional to satisfy the unrelaxed Koopman's condition. When the functional is computed at the equilibrium occupation $f_i = f_i^0$, then the energy is

$$E^{\text{ADF/K}}(f_j^0, f_i^0) = E^{\text{ADF}}(f_j^0, f_i^0) + \delta_i(f_i^0) + \sum_{j \neq i} \delta_j(f_j^0)$$

When we go through our unrelaxed vertical ionization process, we end up with a depleted i-th orbital, and the total energy becomes

$$E^{\text{ADF/K}}(f_j^0, f_i = 0) = E^{\text{ADF}}(f_j^0, f_i = 0) + \delta_i(0) + \sum_{j \neq i} \delta_j(f_j^0)$$

Since we want the corrected functional to have the proper Koopman's behavior, we then wish to have (in the following we drop the dependence on other occupations f_j for clarity)

$$E^{\text{ADF/K}}(f_i = 0) - E^{\text{ADF/K}}(f_i^0) = E^{\text{ADF}}(f_i = 0) - E^{\text{ADF}}(f_i^0) + \delta_i(0) - \delta_i(f_i^0) = -f_i^0 \epsilon_i(f_{\text{ref}})$$

On the other hand we can generally express the first difference on the right-hand side as

$$E^{\text{ADF}}(f_i = 0) - E^{\text{ADF}}(f_i^0) = \int_{f_i^0}^0 \epsilon_i(f) df$$

So we see that the correction $\delta_i(0) - \delta_i(f_i^0)$ needs to satisfy

$$\delta_i(0) - \delta_i(f_i^0) = -f_i^0 \epsilon_i(f_{\text{ref}}) - \int_{f_i^0}^0 \epsilon_i(f) df = -\Pi_i^u(f_i^{\text{ref}}, f_i^0)$$

the last equality comes from the definition of Π_i in equation 7.10 and the superscript u is here to remind us of the "frozen-orbital" approximation (or unrelaxed approximation) that we made. We then choose the function $\delta_i(f)$ such that $\delta_i(0) = 0$, and finally arrive at

$$\delta_i(f) = \Pi_i^u(f_i^{\text{ref}}, f)$$

From this analysis, it is quite clear that the proper definition of the *unrelaxed* Non-Koopman corrected functional is

$$E^{\text{NK}} = E^{\text{ADF}} + \sum_i \Pi_i^u(f_i^{\text{ref}}, f_i) \qquad \boxed{7.11}$$

Similarly to the Perdew-Zunger scheme, we find that

- The Non-Koopman scheme is *general*, in the sense that any approximate density functional can be corrected (it is not bound to one particular form of approximate density functional)

- The Non-Koopman scheme is *orbital dependent*. The corrected Non-Koopman functional is now a functional of all the ρ_i, not just $n(\vec{x})$.

7.2.3   Implementation of the Non-Koopman Scheme

Given the expression of the corrected functional in equation 7.11, we can derive a variational scheme to find the extrema of the corrected functional. For this, we need to compute the functional derivative of 7.11 with respect to ϕ_i and ϕ_i^*. Another way to find an expression for the Non-Koopman orbital dependent hamiltonian is to directly compute the functional derivative of the energy functional with respect to the single electron density $\rho_{i\sigma} = f_i |\phi_{i\sigma}|^2$ (as a matter of generality we now also explicitly include the spin dependence σ). The Non-Koopman corrective terms can be recast into the

explicit functional form[3]

$$
\begin{aligned}
\Pi_{i\sigma}^{u,\mathrm{LSD}}(f_{\mathrm{ref}}) &= f_{i\sigma}(2f_{\mathrm{ref}}-f_{i\sigma})E_{\mathrm{H}}[n_{i\sigma}]-E_{\mathrm{xc}}^{\mathrm{LSD}}[\rho]+E_{\mathrm{xc}}^{\mathrm{LSD}}[\rho-\rho_{i\sigma}] \\
&\quad + \int d\mathbf{r}\rho_{i\sigma}(\mathbf{r})v_{\mathrm{xc},\sigma}^{\mathrm{LSD}}(\mathbf{r};[\rho_{i\sigma}^{\mathrm{ref}}])d\mathbf{r}
\end{aligned}
\tag{7.12}
$$

where $n_{i\sigma}(\mathbf{r}) = |\phi_{i\sigma}|^2(\mathbf{r})$ and $\rho_{i\sigma}(\mathbf{r}) = f_{i\sigma}n_{i\sigma}(\mathbf{r})$. Here, the electronic density $\rho_{i\sigma}^{\mathrm{ref}}(\mathbf{r})$ stands for the reference transition-state density

$$
\begin{aligned}
\rho_{i\sigma}^{\mathrm{ref}}(\mathbf{r}) &= f_{\mathrm{ref}}n_{i\sigma}(\mathbf{r})+\sum_{j\sigma'\neq i\sigma} f_{j\sigma'}n_{j\sigma'}(\mathbf{r}) \tag{7.13} \\
&= \rho(\mathbf{r})+(f_{\mathrm{ref}}-f_{i\sigma})n_{i\sigma}(\mathbf{r}) \tag{7.14}
\end{aligned}
$$

The expression of the functional derivatives reads

$$
\frac{\delta\Pi_{i\sigma}^{u,\mathrm{LSD}}}{\delta\rho_{i\sigma}(\mathbf{r})} = (f_{\mathrm{ref}}-f_{i\sigma})v_{\mathrm{H}}(\mathbf{r};[n_{i\sigma}])+v_{\mathrm{xc},\sigma}^{\mathrm{LSD}}(\mathbf{r};[\rho_{i\sigma}^{\mathrm{ref}}])-v_{\mathrm{xc},\sigma}^{\mathrm{LSD}}(\mathbf{r};[\rho])+w_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}(\mathbf{r})
\tag{7.15}
$$

In Eq. (7.15), the potential $w_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}$ denotes

$$
\begin{aligned}
w_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}(\mathbf{r}) &= f_{\mathrm{ref}}\Bigg[\int d\mathbf{r}'f_{\mathrm{Hxc},\sigma\sigma}^{\mathrm{LSD}}(\mathbf{r},\mathbf{r}';[\rho_{i\sigma}^{\mathrm{ref}}])n_{i\sigma}(\mathbf{r}') \\
&\quad - \int d\mathbf{r}'d\mathbf{r}''f_{\mathrm{Hxc},\sigma\sigma}^{\mathrm{LSD}}(\mathbf{r}',\mathbf{r}'';[\rho_{i\sigma}^{\mathrm{ref}}])n_{i\sigma}(\mathbf{r}')n_{i\sigma}(\mathbf{r}'')\Bigg]
\end{aligned}
\tag{7.16}
$$

where $f_{\mathrm{Hxc},\sigma\sigma'}^{\mathrm{LSD}}(\mathbf{r},\mathbf{r}') = \delta^2(E_{\mathrm{H}}+E_{\mathrm{xc}}^{\mathrm{LSD}})/\delta\rho_\sigma(\mathbf{r})\delta\rho_{\sigma'}(\mathbf{r}')$ is the second-order functional derivative of the LSD energy. Focusing then on the cross derivatives, we obtain

$$
\frac{\delta\Pi_{j\sigma'}^{u,\mathrm{LSD}}}{\delta\rho_{i\sigma}(\mathbf{r})} = v_{\mathrm{xc},\sigma}^{\mathrm{LSD}}(\mathbf{r};[\rho-\rho_{j\sigma'}])-v_{\mathrm{xc},\sigma}^{\mathrm{LSD}}(\mathbf{r};[\rho])+\int d\mathbf{r}'f_{\mathrm{xc},\sigma\sigma'}^{\mathrm{LSD}}(\mathbf{r},\mathbf{r}';[\rho_{j\sigma'}^{\mathrm{ref}}])\rho_{j\sigma'}(\mathbf{r}')
\tag{7.17}
$$

where $f_{\mathrm{xc},\sigma\sigma'}^{\mathrm{LSD}}(\mathbf{r},\mathbf{r}')$ is the exchange-correlation contribution to $f_{\mathrm{Hxc},\sigma\sigma'}^{\mathrm{LSD}}(\mathbf{r},\mathbf{r}')$. As a final result, the orbital-dependent NK Hamiltonian can be cast into the form

$$
\hat{h}_{i\sigma}^{\mathrm{NK}} = \hat{h}^{\mathrm{LSD}}[\rho_{i\sigma}^{\mathrm{ref}}]+\hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}+\hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}},
\tag{7.18}
$$

where $w_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}$ stands for the cross-derivative potential

$$
w_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}(\mathbf{r}) = \sum_{j\sigma'\neq i\sigma} \frac{\delta\Pi_{j\sigma'}^{u,\mathrm{LSD}}}{\delta\rho_{i\sigma}(\mathbf{r})}.
\tag{7.19}
$$

---

[3]in the rest of this thesis we will focus exclusively on the LSD functional (Local Spin Density Approximation). Of course the Non-Koopman correction scheme works on any ADF.

In a nutshell, the NK Hamiltonian consists of the uncorrected LSD Hamiltonian calculated at the reference density $\hat{h}^{\mathrm{LSD}}[\rho_{i\sigma}^{\mathrm{ref}}]$ with the addition of two variational potentials. The first additional term $\hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}$ results from the variation of the reference density as a function of $\rho_{i\sigma}$ while the second term $\hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}$ springs from the cross-dependence of the non-Koopmans corrective terms. The effect of the $\hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}$ and $\hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}$ contributions that arise as by-products of variationality is analyzed in the next section.

7.2.4 Assessing the Non-Koopman Correction

In this section, we assess the performance of the NK self-interaction correction, particularly focusing on the effect of variational terms on the accuracy of NK orbital predictions (i.e., on the cancellation of the unrelaxed frozen-orbital self-interaction measure $\Pi_{i\sigma}^{u,\mathrm{NK}}(f)$).

One simple and probably the most direct way to evaluate the influence of $\hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}$ and $\hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}$ is to introduce a non-variational orbital-energy scheme, the NK_0 method, that consists of freezing the dependence of the reference transition-state densities and the cross-dependence of corrective energy terms, thereby eliminating $\hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}$ and $\hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}$ contributions to the effective potential. Computed NK and NK_0 orbital levels can then be compared for the direct assessment of $\hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}$ and $\hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}$ errors. Explicitly, the NK_0 Hamiltonian can be written as

$$\hat{h}_{i\sigma}^{\mathrm{NK}_0} = \hat{h}^{\mathrm{LSD}}[\rho] + \left. \frac{\delta\Pi_{i\sigma}^{u,\mathrm{LSD}}}{\delta\hat{\rho}_{i\sigma}}\right|_{\rho_{i\sigma}^{\mathrm{ref}}=\mathrm{cst}}, \qquad (7.20)$$

$$\hat{h}_{i\sigma}^{\mathrm{NK}_0} = \hat{h}^{\mathrm{LSD}}[\rho_{i\sigma}^{\mathrm{ref}}]. \qquad (7.21)$$

In the NK_0 optimization scheme, the Hamiltonian given by Eq. (7.21) is employed to propagate orbital degrees of freedom at fixed $\rho_{i\sigma}^{\mathrm{ref}}$. Reference transition-state densities are then updated according to Eq. (7.13). The procedure is iterated until self-consistency.

Due to the loss of variationality, the obvious practical limitation of the non variational NK_0 orbital-energy method is that it cannot provide total energies and interatomic forces. However, NK_0 is of great utility in evaluating the intrinsic performance of the NK correction. In itself, the NK_0 formulation is also useful in determining orbital energy properties that are particularly affected by $\hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}$ and $\hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}$ errors.

Focusing now on computational predictions, the occupation dependencies of the LSD, HF, PZ, and NK unrelaxed orbital energies

$$\epsilon_{i\sigma}^{u}(f) = \left. \frac{dE_{i\sigma}^{u}(f')}{df'}\right|_{f'=f} \qquad (7.22)$$

of the highest atomic orbital of carbon are depicted in Figs. 7-1 and 7-2(a). The sali-

ent feature of the LSD graph is the large variation of the orbital energy from -19.40 to -6.15 eV, reflecting the strong nonlinearity of the corresponding unrelaxed ionization curve. The PZ variation is found to be twice lower than for LSD. In contrast,
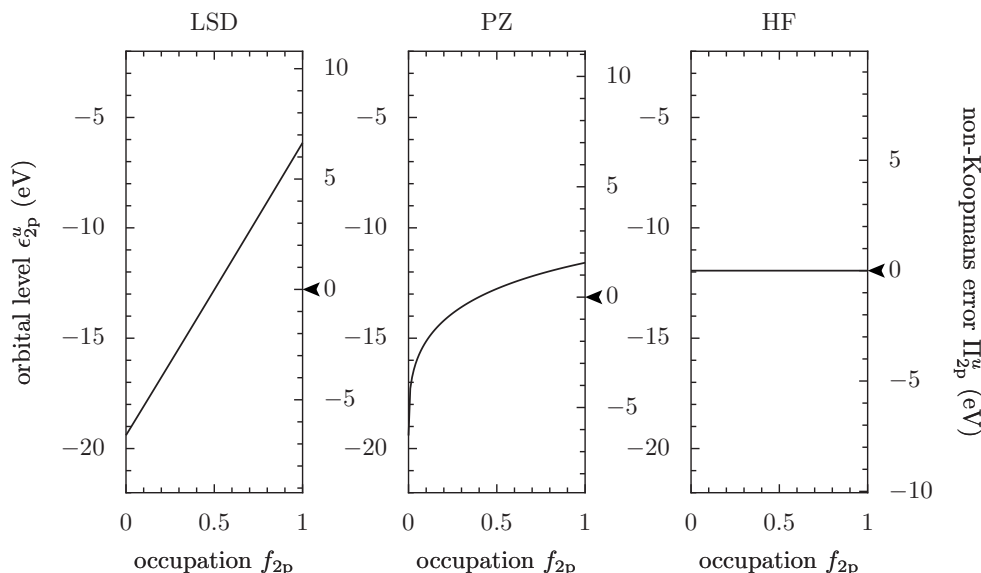

Figure 7-1: LSD, PZ, and HF unrelaxed orbital energies and residual unrelaxed non-Koopmans errors of the highest occupied state of carbon. The black arrow highlights the zero value for the non-Koopmans error scale.

the HF unrelaxed ionization curve exhibits a perfectly linear behavior (i.e., the unrelaxed orbital energy remains constant). This trend is closely reproduced by the NK functional [Fig. 7-2(a)]; on the scale of LSD residual non-Koopmans errors, the eye can barely distinguish any deviation of the NK unrelaxed orbital energies as a function of $f_{i\sigma}$ regardless of the value of the reference occupation.

The above observation is due to the fact that the variational contribution $\hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}}$ affects the orbital energy $\epsilon_{i\sigma}^{\mathrm{NK}}$ indirectly, i.e., only through the self-consistent response of the orbital densities since $\langle \phi_{i\sigma} | \hat{w}_{\mathrm{ref},i\sigma}^{\mathrm{LSD}} | \phi_{i\sigma} \rangle = 0$ at self-consistency. Furthermore, a Taylor series expansion of $\hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}}$ reveals that $\langle \phi_{i\sigma} | \hat{w}_{\mathrm{xd},i\sigma}^{\mathrm{LSD}} | \phi_{i\sigma} \rangle$ does not cause notable departure from the linear Koopmans behavior. In quantitative terms, the dominant term in the expansion of the residual NK Non-Koopman's error is of the fourth order in orbital densities:

$$
\begin{aligned}
\Pi_{i\sigma}^{u,\mathrm{NK}}(f) \;=\; & \frac{1}{4} \sum_{j\sigma' \neq i\sigma} f_{j\sigma'}(2f_{\mathrm{ref}} - f_{j\sigma'})(2f - f_{i\sigma})f_{i\sigma} \times \\
& \times \int d\mathbf{r}_{1234} f_{\mathrm{xc},\sigma\sigma'\sigma'\sigma}^{(4),\mathrm{LSD}}(\mathbf{r}_{1234};[\rho_{j\sigma'}^{\mathrm{ref}}]) n_{i\sigma}(\mathbf{r}_1) n_{j\sigma'}(\mathbf{r}_2) n_{j\sigma'}(\mathbf{r}_3) n_{i\sigma}(\mathbf{r}_4) + \cdots
\end{aligned}
$$

(where $f_{\mathrm{xc},\sigma_{12\cdots n}}^{(n),\mathrm{LSD}}(\mathbf{r}_{12\cdots n})$ denotes the nth order functional derivative of the LSD exchange correlation energy), whereas the PZ correction is found to be less accurate

(a) NK

(b) NK$_0$
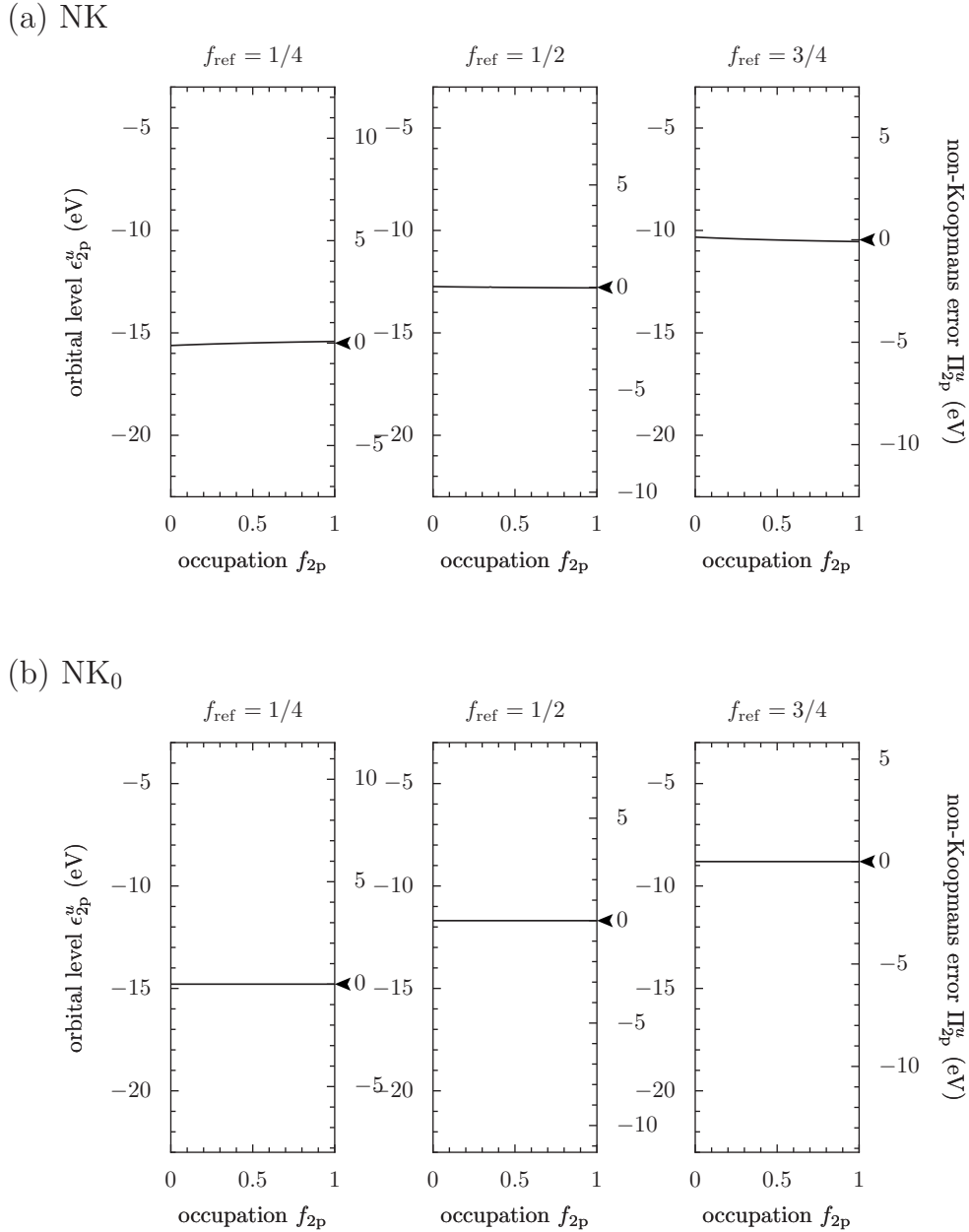
Figure 7-2: Non-Koopmans unrelaxed orbital energies and residual unrelaxed non-Koopmans errors of the highest occupied state of carbon for select arbitrary values of the reference occupation ($f_{\text{ref}} = \frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$) using (a) the NK total-energy method and (b) the NK$_0$ orbital-energy method. The black arrow highlights the zero value for the non-Koopmans error scale.

in minimizing the Non-Koopman's error by one order of precision:

$$\Pi_{i\sigma}^{u,\text{PZ}}(f) = \sum_{j\neq i} \frac{2f - f_{i\sigma}}{2} f_{i\sigma} f_{j\sigma} \int d\mathbf{r}_{123} f_{\text{xc},\sigma\sigma\sigma}^{(3),\text{LSD}}(\mathbf{r}_{123};[\rho - \rho_{i\sigma}]) n_{i\sigma}(\mathbf{r}_1) n_{i\sigma}(\mathbf{r}_2) n_{j\sigma}(\mathbf{r}_3) + \cdots$$

Despite the very good accuracy of the NK correction, direct confrontation with NK_0 results (for which the non-Koopmans measure $\Pi_{i\sigma}^{u,NK_0}(f)$ [Eq. (7.10)] is obviously canceled for any value of f) reveals that NK tends to underestimate orbital energies with deviations of 0.5 to 1.5 eV that gradually increase with f_{ref} [Fig. 7-2(a,b)].

As a conclusion of this preliminary performance evaluation, the NK frozen-orbital correction results in a considerable reduction of residual errors $\Pi_{i\sigma}^{u,NK}(f)$, bringing density-functional approximations in nearly exact agreement with the frozen-orbital linear trend while exhibiting a slight tendency to underestimate orbital energies. In the next section, we present an extension of the NK correction beyond the frozen-orbital paradigm.

7.2.5 Screened Non-Koopman scheme

So far, we have been focusing on imposing a constant behavior of the orbital energies in the "frozen orbital" approximation. Of course when one goes through the self-consistent cycle, the *relaxed* orbital energies will not necessarily stay constant as we change their occupations $f_{i\sigma}$ and re-optimize for the orbitals $\phi_{i\sigma}$. Can we go beyond the "frozen orbital" paradigm and try to shoot for a Koopman's condition with orbital relaxation ? In order to address that issue we will look at the difference between differential ionizations and affinities and total ionizations and affinities in DFT.

By differential ionization/affinity, we mean to compute the ionization/affinity of the system by looking at the highest occupied or lowest unoccupied orbital energies. Indeed we know from section 7.2.1 that the highest occupied orbital is related to the change of the total energy with the total number of electrons of the M-electron system

$$\epsilon_{HOMO} = \frac{dE_0}{dN}$$

We can draw a similar conclusion for the lowest unoccupied orbital of the $M-1$-electron system. On the other hand, we can compute a total ionization/affinity by simply computing the total energy difference between the M-electron system and $M-1$-electron system. But since the total energy function $E_0(N)$ is known to be piecewise linear from section 6.2.2, we see that in exact DFT, we should have a complete agreement between differential ionizations/affinities and total ionizations/affinities. It is an empirical fact, that most if not all of ADFs are pretty good in computing ionizations/affinities through the total ionization/affinity method (as one can see in tables 7.1 and 7.2) but extremely poor at computing the same quantity through the differential ionization/affinity method. The fundamental reason for that behavior is generally related to the *convex* behavior of $E_0(N)$ for ADFs. To illustrate that point, we show in figure 7-3 the computed total energy with the total electron number for two molecules, namely methane CH_4 and ammonia NH_3, in contact with an electron reservoir.

What we observe is a confirmation of the *convex* behavior of the total energy in the LDA case. This behavior explains the discrepancy between total and partial ionizations. The derivatives of the total energy at $N=7$ electrons (to the right) and $N=8$ electrons (to the left) are respectively lower and greater than the total ionization
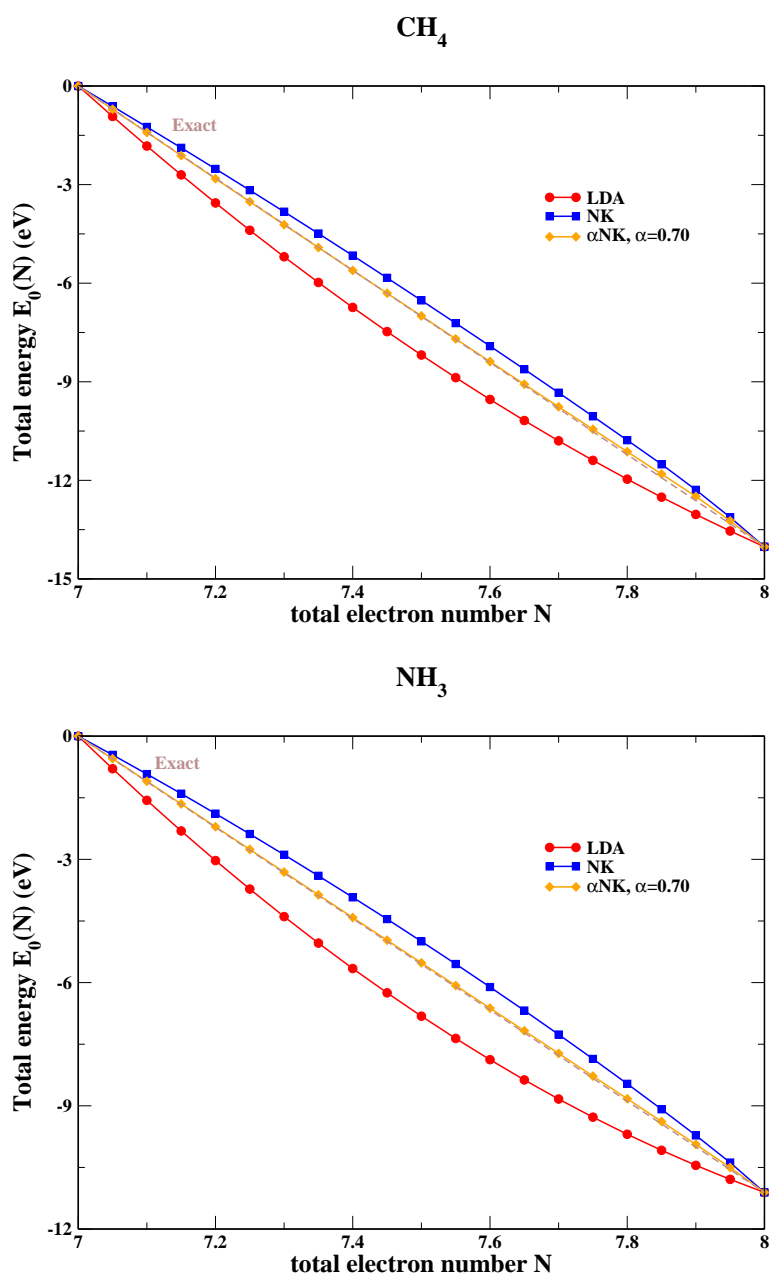
Figure 7-3: Behavior of the total energy $E_0(N)$ with the total number of electron N as computed by the LDA and NK functionals. Other semilocal and even hybrid functionals would exhibit identical trends to LDA. We show in dashed line what the exact DFT result looks like. The top plot corresponds to a methane molecule CH_4 in contact with an electron reservoir, while the bottom plot corresponds to a similar setup but for an ammonia molecule NH_3. We observe a clear *convex* behavior for LDA whereas the NK curve is *concave*. A mix of the two functionals with $\alpha \approx 0.7$ leads to an almost linear behavior of the total energy.

energy represented by the slope of the exact curve (in dashed line). What's more, because of the convexity property, we see that the differential affinity — corresponding to the LUMO of the $M-1$-electron system —will be lower than the differential ionization — corresponding to the HOMO of the M-electron system —. We stress that in exact DFT, these two orbital energies should agree. Looking at the NK total energy, we see that the curve is *concave*, so that the results discussed in the LDA case are inverted. Using that fact, we can envision to mix some part of the LDA functional with some part of the NK functional so as to make sure that the differential ionization and the differential affinity agree (which also leads to an almost linear behavior of the total energy with electron number). Let us then introduce the following functional

$$E^{\alpha\mathrm{NK}} = E^{\mathrm{LSD}} + \alpha\left(E^{\mathrm{NK}} - E^{\mathrm{LSD}}\right) \hspace{2em} \boxed{7.23}$$

The fact that we can find an α for which the differential ionization/affinity energies agree is illustrated for a Carbon atom in figure 7-4. The inclusion of the α parameter

Figure 7-4: LSD differential ionization/affinity energies $\mathscr{I}_N^{\mathrm{LSD}}$, $\mathscr{A}_{N-1}^{\mathrm{LSD}}$, and their average $\frac{1}{2}(\mathscr{I}_N^{\mathrm{LSD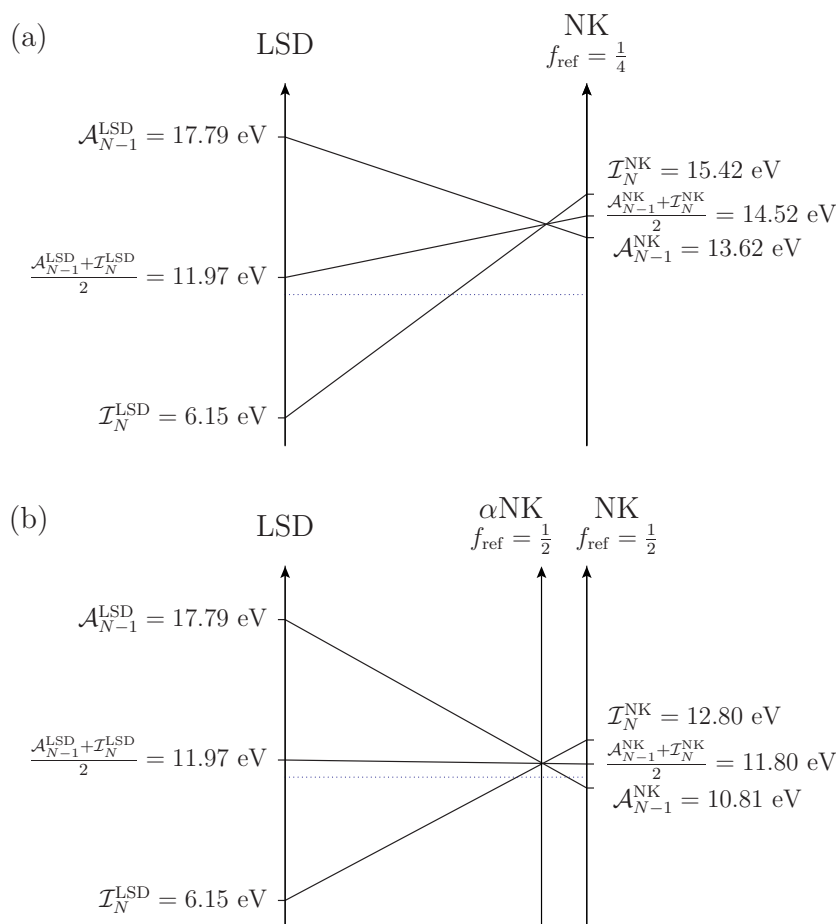}} + \mathscr{A}_{N-1}^{\mathrm{LSD}})$ compared with NK differential ionization/affinity energies for (a) $f_{\mathrm{ref}} = \frac{1}{4}$ and (b) $f_{\mathrm{ref}} = \frac{1}{2}$ for carbon (here N=6 electrons).

in the Non-Koopman scheme is an attempt to deal with orbital relaxation effects and get closer to the relaxed Koopman's condition of equation 7.6. One could think that the introduction of α destroys the *ab-initio* nature of the Non-Koopman method, but as we saw earlier, α is not arbitrary. It is the coefficient for which the LUMO energy of the $M-1$-electron system agrees with the HOMO energy of the M-electron system. α can be seen as a screening coefficient that helps us taking into account the relaxation of the system when one goes from M to $M-1$ electrons. Assuming a linear behavior of the frontier orbital energies with α, we can compute its approximate value through the following formula

$$\alpha \approx \frac{\epsilon_{\text{LUMO}}^{\text{LSD}}(M-1) - \epsilon_{\text{HOMO}}^{\text{LSD}}(M)}{\left(\epsilon_{\text{LUMO}}^{\text{LSD}}(M-1) - \epsilon_{\text{HOMO}}^{\text{LSD}}(M)\right) - \left(\epsilon_{\text{LUMO}}^{\text{NK}}(M-1) - \epsilon_{\text{HOMO}}^{\text{NK}}(M)\right)}. \qquad \boxed{7.24}$$

To give further support to the interpretation of α as a screening coefficient, we calculated that self-consistent coefficient (α^{NK}) for the first five rows of the periodic table (from Hydrogen, H, to Xenon, Xe). The results are shown in figure 7-5. As readily
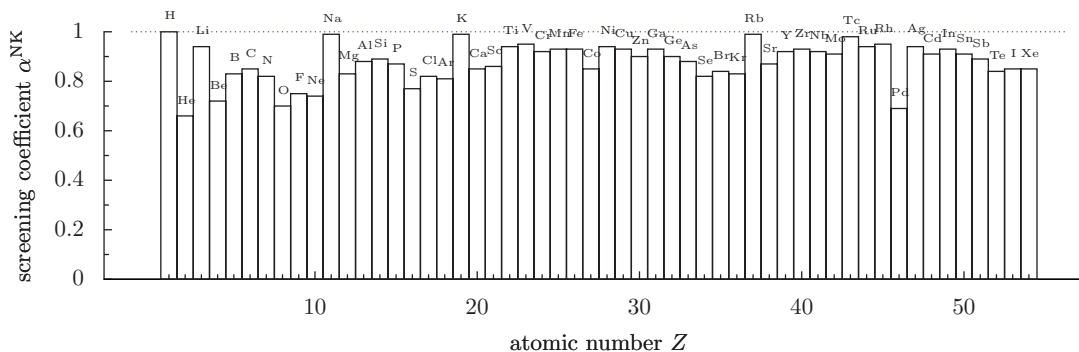
Figure 7-5: Non-Koopmans screening coefficient α for the elements of the first five rows of the periodic table.

seen from the histogram, the elements with a single electron outside of otherwise complete shells (like Hydrogen, Lithium, Sodium, Potassium and Rubidium) are the least affected by the orbitals reconfiguration, and so their screening coefficient is very close to 1 (for example $\alpha^{\text{NK}}(\text{Na})=0.99$). At the other extreme, some of the most sensitive elements to orbital reconfiguration are the noble gases. It is now time to give the final expression of the screened Non-Koopman functional. We will not dwell into the details of why choosing a reference occupation of one half is optimal. We direct the interested reader to section II.F of the reference paper [16] to find out more. The explicit expression of the screened αNK functional with $f_{\text{ref}} = \frac{1}{2}$ reads

$$E^{\alpha\text{NK}} = E^{\text{LSD}} \quad + \quad \alpha^{\text{NK}} \sum_{i\sigma} \left[f_{i\sigma}(1-f_{i\sigma})E_{\text{H}}[n_{i\sigma}] + E_{\text{xc}}^{\text{LSD}}[\rho - \rho_{i\sigma}] \right. \qquad \boxed{7.25}$$

$$\left. + \int d\mathbf{r} v_{\text{xc},\sigma}^{\text{LSD}}\left(\mathbf{r}; \left[\rho + (\frac{1}{2} - f_{i\sigma})n_{i\sigma}\right]\right)\rho_{i\sigma}(\mathbf{r}) - E_{\text{xc}}^{\text{LSD}}[\rho] \right]$$

In the next section, we will apply the Non-Koopman and screened Non-Koopman schemes to the calculation of many atomic, molecular and solid-state properties.

7.3 Predictive Power of the Non-Koopman Correction

In this section, we will present some extensive atomic, molecular and solid-state calculations with the aim of assessing the predictive power of the Non-Koopman correction scheme. We will try to systematically compare NK (in its NK, NK_0, αNK and αNK_0 forms) with LSD, PZ and sometimes HF or higher order schemes like GW when appropriate. Also, whenever possible, we will include experimental results.

7.3.1 Atomic Ionizations and Affinities

We start our exploration of NK's performance by calculating the electron removal energies of a complete range of atomic elements, from hydrogen to xenon, using the all-electron LD1 code of the QUANTUM-ESPRESSO distribution. The LD1 code proceeds by iterative integration of the spherically symmetric electronic-structure problem on logarithmic grids.

we compare αNK differential electron affinity predictions[4] with LSD and experiment in Fig. 7-6. The comparison demonstrates the predictive ability of the αNK
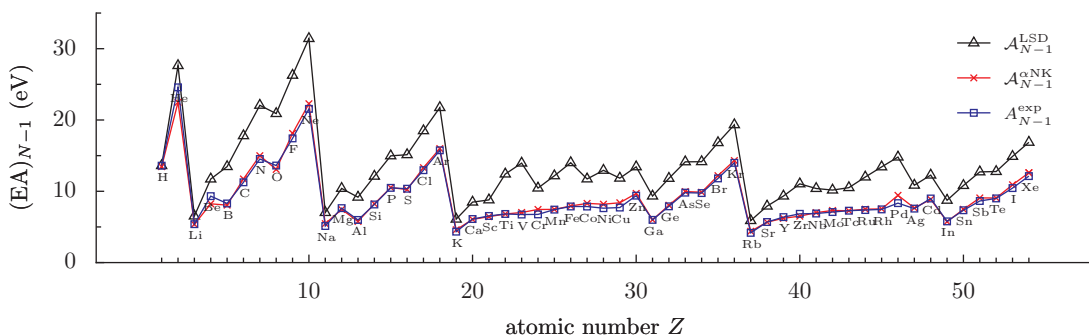
Figure 7-6: LSD and αNK differential electron affinities $\mathscr{A}_{N-1}$ (i.e., opposite energy of the lowest unoccupied orbital of the ionized atom X^+) compared with experiment for the elements of the five first periods.

method, which brings partial electron removal energies $\mathscr{A}_{N-1}$ in very close agreement with experimental total electron removal energies A_{N-1}^{exp}, whereas LSD is found to considerably overestimate $\mathscr{A}_{N-1}$. In quantitative terms, the differential LSD energy $\mathscr{A}_{N-1}^{\mathrm{LSD}}$ is overestimated by more than 4 eV with a standard deviation of 1.85 eV. Comparable deviations to LSD are obtained with the PZ self-interaction correction. The HF energy $\mathscr{A}_{N-1}^{\mathrm{HF}}$ are instead underestimated by a smaller margin of 1.48 eV.

---

[4]as opposite to the LUMO energy of the singly charged cation

	Affinities $\mathscr{A}_{N-1}$					Ionizations $\mathscr{I}_N$				
	LSD	HF	PZ	αNK	αNK$_0$	LSD	HF	PZ	αNK	αNK$_0$
MD	4.26	-1.48	4.08	0.11	-0.43	-3.62	-0.48	-0.06	0.19	-0.36
MAD	4.26	1.48	4.08	0.31	0.44	3.62	0.78	0.34	0.34	0.38
RMS	1.85	0.83	1.74	0.46	0.33	1.49	1.01	0.48	0.46	0.31

Table 7.3: Deviations from experiment of the computed differential affinities and ionizations for the first five periods (from Hydrogen to Xenon). Errors are given in eV.

The αNK correction results in substantial improvement in the calculation of partial electron removal energies, reducing the error to 0.31 eV. Here, it is quite interesting to note that the αNK variational contributions counterbalance the slight tendency of the αNK$_0$ correction to underestimate electron removal energies within LSD. A marked difference with the differential electron affinity results is the enhanced accuracy of the HF and PZ theories for differential ionizations. The improved performance in predicting atomic ionization potentials results from the fact that orbital relaxation compensates the absence of correlation contributions in HF and cancels residual non-Koopmans errors in PZ [77]. Nevertheless, even with beneficial error cancellation in favor of HF and PZ, the αNK deviation is still the lowest, approximately equal to the PZ mean absolute error of 0.34 eV. Table 7.3 summarizes the mean deviation, mean absolute deviation and root mean square deviation from experimental electron removal energies of differential affinities and ionizations for the first five periods as computed by LSD, HF, PZ, αNK and αNK$_0$.

7.3.2 Molecular Ionizations and Affinities

In this section, we focus on the study of molecular systems. For this purpose, we have implemented the HF, PZ, and αNK methods in the plane-wave pseudopotential CP (Car-Parrinello) code of the QUANTUM-ESPRESSO distribution [35]. In this code, orbital optimization proceeds via fictitious Newtonian damped electronic dynamics.

The main difficulty in the CP implementation of the HF, PZ, and αNK functionals is the correction of periodic-image errors that arise from the use of the supercell approximation [74]. Such numerical errors preclude the accurate evaluation of exchange terms and orbital electrostatic potentials. To eliminate periodic-image errors in the plane-wave evaluation of exchange and electrostatic two-electron integrals, we employ countercharge correction techniques [17]. In addition to this difficulty, explicit orthogonality constraints must be considered for the accurate calculation of the gradient of the orbital-dependent PZ and αNK functionals [37]. To incorporate these additional constraints, we use the efficient iterative orthogonalization cycle imple-

Table 7.4: LSD, HF, PZ, αNK, and αNK$_0$ differential molecular electron removal energies compared with experimental vertical electron removal energies. Mean deviations (MD), mean absolute deviations (MAD), and root mean squared deviations of the error (RMS) in absolute and relative terms are also reported. The adiabatic ionization lower bound is given when the experimental vertical ionization energy is not available. Energies are in eV.

	LSD		HF		PZ		αNK		αNK$_0$		Exp.[1]
	$\mathscr{A}_{N-1}$	$\mathscr{I}_N$	$\mathscr{A}_{N-1}$	$\mathscr{I}_N$	$\mathscr{A}_{N-1}$	$\mathscr{I}_N$	$\mathscr{A}_{N-1}$	$\mathscr{I}_N$	$\mathscr{A}_{N-1}$	$\mathscr{I}_N$	
H_2	18.84	10.16	14.78	16.24	19.01	16.99	14.79	14.83	15.74	15.69	>15.43
N_2	20.85	10.37	12.77	17.15	21.55	17.78	16.14	16.20	15.38	15.52	>15.58
O_2	18.94	7.20	9.27	14.71	18.11	15.43	13.85	13.99	13.04	12.51	12.30
P_2	14.28	7.26	8.56	10.84	14.40	11.53	10.86	10.85	10.35	10.35	10.62
S_2	13.28	5.81	7.49	10.49	13.05	11.06	10.17	10.19	9.59	9.54	9.55
PH	13.92	5.81	8.39	10.25	13.81	10.84	9.94	10.02	9.68	9.68	>10.15
HCl	17.81	8.11	10.33	13.05	17.58	13.82	13.09	13.17	12.30	12.32	>12.75
CO	18.70	9.14	11.03	15.06	18.45	15.40	14.15	14.24	13.68	13.74	14.01
CS	15.20	7.42	7.30	12.83	14.53	13.29	11.63	11.76	10.83	10.89	>11.33
H_2O	18.97	7.33	8.97	13.81	18.50	14.77	13.17	13.45	11.75	11.94	>12.62
H_2S	14.87	6.39	8.17	10.55	14.70	11.52	10.77	10.86	10.15	10.17	10.50
NH_3	16.07	6.23	7.61	11.55	15.75	12.50	11.15	11.40	10.18	10.31	10.82
PH_3	14.50	6.84	8.29	10.46	14.15	11.45	10.77	10.83	10.41	10.41	10.59
CH_4	18.69	9.46	11.95	14.92	18.98	16.20	14.46	14.51	13.91	13.86	13.60
SiH_4	15.86	8.50	10.84	13.26	16.46	14.35	12.67	12.68	12.53	12.40	12.30
C_2H_2	16.19	7.40	8.79	11.40	16.19	12.97	12.05	12.14	11.11	11.15	11.49
C_2H_4	15.12	7.01	7.91	10.43	15.07	12.62	11.43	11.53	10.50	10.54	10.68
MD	4.57	−4.35	−2.46	0.75	4.47	1.66	0.40	0.49	−0.19	−0.19	—
	38.5%	−36.4%	−21.0%	5.9%	37.4%	13.7%	3.4%	4.2%	−1.7%	−1.7%	—
MAD	4.57	4.35	2.46	0.80	4.47	1.66	0.50	0.58	0.38	0.29	—
	38.5%	36.4%	21.0%	6.4%	37.4%	13.7%	4.1%	4.8%	3.2%	2.5%	—
RMS	0.95	0.63	0.83	0.73	0.89	0.66	0.46	0.48	0.40	0.28	—
	7.7%	4.0%	7.4%	5.9%	6.3%	5.0%	3.7%	3.9%	3.3%	2.4%	—

[1]Reference http://cccbdb.nist.gov/.

mented in the original CP code [55]. In terms of computational performance, the cost of αNK calculations is here only 40% higher than that of PZ and lower than that of HF. In Table 7.4, we compare LSD, HF, PZ, and NK partial electron removal energy predictions for a representative set of molecules. In each case, molecular geometries are fully relaxed (the accuracy of equilibrium geometry predictions will be examined in Sec. 7.3.3). To perform our calculations, we employ LSD norm-conserving pseudopotentials [2] with an energy cutoff of 60 Ry for the plane-wave expansion of the electronic wavefunctions. With this calculation parameter, we verify that $\mathscr{A}_{N-1}$ and $\mathscr{I}_N$ are converged to within less than 50 meV. More details can be found in appendix B.

It is frequently argued that substituting LSD pseudopotentials for their HF, PZ, and NK counterparts has minor effect on the predicted energy differences [37]. Comparing our pseudopotential calculations with all-electron atomic results (see Sec. 7.3.1), we actually found that the use of LSD pseudopotentials yields HF, PZ, and NK electron removal energies with a typical error of 0.1 to 0.2 eV. However, since

these moderate deviations affect HF, PZ, and NK predictions in identical manner, the pseudopotential substitution does not alter the validity of the present comparative analysis.

As expected, one conspicuous feature in Table 7.4 is the poor performance of LSD that predicts molecular partial electron removal energies with an average error of ±40%. As was the case for atoms, the PZ self-interaction correction reduces the error in predicting $\mathscr{I}_N$ to less than 14%, which corresponds to an average deviation of 1.68 eV, whereas $\mathscr{A}_{N-1}$ predictions are not improved. In comparison, αNK partial ionization energies are predicted with a remarkable precision of 0.50 eV (4.1%) and 0.58 eV (4.8%) for $\mathscr{A}_{N-1}$ and $\mathscr{I}_N$, respectively. The αNK accuracy in predicting molecular vertical ionization energies compares favorably (arguably, even more accurately) with that of recently published fully self-consistent GW many-body perturbation theory calculations [84].

7.3.3 Equilibrium Molecular Structure Prediction

The accuracy of common local and semi-local density functionals in predicting equilibrium molecular structures has been one of the main success of DFT. If one wish to correct ADFs for some deficiencies, one should not impair the already established qualities of the functionals. Unfortunately, it is not easy to know ahead of time if a correction scheme will leave established properties unchanged. For example, we have seen that the Perdew-Zunger correction scheme had a substantial impact on the quality of equilibrium structures. As a test of the new methodology, we compared LSD, PZ, and αNK structural predictions to experimental bond lengths in Fig. 7-7 and we present LSD, HF, PZ, and αNK errors in table 7.5

The first important observation is the very good accuracy of LSD predictions with a mean absolute relative error of 1.1% for the seventeen molecules listed in Table 7.4. PZ bond lengths are instead sensibly underestimated with a mean uncertainty of 2.8%. In contrast with PZ calculations, αNK results deviate from experiment by a relative error margin of 0.8%, which is lower than that of LSD, demonstrating that the αNK self-interaction correction does not deteriorate and even improves LSD structural predictions, at variance with the conventional PZ self-interaction correction. These results illustrate the tendency of PZ to overbind molecular structures, and

	Bond length deviation (%)		
	MD	MAD	RMS
LSD	0.7%	1.1%	1.1%
HF	-1.4%	1.4%	0.8%
PZ	-2.8%	2.8%	0.8%
αNK	0.3%	0.8%	1.0%

Table 7.5: Deviation of predicted equilibrium bond lengths from experiment for LSD, HF, PZ and αNK for the 17 molecules in table 7.4.

confirm the systematic improvement brought about by the αNK correction.
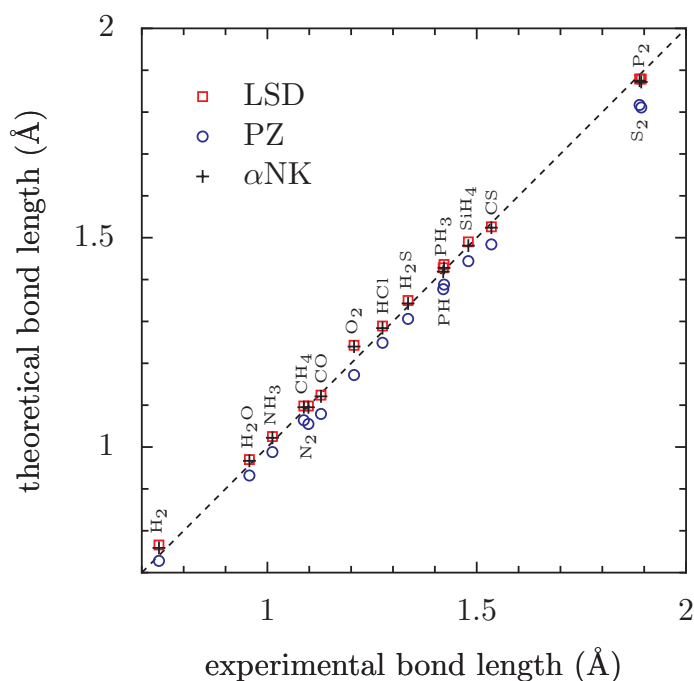
Figure 7-7: LSD, PZ, and αNK molecular bond lengths compared with experiment (Ref. CRC *Handbook of Physics and Chemistry* (2009)).

	LDA	NK	PZ
d	1.3845±0.0001Å	1.3850±0.0002Å	1.3788±0.0001Å
d'	1.3845±0.0001Å	1.3773±0.0010Å	1.3358±0.0001Å
γ	120°±0.001°	120°±0.03°	120°±0.07°
$\frac{|d-d'|}{d}$	0.0	0.005	0.031
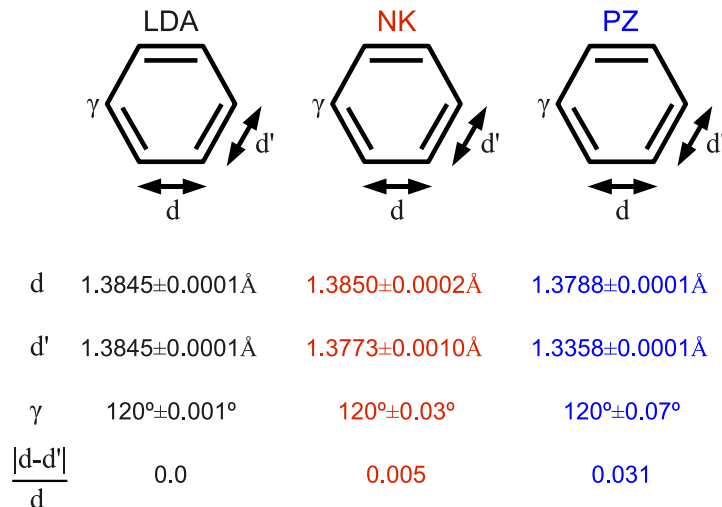
Figure 7-8: The equilibrium structure of Benzene as computed from LDA, Non-Koopman (NK) and Perdew-Zunger (PZ). γ is the average internal angle between three neighboring carbon atoms and $\frac{|d-d'|}{d}$ measures the degree of dimerization of the carbon-carbon bonds in the structure.

We also performed a high accuracy calculation on a benzene molecule to see if NK could properly deal with delocalized π systems. The result is shown on figure 7-8. We observe that, because of the greater freedom in their orbital potentials, orbital-

Table 7.6: LSD, HF, PZ, and αNK_0 orbital energies of benzene compared with experimental photoemission energies. Relative mean absolute deviations (MAD) with respect to experimental photo ionization results are also reported. Energies are in eV.

	LSD	HF	PZ	αNK	αNK_0	Exp.
e_{1g}	6.59	9.18	9.43	10.39	9.39	9.3
e_{2g}	8.28	13.54	15.46	12.66	12.48	11.8
a_{2u}	9.43	13.64	12.99	13.25	12.60	12.5
e_{1u}	10.33	16.02	17.67	14.75	14.56	14.0
b_{2u}	11.02	16.95	18.40	15.46	15.15	14.9
b_{1u}	11.26	17.51	18.82	15.65	15.69	15.5
a_{1g}	13.10	19.26	20.60	17.58	17.30	17.0
e_{2g}	14.85	22.39	22.72	19.27	19.34	19.2
MAD	26.1%	12.0%	18.4%	4.9%	2.1%	—

dependent functionals can lead to symmetry breaking, as seen by the non-zero degree of dimerization in figure 7-8. Nevertheless, even in this challenging case, NK predicts geometries in excellent agreement with LDA. PZ on the other hand, strongly favors the broken symmetry geometry (with a dimerization 6 times greater than NK one).

7.3.4 Molecular Spectroscopy

In section 7.2.1, we showed how the Non-Koopman correction was based on a construction that tries to re-establish physical meaning to orbital energies. In this section we will review some benchmark calculations made on Benzene and C_{60} in order to assess the accuracy of NK and NK_0 in predicting vertical ionization energies from orbital energies. Table 7.6 shows the computed Photoemission spectrum of Benzene compared to experiment [81] while Table 7.7 displays the same type of data for C_{60}. In those molecular photoemission calculations, we use the CP code with the computational procedure described in Sec. 7.3.2. We employ fully relaxed geometries for benzene and the LSD atomic structure of C_{60}, which is found to be in excellent agreement with the NMR experimental geometry. [32]

Focusing first on benzene, we observe that LSD underestimates electron binding energies with errors as large as 4.35 eV for low-lying states. In contrast to LSD, the HF theory provides overestimated photoemission energies with absolute deviations that increase gradually from 0.12 to 3.19 eV when approaching the bottom of the PES. Similar trends are observed for PZ with the difference that the errors do not systematically increase with increasing photoemission energies, leading in particular to the incorrect ordering of the e_{2g} and a_{2u} levels. In contrast, αNK restores the correct

Table 7.7: LSD, HF, PZ, and αNK orbital energies of fullerene C_{60} compared with constrained LSD total energy differences (Δ_cLSD) [94] and experimental photoemission energy bands [3]. Energies are in eV.

band	LSD		HF		PZ		αNK		Δ_cLSD		Exp.
I	h_u	5.84	h_u	7.49	h_u	8.77	h_u	7.45	h_u	7.61	7.60
II	g_g	7.03	h_g	9.42	g_g	9.80	g_g	8.64	g_g	8.78	8.95
	h_g	7.15	g_g	9.64	h_g	10.48	h_g	8.75	h_g	8.90	
III	h_u	8.72	g_u	12.42	g_u	12.21	h_u	10.31	h_u	10.47	10.82–11.59
	g_u	8.74	t_u	12.99	t_u	12.69	g_u	10.35	g_u	10.50	
	h_g	9.03	h_u	13.08			h_g	10.64	h_g	10.79	
	t_u	9.28					t_u	10.91	t_u	11.03	
IV	g_u	10.05	h_g	13.46	h_u	14.13	g_u	11.66	g_u	11.79	12.43–13.82
	t_g	10.52	g_u	15.06	t_g	15.41	t_g	12.12	t_g	12.28	
	h_g	10.59	t_g	15.20	h_g	15.81	h_g	12.20	h_g	12.33	
			h_g	15.66							

relative peak positions and yields slightly overestimated electron binding energies with an absolute precision of 4.9%. The slight tendency of αNK to overestimate electron binding energies is here again due to the influence of variational contributions, as directly confirmed by the performance of the αNK$_0$ orbital-energy method, which predicts photoemission energies in remarkable agreement with experiment.

Similarly to benzene, LSD energy predictions for fullerene are significantly underestimated. However, since the dispersion of the errors is much narrower than in the case of benzene, a simple shift of LSD photoemission bands, equal to the difference between the theoretical and experimental HOMO levels, can bring the predicted PES in close agreement with experiment [32]. Despite the excellent precision of HF in the top region of the spectrum, HF photoemission energies are largely overestimated for low-lying states. In addition, HF inverts the h_g and g_g states in the second photoemission band although it predicts the correct peak ordering in the third and fourth bands [94]. The performance of PZ is found to be slightly worse than that of HF with significant qualitative errors in the grouping and ordering of the states. In contrast, αNK correctly shifts the spectrum and brings photoemission energies in very good agreement with experiment. Predicted αNK binding energies are also in excellent agreement with constrained LSD total energy differences, [94] providing a further validation of the performance of the αNK correction in bringing physical meaning to orbital energies — i.e., in identifying orbital energies as opposite total

Table 7.8: αNK_0 self consistent orbital energies compared to LDA, GGA (PBE) and GW for the HOMO and HOMO-1 levels of the PTCDA molecule. We observe that αNK_0 properly predicts the C state to correspond to the HOMO-1 level consistent with GW predictions and at variance with the LDA and GGA predictions, which assign state A to be the HOMO-1 level (the A and C denominations refer to the ones in table 6.1). Energies are in eV.

Method	HOMO energy	HOMO-1 energy	HOMO-1 symmetry
GW	0.0	-1.43	C
LDA	0.0	-0.59	A,A'
GGA	0.0	-0.58	A,A'
αNK_0	0.0	-1.25	C

electron removal energies.

We finish our exploration of the predictive power of the Non-Koopman formalism in describing photoemission spectra by going back to the PTCDA molecule introduced in section 6.3.1. For this molecule we saw that the LDA and GGA functionals wrongly predict the second delocalized state (state C) to be the HOMO-5 level, whereas GW predicts it to be the HOMO-1 level. What's more, the energy of the HOMO-1 state in LDA and GGA is largely overestimated. We would like to know what the Non-Koopman correction brings about in this difficult case. For this, we performed a full self-consistent αNK_0 calculation of the PTCDA molecule with $\alpha = 0.665$ (using equation 7.24 to determine α). The results are shown in table 7.8. As can be seen from the table, αNK_0 re-establishes the C state to be the correct HOMO-1 level agreeing with the GW results, while LDA and GGA systematically predict the A and A' states to be the HOMO-1 levels. The orbital energy is also very much improved with less than 0.2 eV deviation from GW results. Another important prediction of αNK_0 is the absolute value of the HOMO level. As we saw earlier the absolute HOMO orbital energy is expected to be quite close to the exact ionization of the system and indeed it is. The predicted HOMO absolute energy is -8.5 eV whereas the actual experimental first ionization of PTCDA is found at about -8.2 eV [21]. These results are strongly supporting the interpretation of NK's orbital energies as total removal energies.

7.3.5 Band Structure of Crystalline Solids

This section will outline an application of the Non-Koopman correction scheme to fully periodic systems. In this latter case, it is often argued that the correction goes to zero because of the delocalized nature of the minimizing orbitals. Indeed if we make the assumption that orbitals delocalize all over the system, then since the single particle densities $\rho_{i\sigma}$ become locally small in value (it will typically scale as $\frac{1}{V}$ where V is the system's volume), we see that the correction terms in either Non-

Koopman or Perdew-Zunger (which look like $E_H[\rho_{i\sigma}]$ and $E_{XC}[\rho_{i\sigma}]$) will go to zero as shown on figure 6-1. But such a reasoning implicitly *assumes* delocalized minimizing orbitals $\phi_{i\sigma}$. Here we will show that empirically this is not the case.
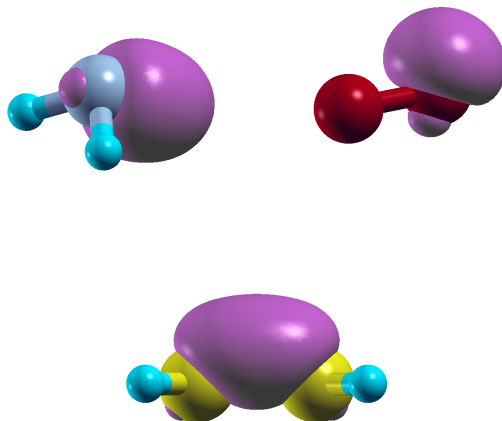

Figure 7-9: An illustration of the localized character of the minimizing orbitals in Non-Koopman. Top to bottom clockwise : a σ bond between Nitrogen and Hydrogen in Ammonia NH_3, a lone pair in di-Oxygen O_2, and a π-like bonding orbital in ethylene C_2H_4.

In all our calculations done so far, the minimizing orbitals have always been found to be localized. In fact, in accordance with Vydrov's conclusion [101] concerning Perdew-Zunger's ability to localize orbitals, we confirm that Non-Koopman's minimizing orbitals always look like Boys' localized orbitals [34] in molecular systems and Maximally Localized Wannier Functions in the case of periodic solids. To illustrate that point we show some typical minimizing orbitals of molecular and periodic systems in figures 7-9 and 7-10. NK's strength to localize orbitals is actually quite large,
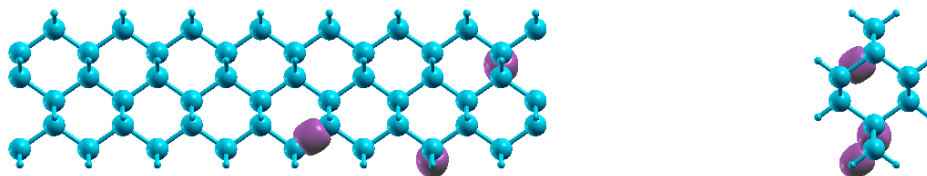

Figure 7-10: Another illustration of the localized character of the minimizing orbitals in Non-Koopman. Here a $\langle 110 \rangle$ Silicon nanowire along with some of its minimizing orbitals at NK_0 self consistency.

since localization even happens in the case of metals or semi-metals like graphene, as shown on figure 7-11. Because of the localized nature of the minimizing orbitals
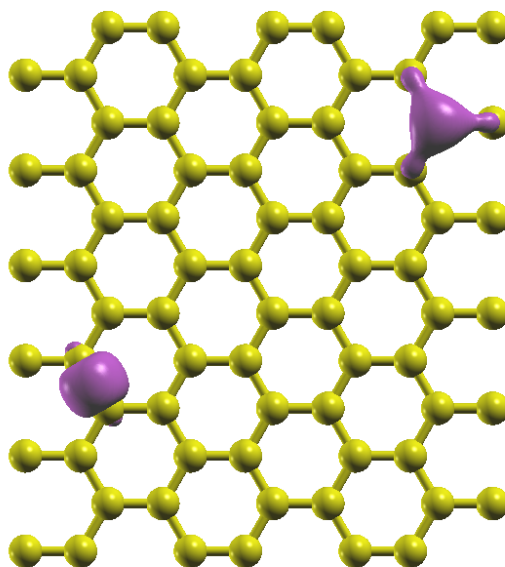
Figure 7-11: Even for metallic systems, the NK orbital dependent potential still localizes the minimizing orbitals. Here we illustrate our claim by displaying some typical σ-like and π-like minimizing orbitals of graphene.

at self-consistency, we expect the Non-Koopman correction to carry through in the case of crystalline solids.

Owing to the good overall prediction of photoemission spectra by NK and NK_0, we decided to investigate the effects of those corrections onto the band structure of crystalline solids. The computational procedure consisted first in introducing a supercell of crystalline solid (just like in figure 7-11), and compute the NK or NK_0 ground state. In a second step, the minimizing orbitals (and sometimes some of the empty states) were unitarily transformed into an equivalent basis of Maximally Localized Wannier Functions. From there, we used the computational procedure introduced in section 5.3 to extract the hopping parameters and compute the band structure from a Wannier function interpolation. Even though we proved in section 5.3 that the "supercell method" works for computing band structures, we wish to present a test case, where we can observe the rate of convergence of the band structure with the system size. In figure 7-12 we show the convergence of the valence bands of *graphane* (a modern day most important 2D material for potential nanoelectronic applications [25]).

The most striking feature of figure 7-12 is the relatively small supercell size required to converge the bands. In quantitative terms, the deviation in band energies of the 6×6 supercell with respect to the 8×8 supercell has a maximum value of 0.09 eV, a minimum value of -0.01 eV, a mean value of 0.04 eV and a standard deviation of 0.018 eV. This demonstrates that a 6×6 supercell is enough to converge band energies to about 40 meV. All the calculated band structures in this thesis have been checked
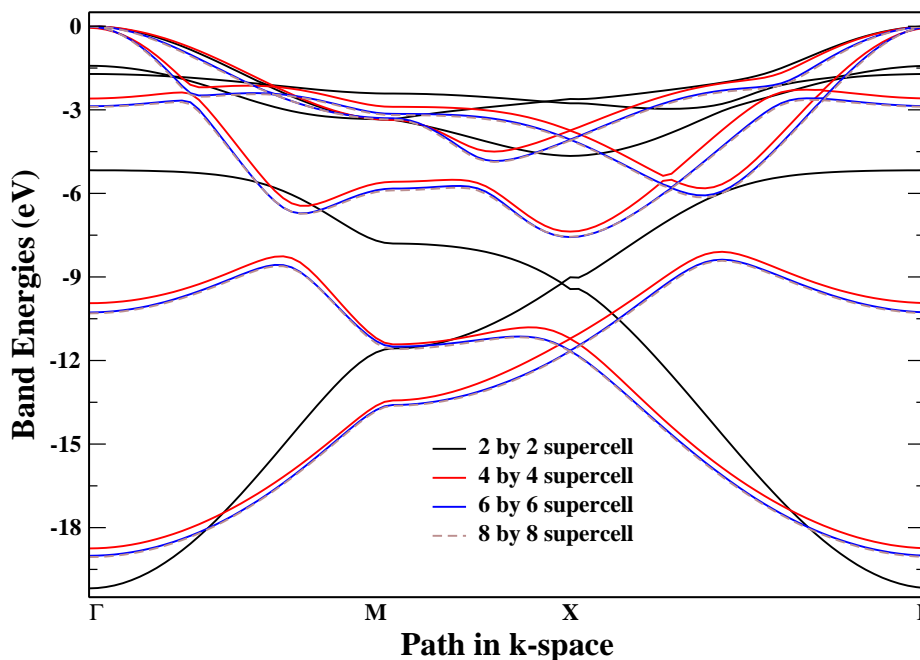
Graphane Valence Bands Convergence

Figure 7-12: Convergence of Graphane's valence bands with supercell size as computed by NK_0. The fundamental building block for the supercell consisted in a cubic cell containing two of Graphane's elementary cell.

against supercell size convergence with similar accuracy reached for the band energies. What's more, because NK predicts equilibrium geometries very close to the LDA ones (see section 7.3.3), we fully relaxed all our crystal structures at the LDA level.

7.3.5.1   1D periodic systems

As a prototypical example of a one dimensional periodic solid, we chose to study the all-trans polyethylene polymer. The main reason for choosing that system had to do with the availability of ARPES[5] measurements for the valence bands and a good knowledge of the experimental band gap. Figure 7-13 displays the calculated valence bands of all-trans polyethylene as computed by LDA, NK, NK_0 and PZ. We also add on top of the plots the experimental data from reference [88].

Despite the fact that the experimental uncertainty is at times quite large (up to 1eV), we see that the best agreement with experiment is realized for the NK_0 and PZ methods. It is quite satisfying to see that NK_0 improves upon valence bandwidths and more generally shifts the bands in the right direction with respect to experiment.

---

[5]Angle Resolved PhotoEmission Spectroscopy is a delicate but powerful experimental tool to investigate the band structure of periodic solids for fully occupied bands.
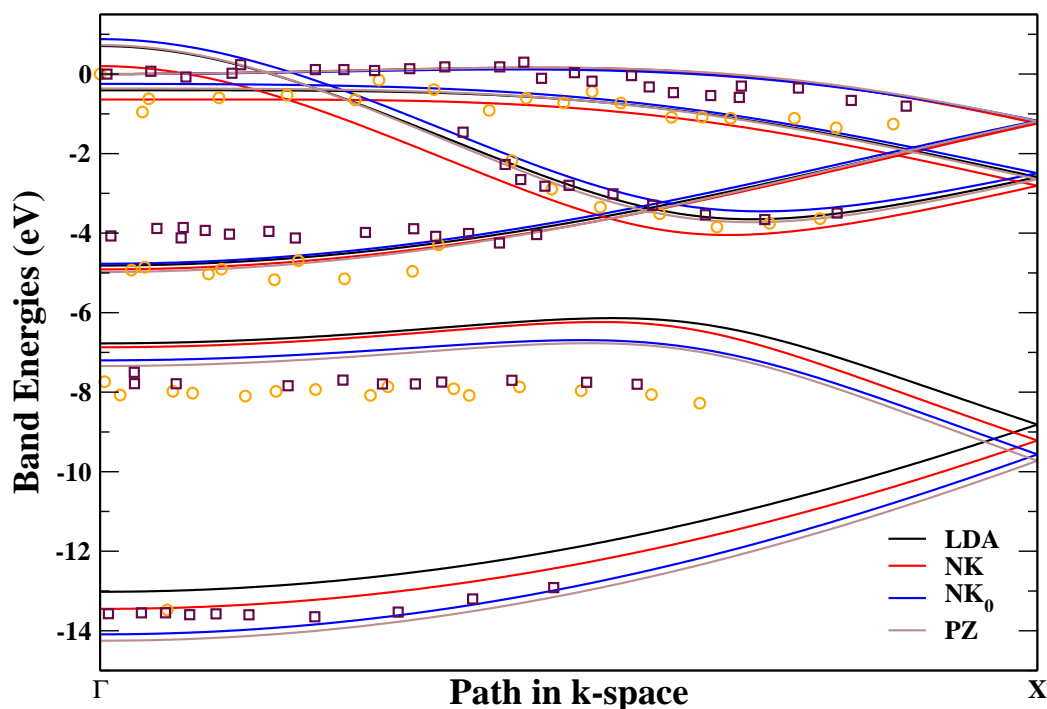
Polyethylene Valence Bands

Figure 7-13: Computed valence bands of all-trans polyethylene. Four different methods were used : LDA, NK, NK_0 and PZ. The experimental data (squares and circles) are from Ref. [88].

In particular, NK_0 preserves the LDA good description of the upper bands whereas it improves the description of the lower bands, where LDA is too high in energy. Besides the valence band structure, we computed the band gap (which lies at the Γ point) for the same methods. The result is shown in table 7.9. We observe a large

Table 7.9: Calculated Band Gap of polyethylene for LDA, NK, NK_0 and PZ compared to experiment. Values are in eV

Method	Band Gap (eV)
LDA	6.13
NK	15.12
NK_0	13.18
PZ	12.88
Exp.	8.8

systematic overcorrection of the band gap with every orbital-dependent method. This overcorrection comes mostly from a lowering of the valence bands compared to the vacuum level, because empty states are almost left untouched by those methods. Des-

pite the large overcorrection, it is important to note that introducing the α parameter will certainly improve on the description of the band gap because the LDA band gap is lower than the experimental one, while the NK and NK_0 are systematically larger. We give in table 7.10 the values of the calculated gap in αNK and αNK_0. The optimal α value for aligning the computed band gap and the experimental one are respectively $\alpha^{NK} \approx 0.30$ for NK and $\alpha^{NK_0} \approx 0.38$ for NK_0.

Table 7.10: Calculated Band Gap of polyethylene for different values of the α parameter in NK and NK_0 corrections. Values are in eV

NK		NK_0	
α	Band Gap (eV)	α	Band Gap (eV)
0.00	6.13	0.00	6.13
0.05	6.58	0.05	6.48
0.10	7.04	0.10	6.84
0.15	7.49	0.15	7.19
0.20	7.95	0.20	7.54
0.25	8.40	0.25	7.90
0.30	8.85	0.30	8.25
0.35	9.30	0.35	8.60
0.40	9.75	0.40	8.95
0.45	10.20	0.45	9.30
0.50	10.65	0.50	9.66
0.55	11.10	0.55	10.01
0.60	11.55	0.60	10.36
0.65	11.98	0.65	10.71
0.70	12.45	0.70	11.05
0.75	12.90	0.75	11.49
0.80	13.34	0.80	11.85
0.85	13.79	0.85	12.18
0.90	14.23	0.90	12.45
0.95	14.68	0.95	12.80
1.00	15.12	1.00	13.18

As the next stage of complexity, we studied the valence bands of a quasi one-dimensional system, namely a $\langle 110 \rangle$ silicon nanowire. The computed band structure is shown on figure 7-14. Similarly to the case of polyethylene, the Non-Koopman correction introduces a non-uniform scaling of the valence manifold which leads to larger bandwidths.

We note in passing that a proper description of the bands seem to require the use of *electrostatic correction* techniques. We demonstrate this by showing side-by-side the computed band structures at the LDA and NK_0 level with and without electrostatic corrections. Figure 7-15 summarizes the findings. As evident from the plots, noticeable changes in the bands topology happens both around -3 to -5 eV and -7 to -9 eV. Specifically, the location of the band crossings in the Brillouin Zone are clearly
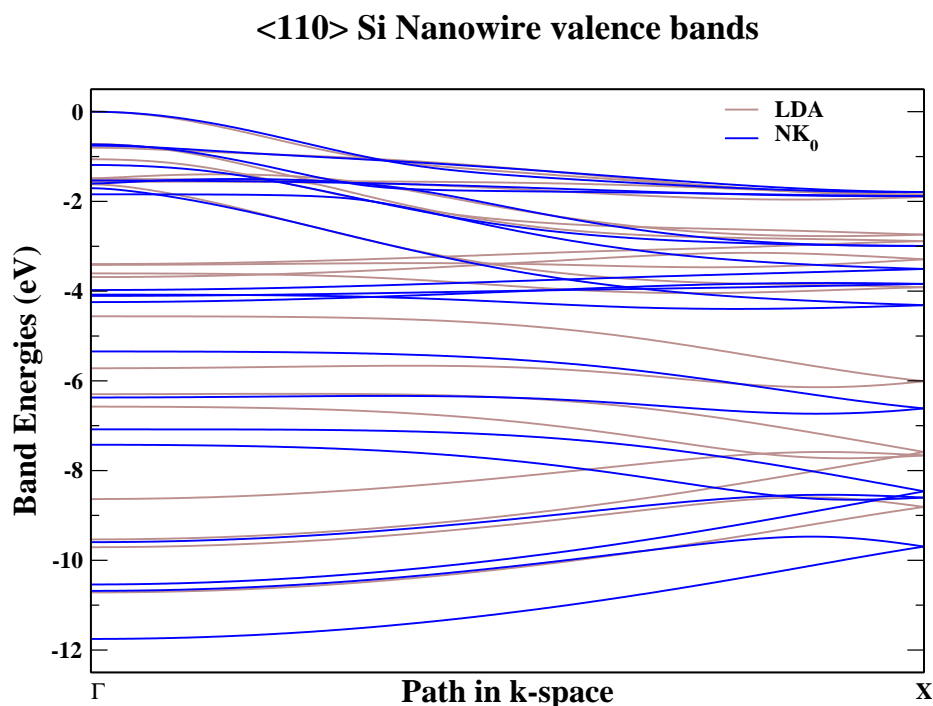
\<110\> Si Nanowire valence bands

Figure 7-14: Computed valence bands of the smallest ⟨110⟩ silicon nanowire (diameter of about 7.8 Å). Two different methods were used : LDA, and NK_0. Similarly to the case of polyethylene, we can observe an increase in the valence bandwidth which results from a non-homogeneous scaling of the bands.

displaced.

7.3.5.2 A 2D periodic system

Increasing the dimensionality of the system, we decided to focus our attention onto a promising two dimensional material for nanoelectronic applications, namely graphane. We computed the valence bands of graphane, and systematically compared our predicted valence manifold to recent GW calculations [57]. The results are shown in figure 7-16. We observe an overall excellent improvement of the bands with a particularly good description of the lower levels. NK_0 also predicts a larger bandwidth but apparently does not correct much of the higher energy bands. As regard to the Band Gap, NK_0 once again overestimates it with a value of 8.6 eV, while LDA underestimates it with a value of 3.6 eV. The reference Gap is taken to be the GW result from ref [57] with a 5.4 eV direct band Gap.

7.3.5.3 A 3D periodic system

As a final test of the accuracy of NK's band structure prediction, we computed the full valence manifold of bulk diamond and summarized our findings on figure 7-17. The
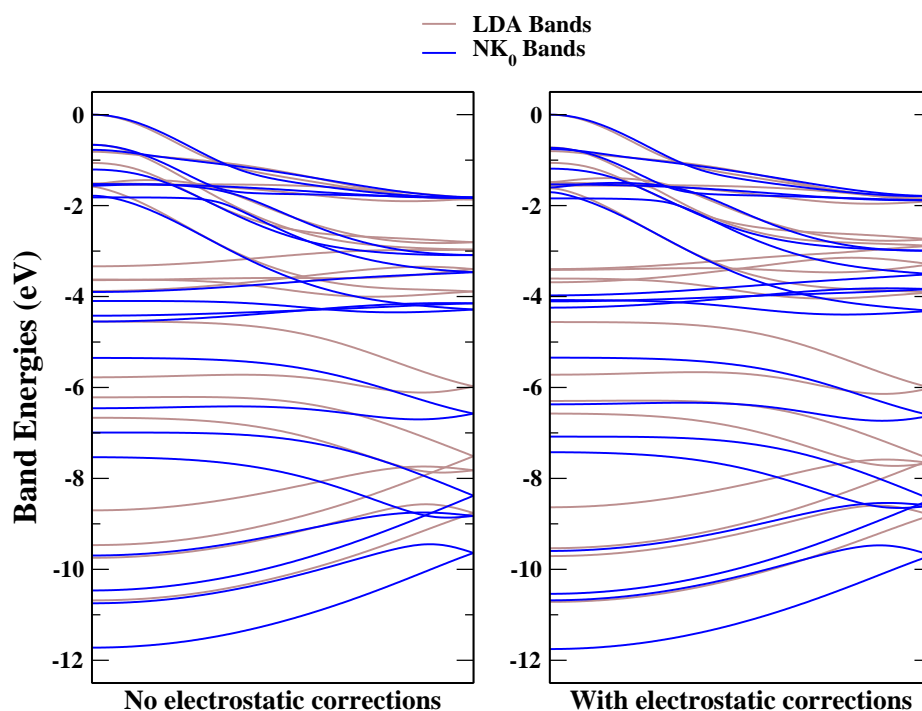
Figure 7-15: Computed valence bands of the smallest $\langle 110 \rangle$ silicon nanowire (diameter of about 7.8 Å). Two different methods were used : LDA, and NK_0. As can clearly be seen from the plots, the use of electrostatic corrections is critical when it comes to band alignments and band crossing points. Important change in bandwidth can be seen around -3 to -5 eV.

figure also displays the experimentally determined valence bandwidth [44]. What we see is yet another enlargement of the bandwidth by NK_0 so as to almost match experiment.

The Gap at Γ[6] was determined and compared to experiment and GW calculations. The results are shown in table 7.11.

As a conclusion we can generally say that in the case of semi-conducting and insulating materials, it seems that Non-Koopman corrections have mainly two effects. The first is to non-homogeneously stretch the bands and so increase the bandwidth of the valence manifold, typically improving the agreement with experimentally determined bandwidths. The second major effect is to overcorrect the band gap. A way to alleviate the overcorrection would be to determine the proper self-consistent α parameter in the Non-Koopman scheme. Since LDA always underestimates the band gap and NK always overestimates it, there should be an α for which the agreement with experiment is ensured. The obvious problem with such an α is the arbitrariness of

---

[6]This Gap does not correspond to the minimum gap of 5.48 eV of diamond, which sits elsewhere in the Brillouin Zone.
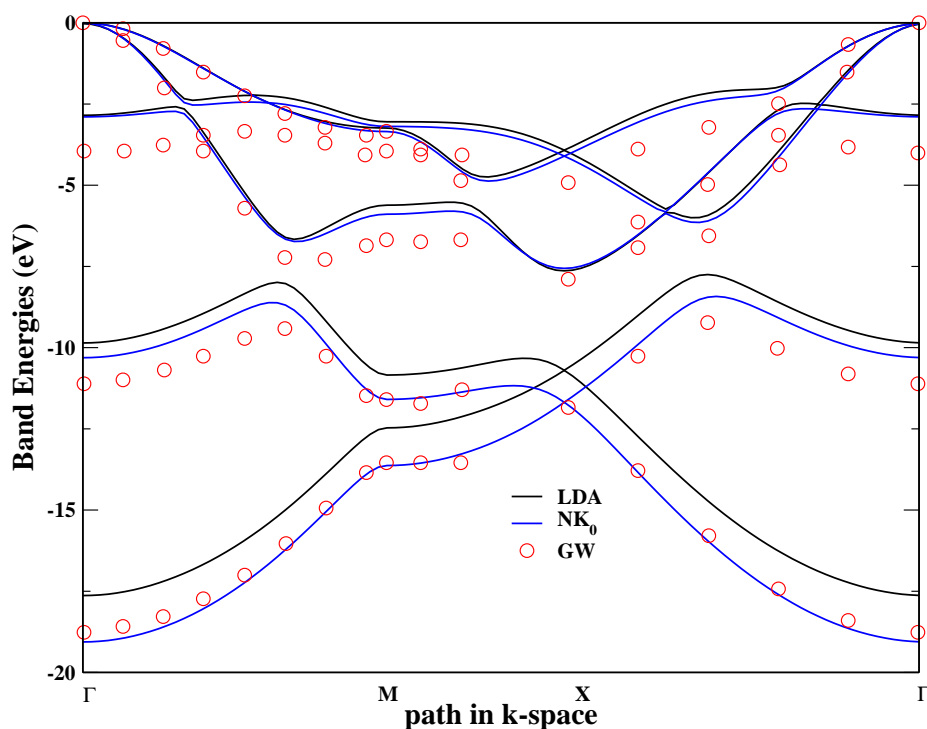
Figure 7-16: Computed NK_0 valence bands of graphane compared to LDA and GW from ref [57]. We observe a substantial improvement of the description all throughout the Brillouin Zone with, in particular, a correct bandwidth. For the higher bands, the NK_0 correction is somehow too weak.

the choice, contrary to the self-consistent α that can be determined on some physical ground (see section 7.2.5). What's more, reducing the value of α from 1 (in the NK and NK_0 case) to some number between 0 and 1 can only decrease the bandwidth, such that a balance needs to be reached between the accuracy of the bandwidth and the one of the band gap.

Table 7.11: Calculated Band Gap at Γ and valence bandwidth of diamond for LDA, NK_0 and GW compared to experiment. Exp. is from [44] and GW from [42]. Values are in eV.

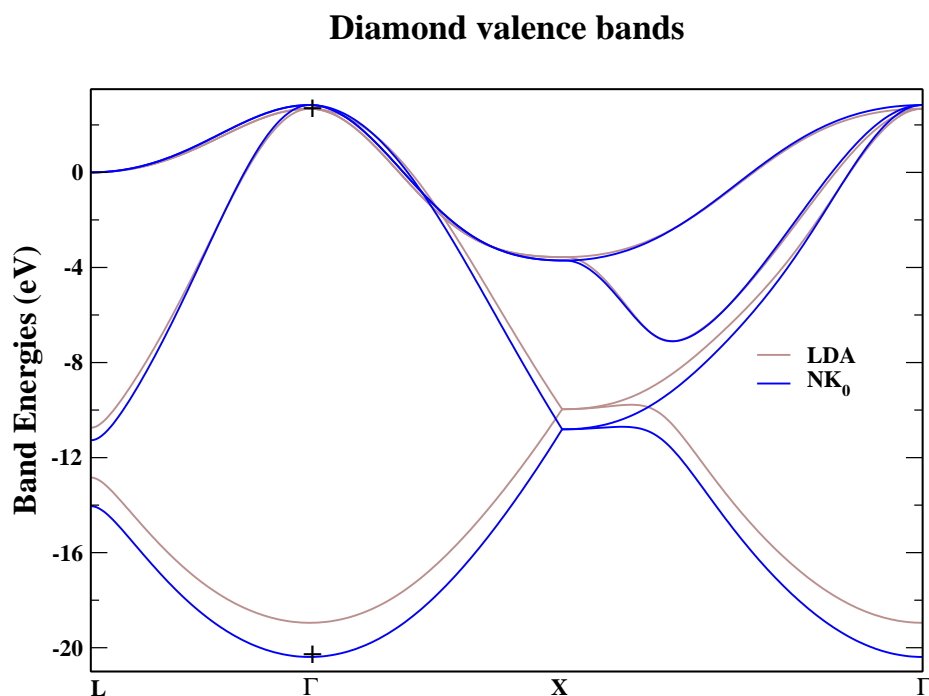| Method	Band Gap at Γ (eV)	Valence Bandwidth (eV)
LDA	4.12	21.63
NK_0	9.94	23.22
GW	7.38	23.0
Exp.	7.3	23.0

Diamond valence bands

Figure 7-17: Computed valence bands of diamond in LDA, NK and NK_0. As usual the NK_0 correction increases the valence bandwidth so as to almost match experiment (as depicted by the two crosses from Ref. [44]).

7.3.6   Dissociations of Homonuclear Diatomic Cations

In continuing our assessment of the predictive power of the NK scheme, we now enter a generally dangerous territory for DFT, namely the dissociation of homonuclear diatomic cations. The reasons for the difficulty of the theory to properly treat those systems have to do with the fact those dissociation processes involve sensitive charge transfers as the nuclei are separated from each other and a competition between ionic and covalent ground states [38]. But let us jump in and compute the potential energy curves of two diatomic cations, H_2^+ and He_2^+. Figure 7-18 displays the potential energy curves of H_2^+ and He_2^+ as computed from LSD, NK, PZ and NKPZ (Non-Koopmans corrections on top of the Perdew-Zunger functional). We also give the exact potential energy curves as computed by us for the H_2^+ case (see appendix D) and extracted from Ref. [47] for the He_2^+ case. As we can see, LSD but also NK are quite off from the exact result. Not only the equilibrium separation is wrong (LSD predicts 1.158Å for H_2^+ and 1.192Å for He_2^+, while NK predicts 1.134Å for H_2^+ and 1.189Å for He_2^+ compared to exact values of 1.055Å for H_2^+ and 1.081Å for He_2^+), but the dissociation energy (i.e. depth of the potential well) also is too small.
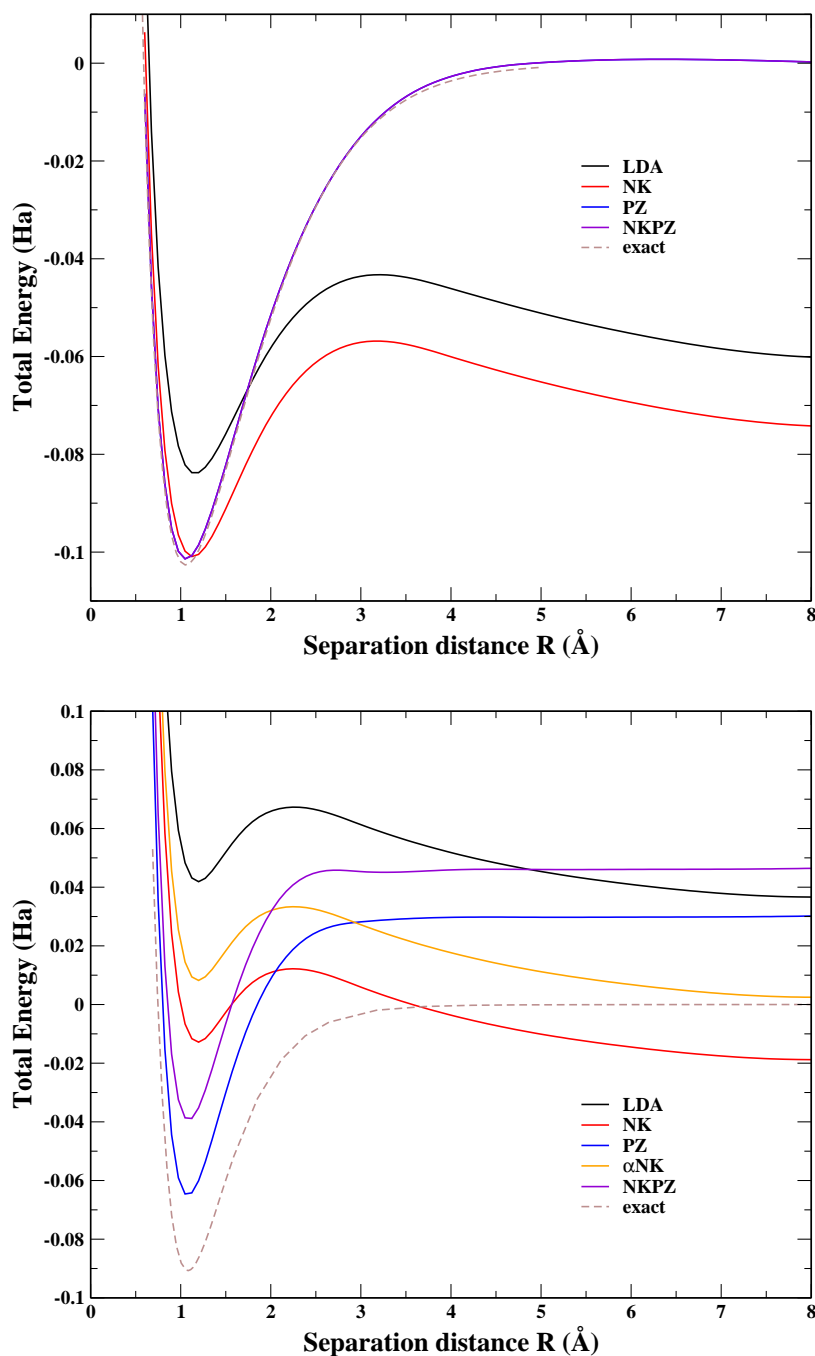
Figure 7-18: Calculated potential energy curves for the dissociations of H_2^+ and He_2^+ from LSD, NK, PZ and NKPZ. The energies are given in Hartrees and are "absolute" energies with a zero reference corresponding to the exact energy of the system in the infinite dissociation limit. Those graphs were obtained by computing the total energy at different separation distances R in $D_{\infty h}$ symmetry. In the region 2-3 Å, the computed energies were wrongly overestimated as shown in [14], and so we used a spline interpolation of the curves.

In the case of He_2^+, the cation is even predicted to be metastable in LSD and NK ! In contrast, the PZ and NKPZ are clearly predicting both the equilibrium separation and the dissociation energies (i.e. the depth of the potential wells). Is this surprising at all ? Actually it is not. Indeed, those two systems, are really one-electron systems in disguise. In the case of H_2^+ it is literally true, while for He_2^+, two of the electrons are fully saturating the 1s orbital, with a remaining 2s electron free to move around. This fact explains the excellent accuracy of PZ which is a correction designed to work exactly for one electron systems. We also observe that if we impose Non-Koopman corrections on top of Perdew-Zunger, the predictions are still excellent. This is a very strong validation of the non-destructive character of the Non-Koopman methodology. Whenever a functional gets a limit right, NK keeps most of the accuracy intact, while correcting the functional in other limits for which the NK methodology has been designed to work.

Despite the impressive accuracy of PZ and NKPZ to predict a proper dissociation curve, we prove here, that not *everything* is perfectly predicted. For this we look at the total energies of split charge configurations, for which the exact theory predicts a constant energy. By split charge we mean to split the positive charge in fractions q and $1-q$ on each atom in the infinite separation limit. If q is the positive charge on one of the atoms, then $1-q$ should be the net charge on the remaining atom. In the infinite separation limit, the total energy of the system is then $E(A^{+q}) + E(A^{1-q})$. Quantum Mechanics predicts that since the $A\ldots A^+$ and $A^+ \ldots A$ states are degenerate ground states, then any linear combination of those states are also ground states with the same total energy. This means that all of the above mentioned split charge states with a charge q on one atom and $1-q$ on the other should also have the same energy. By this analysis, we expect the following result

$$E(A^{+q}) + E(A^{1-q}) - \left(E(A) + E(A^+)\right) = 0 \tag{7.26}$$

To verify that claim, we calculated the energies in equation 7.26 for Hydrogen and Helium at the LSD, NK, PZ and NKPZ levels. The results are shown on figure 7-19. The most striking feature of those graphs, is the systematic overstabilization of the 1/2 split charge configuration with respect to any other. This is very counterintuitive especially in the PZ and NKPZ cases. Indeed we saw earlier than the dissociation limit was right. But here, we see that even those methods, wrongly favor the 1/2 split charge configuration compared to the A...A$^+$ configuration. This just proves that even though the description of systems with *integer* numbers of electrons can be right, the systems with fractional number of electrons are imperfectly described. This tendency of any functional to overstabilize or over-destabilize (in the Hartree-Fock case for example not shown here) split-charge configurations comes from the non-linear behavior of the total energy of the system with the electron number.

As a consolation, we observe that orbital dependent corrections reduce by a factor of 3 to 4 the unphysical overstabilization of split charge configurations.
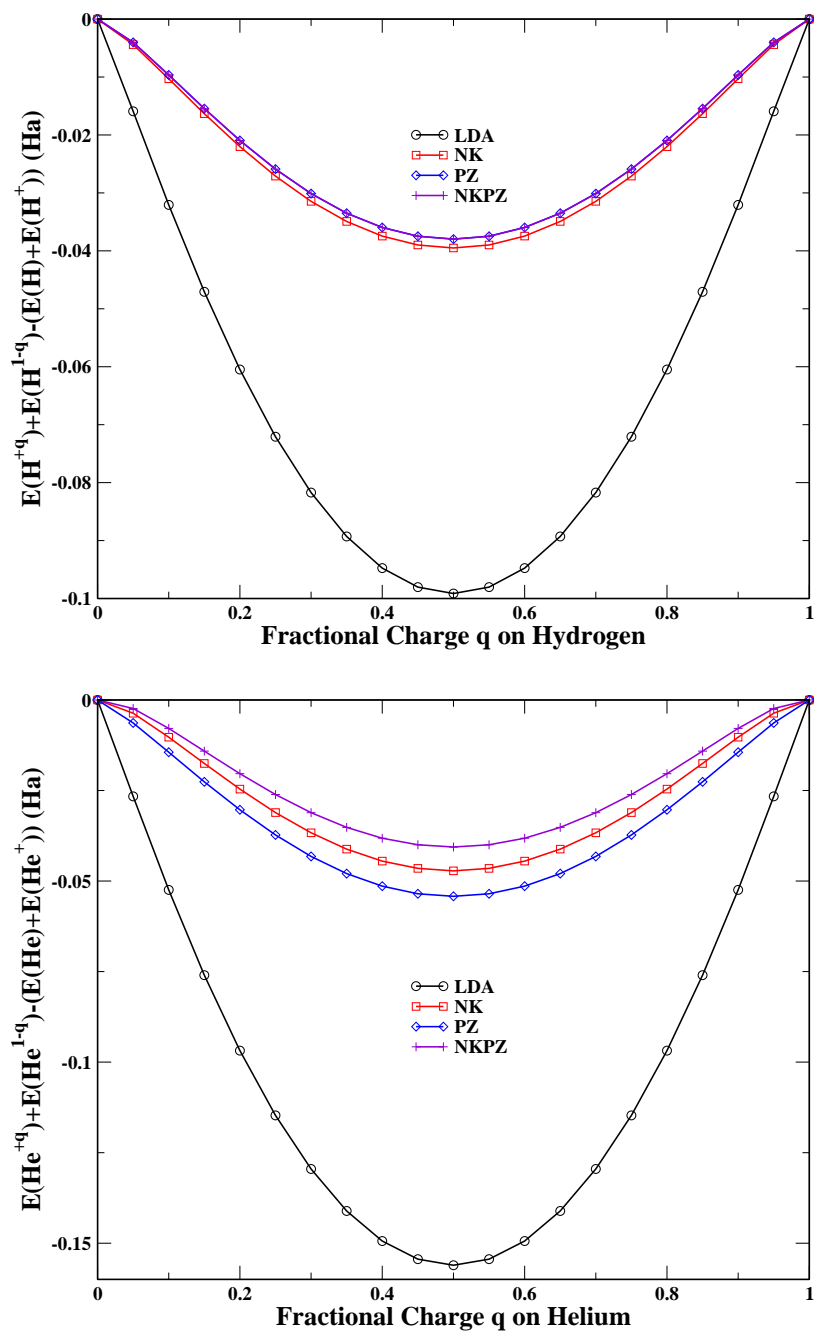
Figure 7-19: Test of the validity of equation 7.26 at the LSD, NK, PZ and NKPZ levels. We observe that none of the functionals satisfy the exact condition of equation 7.26, and that split charge configurations with fractional charges q and $1 - q$ are always systematically overstabilized.

7.3.7 Description of Reaction Barriers

In this final section, we wish to compare the predictive power of LSD, NK, PZ and NKPZ on estimating the forward and reverse reaction barriers of a chemical reaction. We choose to study the following chemical reaction

$$H + SH_2 \leftrightharpoons H_2 + SH \qquad (7.27)$$

Very accurate high-level quantum chemistry data are available for that reaction in terms of equilibrium structures, structure of the activated complex and total energies of the reactants and products [64]. The total energies of the different molecules were computed so as to extract the forward and reverse reaction barriers. The results are summarized in table 7.12.

Table 7.12: Calculated forward and reverse reaction barriers for the chemical reaction of equation 7.27 at the LSD, NK, PZ and NKPZ levels. Accurate Coupled Cluster CCSD(T) calculations of reference [64] are also shown. Values are in eV.

	Forward Barrier (eV)	Reverse Barrier (eV)
LDA	-0.247	0.002
NK	-0.005	0.257
PZ	0.104	0.444
NKPZ	0.269	0.581
CCSD(T)	0.156	0.790

Once again, we observe that LSD and NK are not accurate at all for these delicate charge transfer reactions. This is mostly due to the activated complex, which, as a transition state, lies in a regime of atomic separation where LSD and NK are inaccurate (see the dissociation plots of figure 7-18). At variance with those, PZ correctly predicts the activated complex to be of higher energy than the reactants and products. What's more, the products are also predicted to be more stable than the reactants. NKPZ even improves upon PZ's description and brings the reaction barriers to about 0.1-0.2 eV of the practically exact results of CCSD(T). This provides further support to the conclusion that the NK methodology improves already established properties of approximate functionals.

Part IV

Quantum Transport

8

Basics of Coherent Transport

In this chapter, we will present a somehow thorough introduction to the Landauer-Buttiker formalism for coherent transport. In a first section, our focus will be to outline some fundamental results in one-body Green's function theory, illustrating the power of the Green's function approach. We will also introduce a general expression for the current operator in a localized basis representation, allowing us to compute the electron flow in the system. In a second section, we will take advantage of the general setup for Lead-Conductor-Lead (or two-probe) systems – typical in quantum transport experiments – in order to derive a very general expression for the electron current. This relation, often called the Fisher-Lee formula [33], is of central importance in the whole field of mesoscopic transport. The Fisher-Lee formula gives a constructive approach to the calculation of the transmission function of the system, and leads naturally to the Landauer formula. In appendix C, we propose an alternative way of deriving the Landauer formula, based on a simple real-space approach. Unfortunately, that latter approach does not tell us what the transmission function is, and this is why the Fisher-Lee formula is superior in that respect. But before to jump in, let us speak a little about the validity range of coherent transport and when this regime matters in practice.

8.1 Coherent Transport and Density Functional Theory

The purpose of this short introduction is to give the reader an idea of what coherent transport is. We will not dwell in much details onto the different transport regimes and we refer to excellent reviews like [54] and chapter 3 of [90]. Coherent transport is characterized by the existence of a "phase coherence length", L_ϕ, which gives the average distance between two *inelastic* scattering events. An inelastic scattering event is characterized by a *change* of the electron's energy, E, as it interacts with another particle (electron or phonon for example). The existence of a phase coherent regime has been confirmed experimentally by direct observation of a quantized conductance in semiconducting and metallic samples of nanometric dimensions [98]. Typical experimental coherence lengths range from 39 nm @ 4K to 136 nm @ 50mK in 2D Silicon samples [85] and up to 3 to 5 μm in graphene @ 260mK [69]. A theoretical treatment of coherent transport must then try to deal with such spatial dimensions

for the systems of interest while keeping a full quantum mechanical description of electrons in order to describe wave interference effects. What's more, in the coherent transport regime, the electron's energy is conserved and so is a good quantum number. This tells us that the electron's wavefunction must then be an eigenstate of the system's Hamiltonian. In the following chapters, we will present an attempt to compute the eigenstates of the system using Density Functional Theory, and then use the Landauer-Buttiker formalism to compute the electronic current flowing through the system.

There has much debate as to whether static Density Functional Theory is in a position to properly describe transport phenomena like coherent transport. A somehow comprehensive discussion is presented in the book by Di Ventra [20]. More recently, it has been suggested that as long as the leads transporting the electrons from the reservoirs to the central conductor[1] support a single conducting channel, then exact static Kohn-Sham DFT is predicted to be quite accurate [67]. Notwithstanding all those concerns, we will use DFT to compute the real-space Hamiltonian matrix of typical Lead-Conductor-Lead systems, and then use that matrix along with the formalism about to be described, in order to compute the current intensity. The typical work flow for a DFT-Landauer calculation is shown on figure 8-1. In the following,
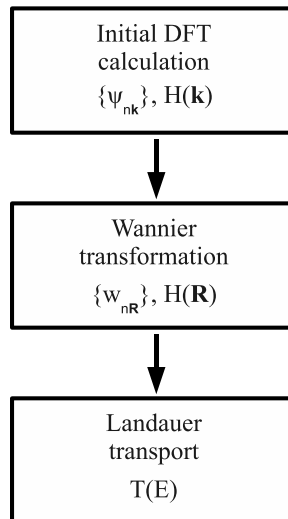

Figure 8-1: A flow diagram depicting the key steps in a typical DFT-Landauer Quantum Conductance calculation. We start with an *ab-initio* DFT calculation for the Ground state of the system. Then we perform a unitary transformation so as to construct the Wannier functions and express the Hamiltonian matrix in this basis (matrix $H(\mathbf{R})$). In a final step, we use the real-space Hamiltonian matrix to compute the Transmission function T(E) of the system, which is the central quantity in Landauer transport.

we will assume that the first two steps in figure 8-1 have been performed and so we end up with a matrix H expressed in a localized basis of Wannier functions.

---

[1]we will give a precise meaning to all those terms shortly

8.2 A Primer on Green's Functions

Before to move on to the calculation of electron current, we first need to digress to a short review of Green's function theory applied to one-body quantum systems. We will mostly follow Economou [23] and present an abstract representation of Green's functions which, as such, will be independent of the basis in which operators and vectors are expressed (e.g. a localized Wannier basis or delocalized Bloch basis). For this, we will make heavy use of Dirac's "bra"-"ket" notation [2].

8.2.1 General Linear Partial Differential Equation

Green's functions are fundamentally bound to partial differential equations. Those "functions" – or operators as we will see – are defined to be the solution of *linear* partial differential equation with an *elementary* (or "impulse") source term $\delta(\vec{x} - \vec{x}')$. We then define a linear operator that we will assume *hermitian*[3], $H(\vec{x})$, and write

$$(z - H(\vec{x}))G(\vec{x}, \vec{x}', z) = \delta(\vec{x} - \vec{x}') \tag{8.1}$$

In the above equation, z is a complex number ($z \in \mathbb{C}$), and the partial differential equation is solved over a definite volume Ω with clearly defined boundary condition on $\partial\Omega$. The Green's function is assumed to satisfy those boundary conditions. Since $H(\vec{x})$ is taken hermitian, we know that there exists a complete and orthonormal set of eigenvectors in which any function defined over the volume Ω can be expanded. We will denote by $\phi_n(\vec{x})$ or simply ϕ_n, those eigenfunctions and by λ_n the associated eigenvalue. We then have

$$H(\vec{x})\phi_n(\vec{x}) = \lambda_n \phi_n(\vec{x}) \tag{8.2}$$

The spectrum of $H(\vec{x})$ can be discrete, continuous or both. In the continuous case, λ_n will vary in a given interval $[\lambda_{\min}, \lambda_{\max}]$.

8.2.2 Abstract Green's Functions

Let us now introduce Dirac's notation. In Dirac's bracket language, an eigenfunction $\phi_n(\vec{x})$ is nothing else than the scalar product of the abstract vector $|\phi_n\rangle$, or "ket", with the linear form $\langle\vec{x}|$, or "bra"

$$\phi_n(\vec{x}) = \langle\vec{x}|\phi_n\rangle$$

---

[2] A thorough introduction to Dirac's notation can be found in Cohen-Tannoudji [12].

[3] we will only treat stationary systems here since the time-dependent Green's functions can be computed from the knowledge of the stationary Green's functions. What's more, we are only interested in steady-state currents so we only need the stationary Schrödinger equation.

bras are hermitian conjugates to kets, so $\left(|\phi_n\rangle\right)^\dagger = \langle\phi_n|$, where $\dagger$ denotes hermitian conjugation. In the same spirit we have the following properties

$$
\begin{aligned}
\int_\Omega |\vec{x}\rangle\langle\vec{x}|\,d\vec{x} &= \mathbb{1} \\
\sum_n |\phi_n\rangle\langle\phi_n| &= \mathbb{1} \\
\langle\phi_n|\phi_m\rangle &= \delta_{nm} \\
\langle\vec{x}|\vec{x}'\rangle &= \delta(\vec{x}-\vec{x}') \\
G(\vec{x},\vec{x}',z) &= \langle\vec{x}|G(z)|\vec{x}'\rangle \\
H(\vec{x}) &= \delta(\vec{x}-\vec{x}')\langle\vec{x}|H|\vec{x}'\rangle
\end{aligned}
$$

The first two equations express the completeness relation of basis $\{|\vec{x}\rangle\}$ and $\{|\phi_n\rangle\}$, while the third and fourth are the orthonormality relations of those basis. The last two equations tell us that $G(z)$ and H are operators and that $G(\vec{x},\vec{x}',z)$ and $H(\vec{x})$ are nothing else than their matrix elements in the $\{|\vec{x}\rangle\}$ basis. We see in particular that the operator H is diagonal in the $\{|\vec{x}\rangle\}$ basis. Using that notation we can easily arrive at the operator equation

$$(z-H)G(z) = \mathbb{1} \tag{8.3}$$

Expressing the Green operator $G(z)$ in the eigenfunction basis $\{|\phi_n\rangle\}$ as

$$G(z) = \sum_{nm} \alpha_{nm}|\phi_n\rangle\langle\phi_m|$$

and injecting into equation 8.3, remembering that, by definition of the eigenfunction basis $H = \sum_n \lambda_n|\phi_n\rangle\langle\phi_n|$, we easily arrive at

$$G(z) = \sum_n \frac{1}{z-\lambda_n}|\phi_n\rangle\langle\phi_n| \tag{8.4}$$

In particular, we can go back to the $\{|\vec{x}\rangle\}$ basis by computing the matrix element of $G(z)$ between $\langle\vec{x}|$ and $|\vec{x}'\rangle$ to find

$$G(\vec{x},\vec{x}',z) = \langle\vec{x}|G(z)|\vec{x}'\rangle = \sum_n \frac{1}{z-\lambda_n}\langle\vec{x}|\phi_n\rangle\langle\phi_n|\vec{x}'\rangle = \sum_n \frac{1}{z-\lambda_n}\phi_n(\vec{x})\phi_n^*(\vec{x}')$$

8.2.3 Some General Properties of Green's Functions

Looking at equation 8.3, we immediately see that $G(z)$ is in general an analytic function of z for all z with a non-zero imaginary part. This is just a consequence of the fact that λ_n is real for all n because H is hermitian. For the *discrete* part of the spectrum, we see that $G(z)$ has a simple pole at $z = \lambda_n$, which implies that $G(z)$ is not defined at those points. We realize that the knowledge of the eigenfunctions and eigenvalues of H leads to a complete knowledge of $G(z)$. Vice versa, a knowledge of $G(z)$ allows us to find the discrete eigenvalues by using a most important theorem of complex analysis known as the Residue Theorem. What's more, the interval of continuous

eigenvalues form a branch cut for $G(z)$, which means that we also have access to the continuous eigenvalues through the knowledge of the branch cut of $G(z)$. Just like the case of the discrete eigenvalues we see that $G(z)$ is, *a priori*, not well defined for $z = \lambda_n$ in the continuous spectrum. In general though, if the continuous spectrum is associated with **delocalized** states, like Bloch states for periodic systems, then one can define two limits of most important value as we will see later. Namely one can take the limit of $G(z)$ when z is approached from above the real axis or from below and define respectively the *retarded* and *advanced* Green's functions

$$G^{\pm}(\lambda) = \lim_{s \to 0^+} G(\lambda \pm is) \qquad (8.5)$$

Those two functions are actually not independent from each other. This comes from the fact that $G^{\dagger}(z) = G(z^*)$, as can be easily verified on equation 8.4. With this we see that

$$G^-(\lambda) = \lim_{s \to 0^+} G(\lambda - is) = \lim_{s \to 0^+} (G(\lambda + is))^{\dagger} = G^{\dagger}(\lambda)$$

Using a well-known identity from the theory of generalized functions

$$\lim_{y \to 0^+} \frac{1}{x \pm iy} = \mathscr{P}\left(\frac{1}{x}\right) \mp i\pi\delta(x)$$

where $\mathscr{P}\left(\frac{1}{x}\right)$ denotes the Cauchy principal value, we see that

$$G^+(E) - G^-(E) = \lim_{s \to 0^+} \sum_n \left[\frac{1}{E - \lambda_n + is} - \frac{1}{E - \lambda_n - is} \right] |\phi_n\rangle\langle\phi_n| = \sum_n -2i\pi\delta(E - \lambda_n)|\phi_n\rangle\langle\phi_n|$$

which leads to the definition of the spectral density $A(E)$

$$A(E) = i\left[G^+(E) - G^-(E)\right] = i\left[G^+(E) - \left(G^+(E)\right)^{\dagger}\right] = 2\pi \sum_n \delta(E - \lambda_n)|\phi_n\rangle\langle\phi_n| \qquad (8.6)$$

$A(E)$ is of central importance in Quantum Physics. Indeed, if one takes the diagonal matrix element of $A(E)$ in the $\{|\vec{x}\rangle\}$ basis we see that

$$\langle\vec{x}|A(E)|\vec{x}\rangle = \sum_n \delta(E - \lambda_n)|\phi_n(\vec{x})|^2$$

which is the *density of states* per unit volume of the system. Integrating that quantity over the entire volume Ω leads to the total density of states

$$\int_{\Omega} d\vec{x}\langle\vec{x}|A(E)|\vec{x}\rangle = \text{Tr}[A(E)] = \sum_n \delta(E - \lambda_n) = \rho(E)$$

where use has been made of the orthonormality of the eigenfunctions. One can also easily verify that the total density of states can be obtained from $G^+(E)$ or $G^-(E)$ alone through

$$\rho(E) = \sum_n \delta(E - \lambda_n) = \mp\frac{1}{\pi}\text{Im}\left(\text{Tr}\left[G^{\pm}(E)\right]\right)$$

8.2.4 Use of Green's Functions

Let us show how the Green's function can be used to easily find solutions to inhomogeneous partial differential equations. Indeed if one is facing an equation similar to this

$$(z - H(\vec{x}))u(\vec{x}) = f(\vec{x})$$

and one also knows about the Green's function $G(\vec{x}, \vec{x}', z)$, then the solution $u(\vec{x})$ is readily given by

$$u(\vec{x}) = \begin{cases} \int_\Omega G(\vec{x}, \vec{x}', z) f(\vec{x}') d\vec{x}', & \text{if } z \neq \lambda_n \text{ in the discrete spectra} \\ \int_\Omega G^+(\vec{x}, \vec{x}', z) f(\vec{x}') d\vec{x}' + \phi(\vec{x}), & \text{if } z = E \text{ in the continuous spectra} \end{cases}$$

where in the second case, $\phi(\vec{x})$ is a general solution to the homogeneous equation, which itself is a general linear combination of eigenfunctions of H corresponding to the eigenvalue E. In an abstract space, the solution reads

$$|u\rangle = \begin{cases} G(z)|f\rangle, & \text{if } z \neq \lambda_n \text{ in the discrete spectra} \\ G^+(z)|f\rangle + |\phi\rangle, & \text{if } z = E \text{ in the continuous spectra} \end{cases}$$

8.2.5 Green's Functions and Perturbation Theory

In this last section on Green's functions, we would like to look at how one can relate the Green's function of two operators H_0 and $H = H_0 + V$. The goal here is to compute the Green's function of H from the knowledge of V and the Green's function of H_0. We start from the definition of G and G_0

$$(z - H_0)G_0(z) = \mathbb{1} \quad \text{and} \quad (z - H)G(z) = \mathbb{1}$$

which easily gives

$$(G_0(z)^{-1} - V)G(z) = \mathbb{1}$$

so we finally arrive at

$$G(z) = \frac{1}{1 - G_0(z)V} G_0(z) = G_0(z) + G_0(z)VG_0(z) + G_0(z)VG_0(z)VG_0(z) + \dots \quad \boxed{8.7}$$

where we used the infinite expansion of $\frac{1}{1-x} = 1 + x + x^2 + \dots$. Introducing the so-called T-matrix defined by

$$T(z) = V + VG_0(z)V + VG_0(z)VG_0(z)V + \dots \quad \boxed{8.8}$$

we see that the total Green's function can be obtained from the knowledge of $T(z)$

$$G(z) = G_0(z) + G_0(z)T(z)G_0(z)$$

Despite the academic looking matrix manipulations, and a priori arbitrary definitions mentioned above, we stress that in practice, switching from $G(z)$ to $T(z)$ allows

us to focus on the perturbation part, which quite often, is a local perturbation in real-space (like for example the perturbation coming from an isolated defect in an otherwise perfect carbon nanotube). Pretty much in all situations of interest, the operators $G(z)$ and $G_0(z)$ in their matrix representation in some localized basis are infinite dimensional, whereas V is finite dimensional. As a consequence, equation 8.8 tells us that the T-matrix is also finite dimensional with the same dimensionality as V. Because $G(z)$ and $T(z)$ share the same spectral properties, we see that knowledge of one readily gives the other. But in practice trying to compute $T(z)$ is much more tractable than directly attacking $G(z)$ from 8.7.

Let us now imagine that we know $T(z)$ (or $G(z)$ which is equivalent). Then one can prove [23] that if H_0 and H share the *same* continuous eigenvalues (i.e. $G_0(z)$ and $G(z)$ have the same continuous eigenvalue spectrum), then one can relate in a *one-to-one* fashion, the proper scattering states of H from the knowledge of the scattering states of H_0. If $|\psi_n\rangle$ is a (causal) scattering state of H originating from $|\phi_n\rangle$, scattering state of H_0, then

$$
\begin{aligned}
|\psi_n\rangle &= |\phi_n\rangle + G_0^+(E)T^+(E)|\phi_n\rangle \\
&= |\phi_n\rangle + G^+(E)V|\phi_n\rangle \\
&= |\phi_n\rangle + G_0^+(E)V|\psi_n\rangle
\end{aligned}
\tag{8.9}
$$

In particular, the first two equations relate $|\phi_n\rangle$ to $|\psi_n\rangle$ directly through a linear transformation. The last equation is often referred to as the *Lippmann-Schwinger* equation [63].

8.3 The Fisher-Lee Formula for Lead-Conductor-Lead Systems

In this section, we will first introduce the general setup of Lead-Conductor-Lead systems, and then move on to prove a general expression for the current operator in a localized basis. In the last part, a proof of the Fisher-Lee formula [33] is given. Working out a proof will teach us a great deal about the assumptions underlying the theory along with offering a deep understanding of the physics involved.

8.3.1 The LCR Setup

The so-called Lead-Conductor-Lead setup is a theoretical idealization of an experimental two-probe junction, in which a small system – the conductor – is sandwiched between two semi-infinite leads which themselves are connected "at infinity" to macroscopic electron reservoirs. Those reservoirs impose their own chemical potential and temperature to the leads. We sketch the situation on figure 8-2. The leads are semi-infinite systems with perfect periodicity in the elongated direction. As a consequence, their real-space Hamiltonian matrix has some symmetries. Because we choose to express Hamiltonian matrices in a localized basis (like for example the
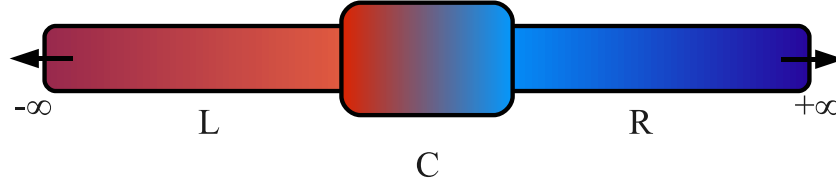
Figure 8-2: Typical LCR setup for Quantum Conductance calculations in quasi-one dimensional systems. The left lead (L) is connected at $-\infty$ to an electron reservoir with chemical potential μ_L and temperature T. In the same way, the right lead (R) is connected at $+\infty$ with an electron reservoir with chemical potential μ_R and temperature T. The temperatures of the two reservoir could also be different in general.

Wannier basis), we know that the matrix elements will be localized around each basis function. For the leads, we can hence define a **principal layer**, which consists of an agglomeration of a certain amount of lead unit cells, such that the interaction among principal layers is restricted to nearest neighbors. We illustrate the fundamental concept of a principal layer in figure 8-3. This figure features a simple semi-infinite linear chain of periodically repeated atoms so as to represent a model of right lead from figure 8-2. A single Wannier function w_n is attached to each atom n, and the localization of those Wannier functions is such that only the nearest and second nearest neighbors can interact through hopping integrals

$$\langle w_n|H|w_{n+1}\rangle = \langle w_n|H|w_{n-1}\rangle = t_1 \text{ and } \langle w_n|H|w_{n+2}\rangle = \langle w_n|H|w_{n-2}\rangle = t_2$$

respectively. The on-site matrix elements of the Hamiltonian are $\langle w_n|H|w_n\rangle = \epsilon$. We then observe on the Hamiltonian matrix that if the basis of Wannier functions is properly arranged (or *sorted* as we will say from now on) so as to reflect the real-space symmetry of the system, two fundamental building blocks emerge, namely the Principal Layer H^{00} and the coupling between adjacent Principal Layers H^{01}. Those blocks repeat themselves *ad infinitum* to form the Hamiltonian matrix of the right lead. With the splitting of the system into left lead, conductor and right lead, along with the assumed periodicity of the leads, we can write quite generally the Hamiltonian of the LCR system in a *sorted* localized basis as

$$H = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & H_L^{00} & H_L^{01} & 0 & 0 & 0 & \cdots \\ \cdots & H_L^{01\dagger} & H_L^{00} & H_{LC} & 0 & 0 & \cdots \\ \cdots & 0 & H_{LC}^{\dagger} & H_C & H_{CR} & 0 & \cdots \\ \cdots & 0 & 0 & H_{CR}^{\dagger} & H_R^{00} & H_R^{01} & \cdots \\ \cdots & 0 & 0 & 0 & H_R^{01\dagger} & H_R^{00} & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (8.10)$$
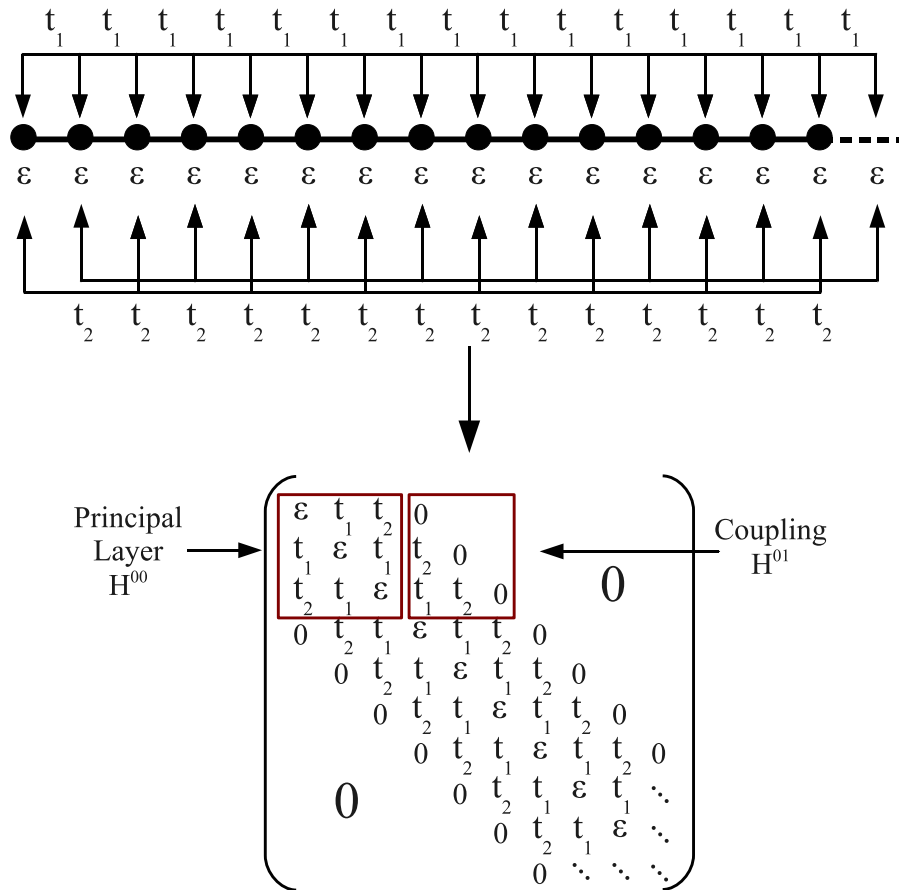
Figure 8-3: The concept of the Principal Layer (PL) is illustrated on a semi-infinite one-dimensional linear chain emulating a semi-infinite right lead. In this model an electron on a site n can hop onto the nearest and second nearest neighbor sites through hopping integrals t_1 and t_2. Moreover the on-site matrix elements are given by ϵ. When the Wannier basis (or any localized basis for that matter) is arranged so as to reflect the intrinsic periodicity of the system, we see that the Hamiltonian matrix is fundamentally the repetition of two basic blocks. The first block is centered on the main diagonal and called the Principal Layer H^{00}, while the second block represents the interaction between two principal layers H^{01}.

This form for the Hamiltonian of the full LCR system will be our starting point for the rest of the chapter.

8.3.2 A General Expression for the Current Operator

In order to compute the electronic current in the system, we obviously need a workable expression for the current operator. For this we will use our physical intuition for what that current should be. We start off from the time-dependent Schrödinger

equation

$$i\hbar \frac{d|\psi(t)\rangle}{dt} = H|\psi(t)\rangle$$

from which, in theory, we can obtain the state vector at time t, $|\psi(t)\rangle$. If we want to know what is the probability to find an electron in the right lead at time t, all we need to do is compute the expectation value of the right-lead projector operator for a state vector $|\psi(t)\rangle$

$$\rho_R(t) = \langle\psi(t)|P_R|\psi(t)\rangle$$

There is no magic in the above formula. In terms of matrices, P_R is the identity matrix restricted to the right lead subspace only. If state vectors are normalized, we know that we are *sure* to find an electron in the system, and so in general

$$\rho_{\text{system}}(t) = \langle\psi(t)|P_L + P_C + P_R|\psi(t)\rangle = 1$$

$\rho_R(t)$ then, just tells us what is the fraction of electron found in the right lead.

The algebraic probability current into the right lead is then simply the time derivative of $\rho_R(t)$. This results from a simple argument of conservation of charge. Let us then calculate the electron current into the right lead from the expression for $\rho_R(t)$ (the electron current is simply $-e$ times the probability current, where e is the absolute value of the electron charge)

$$j_R(t) = -e\frac{d\rho_R(t)}{dt} = -e\frac{d\langle\psi(t)|}{dt}|P_R|\psi(t)\rangle - e\langle\psi(t)|P_R|\frac{d|\psi(t)\rangle}{dt}$$

Using the Schrödinger equation, we arrive at

$$j_R(t) = (-e)\left(-i\hbar\langle\psi(t)|HP_R|\psi(t)\rangle + \langle\psi(t)|P_R|i\hbar H|\psi(t)\rangle\right)$$

so, introducing the commutator $[A,B] = AB - BA$ we finally end up with

$$j_R(t) = -e\frac{i}{\hbar}\langle\psi(t)|[H,P_R]|\psi(t)\rangle$$

which directly leads to the expression for the current operator J_R

$$J_R = -e\frac{i}{\hbar}[H,P_R] \tag{8.11}$$

To close the discussion, let us use the matrix in equation 8.10 to express the current operator in terms of the fundamental blocks of H. Using equation 8.11, and performing the appropriate matrix multiplications, we easily arrive at a most useful formula for the current operator

$$J_R = -e\frac{i}{\hbar}\left(H_{CR} - H_{CR}^\dagger\right) \tag{8.12}$$

From this formula, it is obvious that if the coupling between the central conductor and the right lead is zero, then no current can flow into the right lead, which makes

intuitive sense.

8.3.3 Modeling the Physics of Electron Flow

We are now ready to discuss how we will tackle the problem of computing the net current flowing from the left lead to the right lead through the conductor. This derivation is inspired in part from an insightful discussion by Janne Viljas [100]. The first assumption used is that the electron reservoirs are thermalized with their own temperature T and chemical potential μ. Moreover, the contacts between the leads and the reservoirs are such that when an electron is injected into the reservoirs from the leads, no reflection of the electron is possible. This fact is the consequence of the great disparity of available electronic states between the macroscopic reservoirs and the nanoscopic leads, which basically sets the probability for an incoming electron to be reflected back into the lead to zero. Because of the thermal equilibrium of the electron reservoirs, the probability for an electron to be injected into one eigenstate of the leads with an energy E is given by the Fermi function

$$f_X(E) = \frac{1}{1 + e^{\frac{E - \mu_X}{k_B T}}}, \quad \text{where } X = L, R \tag{8.13}$$

Let us consider the Hamiltonian matrix of the LCR system of equation 8.10 to be given (that matrix generally comes from step 2 in figure 8-1). We will split that matrix into two parts. The first part, the "unperturbed" Hamiltonian H_0, will correspond to the full matrix without the coupling matrices H_{LC} and H_{CR}

$$H_0 = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & H_L^{00} & H_L^{01} & 0 & 0 & 0 & \cdots \\ \cdots & H_L^{01\dagger} & H_L^{00} & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & H_C & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & H_R^{00} & H_R^{01} & \cdots \\ \cdots & 0 & 0 & 0 & H_R^{01\dagger} & H_R^{00} & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{8.14}$$

while the "perturbation" matrix V will be

$$V = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & 0 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & H_{LC} & 0 & 0 & \cdots \\ \cdots & 0 & H_{LC}^{\dagger} & 0 & H_{CR} & 0 & \cdots \\ \cdots & 0 & 0 & H_{CR}^{\dagger} & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 0 & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{8.15}$$

We observe in particular that H_0 does *not* couple the leads to the conductor. The

spectrum of the "unperturbed" hamiltonian then consists of isolated eigenvalues corresponding to the eigenstates of H_C, along with a continuous set of eigenvalues corresponding to the eigenstates of the leads H_L and H_R. When the perturbation is added, the qualitative picture is kept for the full Hamiltonian matrix. In particular the localized eigenstates in the conductor will be perturbed but will keep their localized nature. On the other hand, the continuous spectrum of the full Hamiltonian will correspond to the continuous spectrum of H_0. Let us denote the leads eigenstates by $|\phi_n\rangle$, where n is a continuous quantum number labeling the states. In the rest of the chapter, we will be focusing on the scattering eigenstates of the full LCR system arising as a perturbation of the eigenstates of the left lead. The situation being formally equivalent for the right lead, we will just state the result. Using the one-to-one mapping mentioned earlier in section 8.2, we can write the scattering eigenstates of the full system, $|\psi_n\rangle$ (corresponding to an eigenvalue $\lambda_n = E$) as

$$|\psi_n\rangle = |\phi_n\rangle + G_0^+(E)V|\psi_n\rangle \qquad (8.16)$$

where $G_0^+(E)$ is the retarded Green's function of the "unperturbed" system.

In order to compute the electron current we will then simply compute the elementary current carried by each eigenstate $|\psi_n\rangle$ and sum them all up for all quantum numbers n compatible with the given energy E. The Landauer formula then proceeds by weighting those currents by the probability for the electron's energy to equate E (simply given by the Fermi function) and integrate over all possible energy values (typically from $-\infty$ to $+\infty$).

8.3.4 An Expression for the Conductor's Green's Function

Before to compute the electron current, let us prove a very useful expression for the conductor part of the full system's Green's function. We start from the definition of the Green's function $G(z)$ and write

$$(z - H)G(z) = \begin{pmatrix} z - H_L & -H_{LC} & 0 \\ -H_{LC}^\dagger & z - H_c & -H_{CR} \\ 0 & -H_{CR}^\dagger & z - H_R \end{pmatrix} \begin{pmatrix} G_L & G_{LC} & G_{LR} \\ G_{CL} & G_C & G_{CR} \\ G_{RL} & G_{RC} & G_R \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

From the block matrix multiplications, we isolate at least 3 interesting relations

$$\begin{aligned} -H_{LC}^\dagger G_{LC} + (z - H_C)G_C - H_{CR}G_{RC} &= 1 \\ (z - H_L)G_{LC} - H_{LC}G_C &= 0 \\ -H_{CR}^\dagger G_C + (z - H_R)G_{RC} &= 0 \end{aligned}$$

The first equation correspond the central diagonal blocks of $(z - H)G(z)$, whereas the last two are respectively the middle upper block and middle lower block. From the

last two equations, we easily find

$$
\begin{aligned}
-H_{CR}G_{RC} &= -H_{CR}\frac{1}{z-H_R}H_{CR}^{\dagger}G_C \\
-H_{LC}^{\dagger}G_{LC} &= -H_{LC}^{\dagger}\frac{1}{z-H_L}H_{LC}G_C
\end{aligned}
$$

which once injected into the first of the previous set of equations, leads to

$$
G_C = \frac{1}{z-\left[H_C+H_{CR}\frac{1}{z-H_R}H_{CR}^{\dagger}+H_{LC}^{\dagger}\frac{1}{z-H_L}H_{LC}\right]} \tag{8.17}
$$

The formula in equation 8.17 is interesting for at least two reasons. The first is the fact that one can express the conductor part of the Green's function G_C from multiplications and inversions of *finite matrices*. Indeed, despite the fact that $G_L^0(z) = \frac{1}{z-H_L}$ and $G_R^0(z) = \frac{1}{z-H_R}$ are infinite matrices, the finiteness of H_{LC} and H_{CR} implies that only a *finite* part of $G_L^0(z)$ and $G_R^0(z)$ really matters. The submatrices from $G_L^0(z)$ and $G_R^0(z)$ that couple to H_{LC} and H_{CR} are called **surface Green's functions**. A very robust and highly convergent scheme has been devised by Lopez-Sancho *et al.* in order to compute those surface Green's functions. We refer the interested reader to the appropriate literature [86]. The second reason for finding equation 8.17 interesting has to do with the fact that the coupling matrices naturally add two extra terms that can be seen as a "renormalization" of the conductor Hamiltonian H_C. Those terms are called the **lead self-energies**. To simplify the algebra of the next section, and at the same time conform to established rules, we will introduce the following notations

$$
\begin{aligned}
\Sigma_L(E) &= H_{LC}^{\dagger}\frac{1}{E-H_L+i0^+}H_{LC} \tag{8.18} \\
\Sigma_R(E) &= H_{CR}\frac{1}{E-H_R+i0^+}H_{CR}^{\dagger} \\
\Gamma_L(E) &= i\left(\Sigma_L-\Sigma_L^{\dagger}\right) \\
\Gamma_R(E) &= i\left(\Sigma_R-\Sigma_R^{\dagger}\right)
\end{aligned}
$$

where $i0^+$ is here to indicate a very small but positive imaginary part.

8.3.5 The Fisher-Lee Formula

We finally arrive at the last step where we will compute the current flowing through the LCR system from the left lead to the right lead. As we saw earlier, the proper scattering eigenstate of the LCR system corresponding to an electron originating from the left lead and going through the conductor into the right lead is given by

$$
|\psi_n\rangle = |\phi_n\rangle + G_0^+(E)V|\phi_n\rangle
$$

where $G_0^+(E)$ is the retarded Green's function of the unperturbed Hamiltonian H_0 and for a given energy E. Using equation 8.12 for the current operator, we compute the elementary current as

$$j_n(E) = \langle \psi_n | J_R | \psi_n \rangle = -e\frac{i}{\hbar}\langle \psi_n | H_{CR} - H_{CR}^\dagger | \psi_n \rangle = -e\frac{i}{\hbar}\langle \psi_n | H_{CR}P_R - P_R H_{CR}^\dagger | \psi_n \rangle$$

with P_R as the projector onto the right lead. Using the Lippmann-Schwinger equation 8.9 and the fact that the scattering state $|\phi_n\rangle$ only lives in the left lead subspace [4] such that $P_R|\phi_n\rangle = 0$, we find

$$P_R|\psi_n\rangle = P_R G_0^+(E)V|\psi_n\rangle = \left(G_R^0\right)^+(E)H_{CR}^\dagger|\psi_n\rangle$$

where $\left(G_R^0\right)^+(E)$ is the right lead part of the retarded Green's function for the unperturbed Hamiltonian H_0. To convince oneself that the above is true the reader is invited to perform the block matrix multiplications. Injecting that expression into the elementary current we end up with

$$j_n(E) = -e\frac{i}{\hbar}\langle \psi_n | H_{CR}\left(G_R^0\right)^+(E)H_{CR}^\dagger - H_{CR}\left[\left(G_R^0\right)^+(E)\right]^\dagger H_{CR}^\dagger | \psi_n \rangle$$

which, given the definition of $\Gamma_R(E)$ in equation 8.18 reduces to

$$j_n(E) = -\frac{e}{\hbar}\langle \psi_n | \Gamma_R(E) | \psi_n \rangle \tag{8.19}$$

Having the elementary current corresponding to a single scattering eigenstate, we find the total contribution to the current at energy E by summing over all the scattering eigenstates for which the quantum number n is compatible with that energy. We find

$$j(E) = \sum_n \delta(E - \lambda_n)j_n(E)$$

where the delta function is here to ensure that we only select scattering eigenstates with a proper quantum number n such that $\lambda_n = E$. From equation 8.19 we expand to

$$j(E) = -\frac{e}{\hbar}\sum_n \delta(E - \lambda_n)\langle \psi_n | \Gamma_R(E) | \psi_n \rangle$$

The fundamental observation at this point is to realize that $\Gamma_R(E)$ is a matrix that lives in the conductor subspace. As a consequence we can replace it by $P_C\Gamma_R(E)P_C$ and also introduce the completeness relation for that subspace $\sum_c |c\rangle\langle c|$. The result reads

$$j(E) = -\frac{e}{\hbar}\sum_c \langle c | P_C \Gamma_R(E) P_C \left(\sum_n \delta(E - \lambda_n)|\psi_n\rangle\langle\psi_n|\right) |c\rangle$$

---

[4] remember that $|\phi_n\rangle$ is the unperturbed scattering state in the left lead. So in particular, that state does not "spill over" onto the conductor or the right lead

or equivalently

$$j(E) = -\frac{e}{\hbar}\mathrm{Tr}_c\left[P_C\Gamma_R(E)P_C\left(\sum_n\delta(E-\lambda_n)|\psi_n\rangle\langle\psi_n|\right)\right]$$

with Tr_c being the trace over the conductor subspace.

The final touch consists in expressing $\sum_n\delta(E-\lambda_n)|\psi_n\rangle\langle\psi_n|$ from known quantities. Actually, we only need the projection of that quantity onto the conductor subspace, namely $P_C\sum_n\delta(E-\lambda_n)|\psi_n\rangle\langle\psi_n|P_C$ since by invariance of the trace with cyclic permutations we have

$$\mathrm{Tr}_c\left[P_C\Gamma_R(E)P_C\left(\sum_n\delta(E-\lambda_n)|\psi_n\rangle\langle\psi_n|\right)\right] = \mathrm{Tr}_c\left[\Gamma_R(E)P_C\left(\sum_n\delta(E-\lambda_n)|\psi_n\rangle\langle\psi_n|\right)P_C\right]$$

We use the second form of equation 8.9 and the fact that $P_C|\phi_n\rangle = 0$, where $|\phi_n\rangle$ is a scattering eigenstate of the left lead and compute

$$
\begin{aligned}
P_C\sum_n\delta(E-\lambda_n)|\psi_n\rangle\langle\psi_n|P_C &= \sum_n\delta(E-\lambda_n)P_C|\psi_n\rangle\langle\psi_n|P_C \\
&= \sum_n\delta(E-\lambda_n)P_CG^+(E)V|\phi_n\rangle\langle\phi_n|V^\dagger\left[G^+(E)\right]^\dagger P_C \\
&= \sum_n\delta(E-\lambda_n)G_C^+(E)H_{LC}^\dagger|\phi_n\rangle\langle\phi_n|H_{LC}\left[G_C^+(E)\right]^\dagger \\
&= G_C^+(E)H_{LC}^\dagger\left(\sum_n\delta(E-\lambda_n)|\phi_n\rangle\langle\phi_n|\right)H_{LC}\left[G_C^+(E)\right]^\dagger \\
&= G_C^+(E)H_{LC}^\dagger\frac{A_L(E)}{2\pi}H_{LC}\left[G_C^+(E)\right]^\dagger
\end{aligned}
$$

where use has been made of the expression for the spectral density in 8.6 for the left lead. We now use the definition of the spectral density and write

$$\frac{A_L(E)}{2\pi} = \frac{i}{2\pi}\left[\left(G_L^0(E)\right)^+ - \left[\left(G_L^0(E)\right)^+\right]^\dagger\right]$$

which once injected into the previous expression leads to

$$P_C\sum_n\delta(E-\lambda_n)|\psi_n\rangle\langle\psi_n|P_C = \frac{1}{2\pi}G_C^+(E)H_{LC}^\dagger\left[\left(G_L^0(E)\right)^+ - \left[\left(G_L^0(E)\right)^+\right]^\dagger\right]H_{LC}\left[G_C^+(E)\right]^\dagger$$

The definition of $\Gamma_L(E)$ lets us conclude that

$$P_C\sum_n\delta(E-\lambda_n)|\psi_n\rangle\langle\psi_n|P_C = \frac{1}{2\pi}G_C^+(E)\Gamma_L(E)\left[G_C^+(E)\right]^\dagger$$

All in all, the final expression for the current is

$$j(E) = -\frac{e}{\hbar}\frac{1}{2\pi}\mathrm{Tr}_c\left[\Gamma_R(E)G_C^+(E)\Gamma_L(E)\left[G_C^+(E)\right]^\dagger\right]$$

which, once re-arranged through cyclic permutations of the operators leads the Fisher-Lee formula

$$j(E) = -\frac{e}{h} \text{Tr}_c \left[G_C^+(E) \Gamma_L(E) \left[G_C^+(E) \right]^\dagger \Gamma_R(E) \right] \qquad (8.20)$$

8.3.6 General Landauer Formula for LCR Systems

From the Fisher-Lee formula, it is a piece of cake to find the Landauer formula for the total current. Indeed, equation 8.20 gives us the total current at energy E coming from the left lead into the right lead. We know that the probability for an electron to be injected by the left reservoir into the left lead with an energy E is given by the Fermi function $f_L(E)$. To compute the total current through the system we just need to subtract from that current the current coming from the right lead into the left lead for which the expression in 8.20 is identical up to a minus sign in front. The total current is then the integrated value of the net total currents at energy E. The Landauer formula then reads

$$I = -\frac{e}{h} \int_{-\infty}^{+\infty} dE \, (f_L(E) - f_R(E)) \text{Tr}_c \left[G_C^+(E) \Gamma_L(E) \left[G_C^+(E) \right]^\dagger \Gamma_R(E) \right] \qquad (8.21)$$

from which we can define the *transmission function* of the LCR system to be

$$\mathscr{T}(E) = \text{Tr}_c \left[G_C^+(E) \Gamma_L(E) \left[G_C^+(E) \right]^\dagger \Gamma_R(E) \right] \qquad (8.22)$$

This last equation concludes our overview of the fundamentals of Quantum Transport.

9

Automated Large Scale Quantum Conductance Calculations

In this chapter, we intend to present some details of the practical challenges and algorithmic solutions to those challenges in actual quantum conductance calculations. In a first section, we will compare the ideal theoretical LCR system to practical LCR systems in order to identify some challenges to be overcome if one wish to predict quantum conductances from first principles. In a second section, we will outline our algorithmic solutions. In a last section, an approach to the problem of tackling experimentally relevant sizes for LCR systems is put forward building on the results from section two. A detailed exposition of the problematic of this chapter can be found in [71].

9.1 Theory versus Practice

The theoretical derivations of the previous chapter left us with the impression that implementing a workable scheme for computing the quantum conductance of LCR systems is a simple matter of coding. In this section, we will outline the implicit assumptions of the theory and a comparison with practical limitations will allow us to arrive at the formulation of a set of problems that need solutions if we ever want to compute quantum conductances in practice.

9.1.1 LCR Systems and Periodic Boundary Conditions

The theoretical LCR system of figure 8-2 is intrinsically an *aperiodic* system, in which two semi-infinite leads, with perfect periodicity, are connected to a central conductor with no *a priori* periodicity. In practice however, *periodic boundary conditions* are most often used for the calculation of the electronic ground state of any system. One of the main reason has to do with the use of a plane wave basis set for which efficient algorithms for the calculation of the ground state energy are available [74]. The first consequence of periodicity is to impose a similar nature for the left and right leads, which is a constraint that was not necessary in the theoretical derivation. Using then a left and right lead consisting of the same material, we still observe on figure 9-1 that we can't actually simulate semi-infinite leads. A solution may be found by

using an appropriate supercell size that will include a sufficient amount of leads in order to recover the lead's "bulk" limit around the midpoint region between two adjacent conductors. This idea is illustrated on figure 9-1. The length of the supercell is increased until the electronic structure at the midpoint recovers its "bulk" value corresponding to an infinite lead. The key question of course is how fast does the elec-
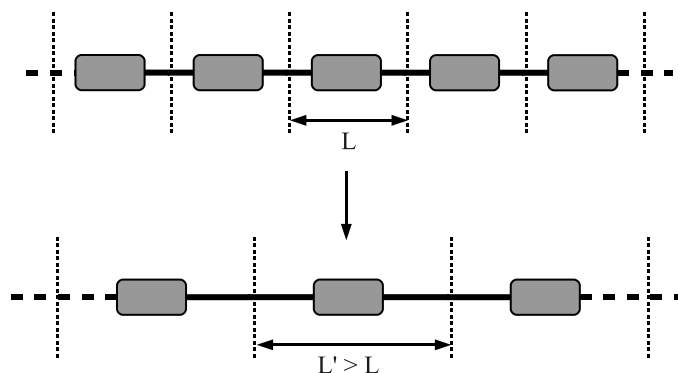

Figure 9-1: A simulated LCR system consists of the repetition of a motif because of periodic boundary conditions. The dashed vertical lines denote the limits of the computational cell (or supercell). In order to approach the limit of semi-infinite leads one may increase the amount of leads to be put inside the supercell. The goal is to converge the electronic structure in the leads so as to recover the "bulk" limit of an infinite pristine lead.

tronic structure converges with the supercell size. If we use a Bloch basis (or equivalently a plane wave basis since the Bloch functions are expanded in a plane wave basis), then it seems difficult to estimate when the Bloch function recovers its "bulk" value, considering that those functions are delocalized over the entire system. At variance with Bloch functions, Maximally Localized Wannier Functions (MLWF) are extremely well suited to this kind of situation. It has been shown quite extensively [60] that MLWFs recover their bulk values quite rapidly. In fact, in most situations, the recovery is exponential, signifying that only a few unit cells of lead are actually necessary to separate a conductor region from its periodic images. To put some numbers in perspective, we find empirically that about 3 unit cells of leads on each side of the conductor region are sufficient to recover "bulk" matrix elements in the leads for typical Armchair Carbon nanotubes[1]. We will come back to this idea of electronic nearsightedness in the MLWF basis in section 9.3. We arrive to the conclusion that using a supercell for the ground state DFT calculation, followed by a transformation to a MLWF basis seem an appropriate way to deal with periodic boundary conditions while extracting an Hamiltonian matrix suitable to mimic an aperiodic infinite system. Still left unsolved is the question of how best to choose the supercell geometry. This will constitute our first challenge.

---

[1]Those numbers are even more impressive when one considers that Armchair nanotubes are metallic, which means that electrons close to the Fermi level are free to roam about the system.

9.1.2 Sorted basis and Parity Consistence

A set of two other very important assumptions have been made in the theoretical derivation of chapter 8. The first assumption pertains to the very fact that we could partition the Hamiltonian matrix in five submatrices, namely H_L, H_{LC}, H_C, H_{CR} and H_R and enforce the lead matrices H_L and H_R to be periodic. For all this to be true, we obviously need the basis of MLWF to be **sorted** in space. Yet another assumption, more subtle but at least as important in practice, was that the hopping matrix elements were taken to be consistent in their signs from one unit cell of lead to another. Both issues are illustrated on figure 9-2. The featured system is a simple
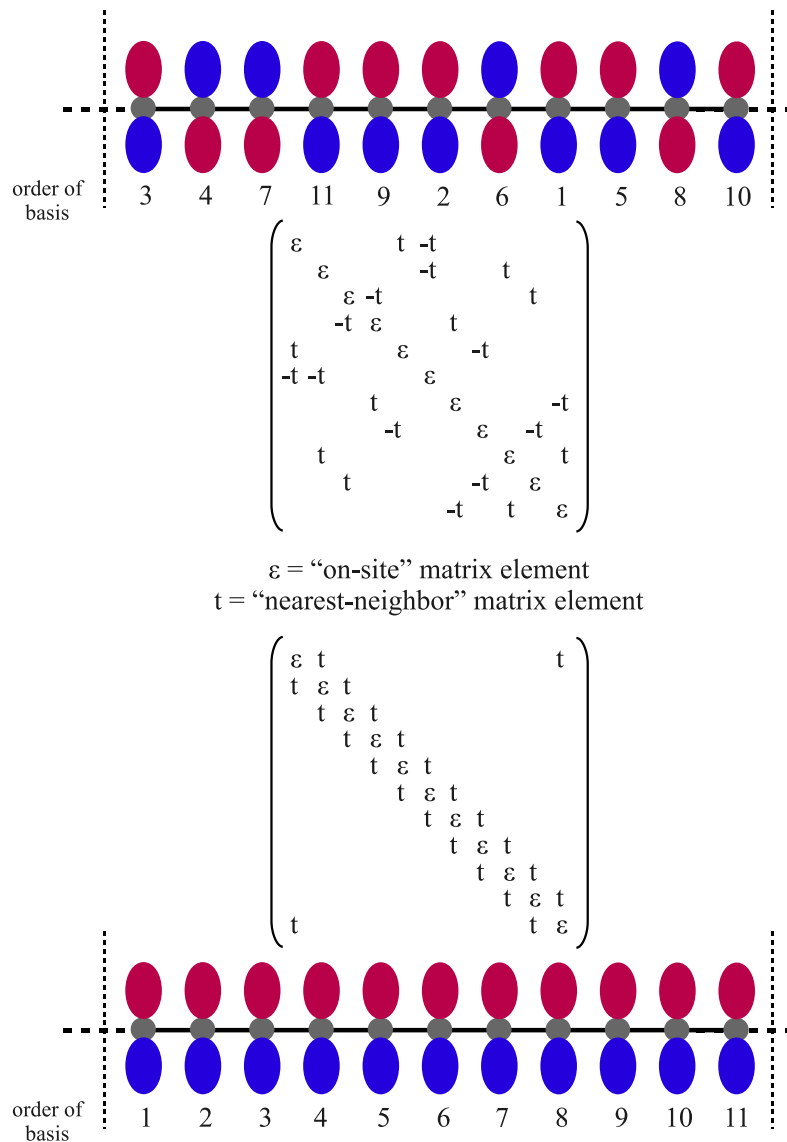

Figure 9-2: illustration of the sorting and parity issues on a simple linear chain of atoms each with a unique p_z-like Wannier function.

linear chain of atoms for which each atom has a unique p_z-like Wannier function attached to it. We take the onsite matrix element of the Hamiltonian to be ϵ, while the nearest neighbor interaction between two adjacent Wannier functions with a consistent sign pattern[2] is t. The vertical dashed lines denote once more the boundary of the computational supercell. The sketch at the top, displays a typical output from a Wannier function calculation. The basis of Wannier functions bears no particular relation to the real-space periodic symmetry of the system. What's more, the Wannier function parities (or overall sign) are given randomly. As a consequence, the Hamiltonian matrix, which in general is sparse because of the localized nature of the Wannier functions, has its matrix elements scattered all over. Moreover, some of the hopping parameters are positive while others are negative because of the parity inconsistency from one Wannier function to another. In contrast, the lower sketch shows the same system for which the basis has been sorted and the parities have been fixed. We then observe that the Hamiltonian matrix looks exactly like what we assumed in the previous chapter. The sorting of the basis and the enforcement of a consistent parity pattern constitute two more challenges that we will need to overcome.

9.2 Practical Algorithms for Quantum Transport

In the preceding section, we raised three fundamental challenges. Time has come to give details about our algorithmic strategies to address those issues.

9.2.1 The Single Supercell Geometry

The first challenge was to find a proper choice for the LCR supercell in order to unambiguously partition the full Hamiltonian matrix into critical submatrices. Our approach consists in using a principal layer of lead for both the left and right leads, along with a conductor consisting of two buffer layers on both side of the defected region. Only a single k point in the Brillouin Zone must be used (Γ point in the center of the Brillouin Zone) so as to make sure that the periodicity of the supercell put the right principal layer (PL2) immediately to the left of the left principal layer (PL1). We call such a supercell the *single supercell geometry* or SSG for short. The structure of the SSG is illustrated on figure 9-3. Depicted by vertical dashed segments are the boundary of the supercell. The different submatrices needed for the transport calculation are obtained as follows

- The left lead matrix H_L consists of two submatrices. The principal layer matrix H_L^{00}, and the coupling matrix H_L^{01}. The first is obtained by extraction of the Hamiltonian matrix elements for the Wannier functions in PL1. The second

---

[2]In general Wannier functions are complex-valued, but in our approach described in the next section, we enforce those Wannier functions to be real. What's more we saw in chapter II that in general, the set of Wannier functions corresponding to the global minimum of the spread functional are real if the minimum is unique. Using real Wannier functions, we see that the unknown phase becomes an unknown sign. This is that overall sign that we will call **parity** from now on
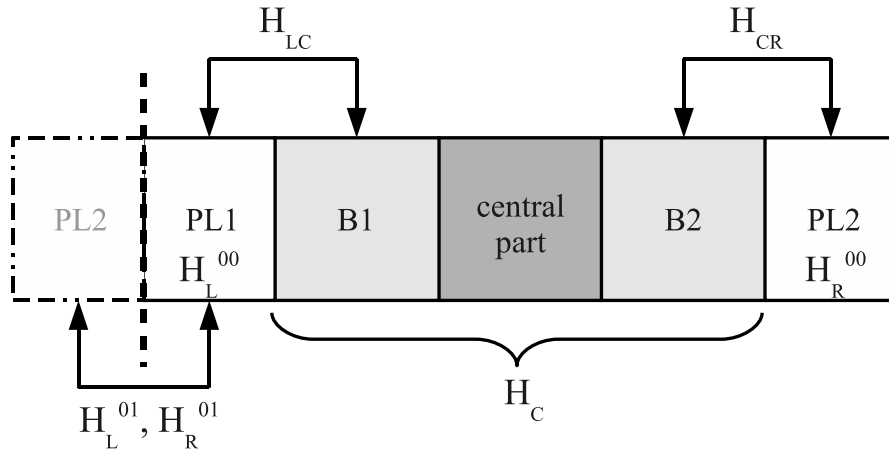
Figure 9-3: illustration of the concept of *single supercell geometry*.

matrix is obtained from matrix elements between Wannier functions in PL2
and Wannier functions in PL1. The right lead matrix is similarly extracted
from PL2 for H_R^{00} and the same coupling matrix as for the left lead.

- The conductor matrix H_C consists of the Hamiltonian matrix elements for the
  Wannier functions in the central part plus the two buffer layers. The role of the
  buffer layers is to confine the defected region to the central part and ensure a
  clean connection of the conductor with the leads.

- At last the coupling matrices of the conductor with the leads are given by the
  matrix elements between Wannier functions in PL1 and B1 and between Wan-
  nier functions in B2 and PL2 for respectively H_{LC} and H_{CR}

9.2.2 Sorting the MLWF

The Hamiltonian sub-matrices attained from partitioning the total Hamiltonian re-
quire a number of operations performed on them before they can be input into trans-
port calculations. First, we need to re-order the MLWFs in real-space so that every
unit cell in PL1, PL2, B1 and B2 has a consistent sequence of MLWFs. This is be-
cause the Hamiltonian corresponding to the semi-infinite leads is constructed from
sub-matrices extracted from the SSG Hamiltonian in the MLWF basis. As we said
earlier, the coupling matrix H_L^{01} is constructed from the Hamiltonian matrix ele-
ments between MLWFs in PL1 and the periodic image of PL2, whereas H_L^{00} is con-
structed from PL1 only. These two matrices are then duplicated along the block
off-diagonal and block diagonal, respectively, of the Hamiltonian of equation 8.10. In
doing so, the implicit assumption is that the sequence of MLWFs in the rows and
columns of the Hamiltonian sub-blocks are the same, which in general is not true.
To overcome this problem we use the positions of the MLWF centers in real-space to
order the elements of the Hamiltonian sub-matrices: the MLWFs in each unit cell

of lead are arranged first according to their position along one direction perpendicular to the transport direction, then in the other perpendicular direction, and finally along the transport direction itself. This ensures that the sub-matrices can be used consistently to build the Hamiltonian of equation 8.10.

The shape of MLWFs are often chemically intuitive and display atomic-like or bonding/anti-bonding orbitals. Thus if more than one MLWF exists with precisely the same center, as may happen with d-like MLWFs on a transition metal site, a second level of ordering based on the orbital character of the MLWF is performed, employing a technique we have developed using spatially-dependent integrals to deduce a unique signature for each MLWF (see Appendix E).

9.2.3 Imposing a Consistent Parity Pattern

As with the issue of ordering the MLWFs, the procedure of building the Hamiltonian from sub-matrices implicitly assumes that the MLWFs in PL2 have the same parity pattern as those in PL1, which in general is not true. To address this issue, we enforce a consistent parity pattern at the level of the unit cell of lead onto the ordered Hamiltonian sub-matrices (PL1, B1, B2 and PL2). The parities of the MLWFs in the leftmost unit cell of lead in the SSG supercell are used as a template. By assessing the relative parity of MLWFs in this unit cell compared to translationally equivalent MLWFs in the other unit cells of the PLs and buffer regions, the pattern is enforced throughout by multiplying by ± 1, as appropriate. The relative parities are determined by using the unique signature associated with each MLWF explained in Appendix E.

We outline three caveats that apply to the current implementation of the SSG method:

- the Bloch states used as input for determining the MLWF basis in the SSG are calculated at the Γ-point only

- the lattice vectors of the SSG must form a orthorhombic set and the direction of conduction must be in the x, y or z direction.

- the system under investigation must be quasi-one-dimensional

To conclude this section, we present on figure 9-4 a detailed work flow of the different steps involved in a quantum conductance calculation.
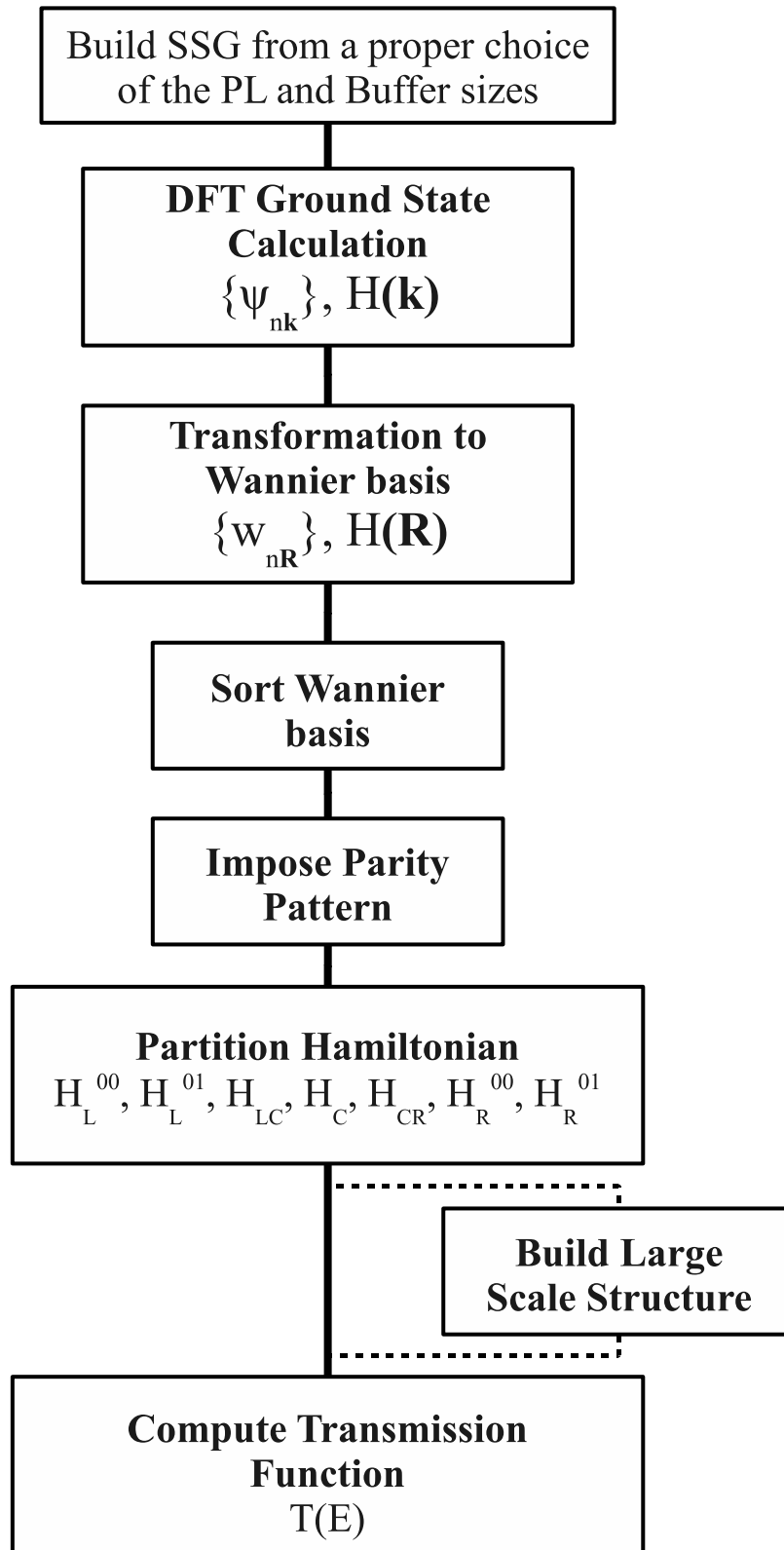
Figure 9-4: Full work flow of a quantum conductance calculation.

9.3 Large Scale Quantum Transport

Moving from a Bloch to a Wannier representation is not only a way to represent electronic structures in a very compact manner. It also opens the possibility to exploit the real-space nature of the basis to build very large systems. Systems so large in fact, that a conventional DFT calculation would be intractable.

The fact that electronic nearsightedness becomes explicitly manifest in the MLWF basis, as demonstrated in figure 9-5, allows them to be used to build the Hamiltonian matrix of a large structure from the smaller Hamiltonian matrices of its constitutive sub-systems.
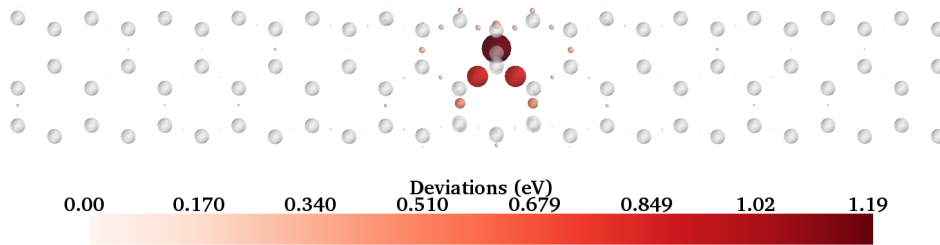

Figure 9-5: Illustration of the inherent electronic nearsightedness in a (3,3) carbon nanotube functionalized with a single hydrogen atom. The white-gray spheres represent the atoms of the structure while the colored spheres represent the deviations from the "bulk" values of the on-site Hamiltonian matrix elements for each MLWF. The size of colored spheres is another indication of the deviation of the matrix element from its "bulk" value. The smaller the sphere, the smaller the deviation. We observe a very rapid decay of the deviations indicating a very local perturbation of the Wannier functions from the defected region.

In order to illustrate the method, consider the schematic lead-conductor-lead system shown in figure 9-6 in which the conductor region has two identical defects separated by a region of lead material in the form of a buffer (B2 and B'1). We could calculate the QC of this structure by making a SSG with the whole conductor (regions corresponding to matrices H_X and $H_{X'}$). In that case, the full conductor matrix would look something like

$$H_C = \begin{pmatrix} H_X & H_{XX'} \\ H_{XX'}^\dagger & H_{X'} \end{pmatrix}$$

(9.1)

where X, X' and XX' represent blocks of Hamiltonian matrix elements among MLWFs in region X, among MLWFs in region X', and between MLWFs in these two regions, respectively. However, we may exploit the nearsightedness of the MLWF basis to find a more computationally efficient approach. If the effect of the defects are localized (in the sense that the local electronic structure and geometry at the junction between B2 and B'1 is sufficiently similar to that seen in the leads), then we may construct the Hamiltonian for the system with two defects (figure 9-6) from
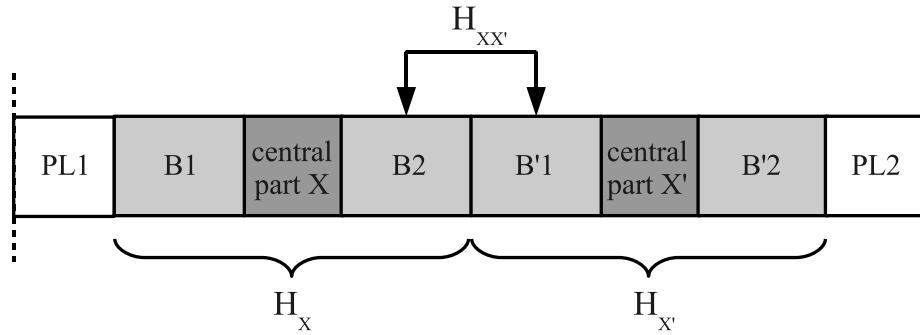
Figure 9-6: Schematic for a SSG with a conductor containing two identical defects. We identify two regions in the conductor, X and X'.

information gathered from only *one* SSG calculation containing just a single defect (see figure 9-7). Since this system is smaller, there is a clear advantage in terms of computational cost for the initial DFT calculation. Given the geometry of the system,
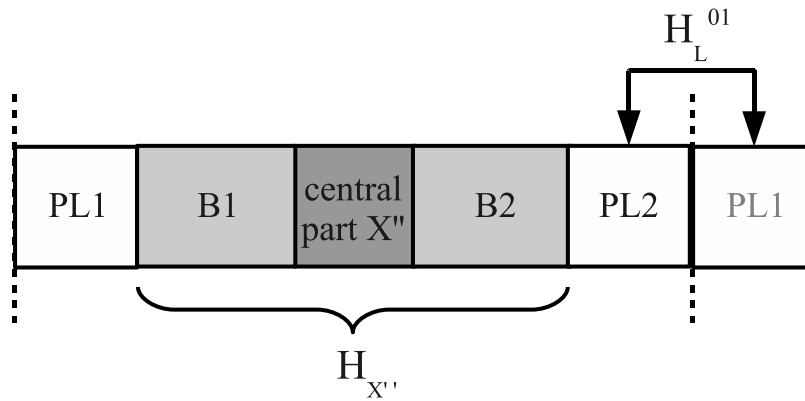
Figure 9-7: Schematic for a SSG with a conductor X'' containing one defect.

and the nearsightedness of the electronic structure, blocks H_X and $H_{X'}$ should be quite close in terms of their matrix elements. Moreover, because of the constraint that a buffer is at least as large as a lead principal layer, we expect the non-zero matrix elements of $H_{XX'}$ to correspond closely to the overlaps between the two adjacent principal layers. This observation stems from the very definition of a principal layer. As a consequence, we can construct a close approximation to H_C by using the matrices extracted from a SSG calculation of the structure shown in figure 9-7. In this approximation, blocks H_X and $H_{X'}$ are replaced with $H_{X''}$, and $H_{XX'}$ is replaced by the overlap matrix between two principal layers of lead (namely H_L^{01} from the

overlap between PL2 and PL1 in figure 9-7).

$$H_C \simeq \begin{pmatrix} H_{X''} & H_L^{01} \\ \left(H_L^{01}\right)^\dagger & H_{X''} \end{pmatrix}.$$

<div align="right">9.2</div>

The approach described above is general and may be applied to any number of isolated defects in the conductor region. In this way, Hamiltonians for systems of arbitrary size may be constructed with first-principles accuracy from one DFT calculation in a SSG with a single defect. We note in passing the importance that the MLWFs parities are consistent between different regions of the system. As mentioned in section 9.2, the parities need to be checked and made consistent to allow seamless connections between Hamiltonian sub-matrices, a task that is automatic in the present approach. Furthermore, the Hamiltonian of a conductor with more than one type of defect may be constructed by combining matrix elements from separate SSG calculations. In this latter case, care must be taken in order to align the Fermi energies of the two or more distinct calculations. This is the consequence of the lack of an absolute reference for the electrostatics in periodic boundary conditions, which can lead to Fermi energies that are shifted by a constant.

As an illustration of the excellent accuracy of the Large Scale methodology compared to the "brute force" approach of a single LCR automated calculation, we compare on figure 9-8 the predicted quantum conductance of a defected Silicon Nanowire as computed by the "direct" and "multiple defect" approaches. The observed near perfect agreement of the two curves constitutes a strong validation of the proposed methodology. A more involved application is presented in the next chapter.
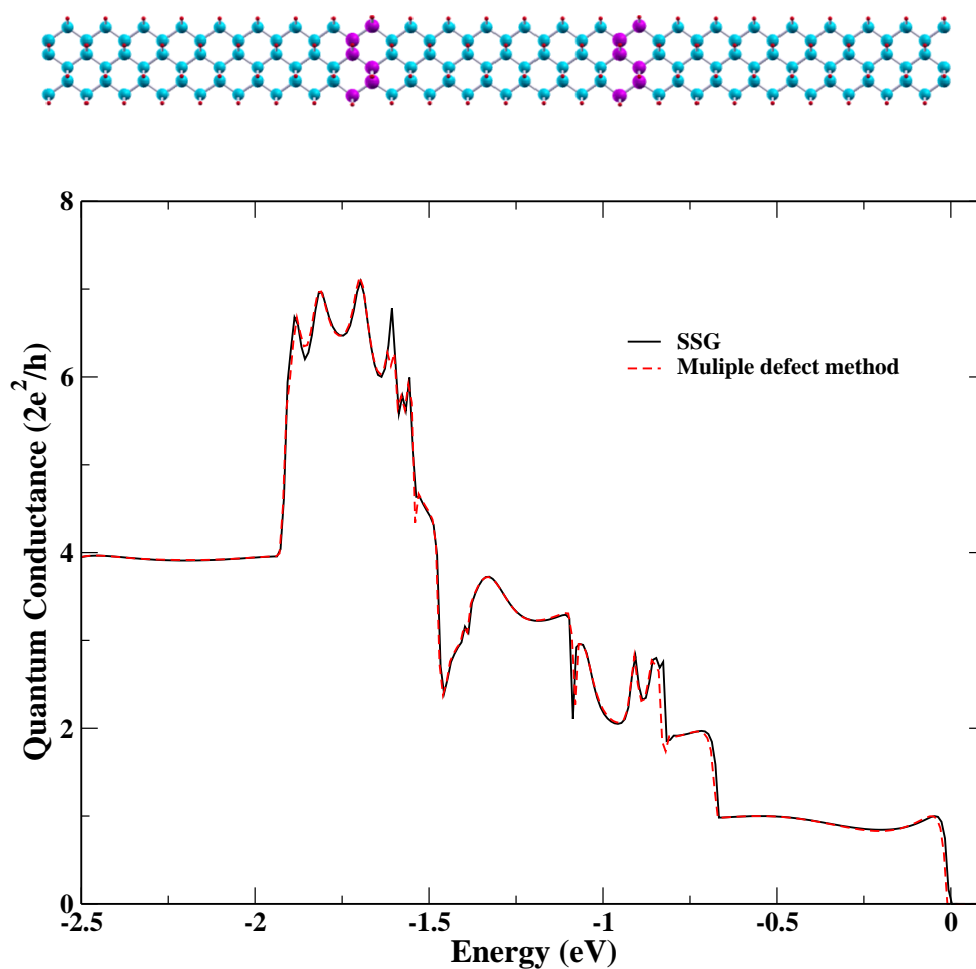
Figure 9-8: Top: Si nanowire with two Ge heterostructure defects. The system is investigated by either manipulating the Hamiltonian of a single defect (multiple defect method) or directly using a SSG. Red, cyan and magenta atoms are H, Si and Ge respectively. Bottom: Comparison of QC for the two methods, showing excellent agreement.

10

An Application of the Formalism

The goal of this chapter is to present an application of the formalism of Large Scale Quantum Transport in a "real life" situation. The problematic is related to the world of carbon nanotubes (CNTs), and more precisely the control of metallic nanotubes' conductance as a path to molecular size switch for nanoelectronic applications. We will start with an introduction to functionalized carbon nanotubes as a relevant field with endless possibilities for technological applications. We will then move on to briefly describe the fundamental competition between the strain energy of carbon nanotubes and the degree of periconjugation of organic addends leading to a possible direct control of metallic nanotubes' quantum conductance. In the last part, we will give ample details concerning the quantum conductance calculations of randomly functionalized metallic nanotubes that allowed us to demonstrate the reality of the switching mechanism. A detailed account of the work outlined in this chapter can be found in [26].

10.1 Functionalized CNTs

Single-walled carbon nanotubes (SWNTs) have been investigated for manifold applications due to their special structural, mechanical and electronic properties [18]. Chemical functionalizations of carbon nanotubes can add to their versatility serving different purposes in chemical sensing, modify the surface properties and solubilities, and facilitate the assembly, separation and purification of CNTs [11, 93, 82]. Covalent chemical functionalizations [91, 22] are especially relevant to manipulating the electronic properties of CNTs in nanoscale electric devices, such as molecular diodes or single molecular transistors.

   In all of these studies, it appears that a molecular switch with the capability of controlling CNT conductance in response to an external optical, chemical or electrical stimulus is highly desirable, and could have applications for molecular devices, chemical sensors, and imaging. Recent theoretical studies have revealed the possibility of tuning bond-cleavage chemistry of [1+2] cycloadditions on CNTs through the orientation of the unsaturated π bonds of the addend with respect to the CNT surface [59]. In previous study on the model system of dinitrocarbene-functionalized CNTs, wit was found that the bond-closed configuration is greatly stabilized when the plane of the addend π system bisects the base of the cyclopropane ring moiety. This stabiliza-

tion effect was originally attributed to the enhanced interaction between the addend π and the cyclopropane Walsh orbitals, which weakens the anti-bonding interaction of the sidewall bond [87]. The bridgehead carbon atoms can *reversibly rehybridize* from sp^2 to sp^3 in response to addend π orientation, implying a switch-like behavior. However, for dinitrocarbene-functionalized CNTs the bond-closed configuration is unstable. The addend prefers to rotate out of cyclopropane conjugation, resulting in only one stable open configuration that is impossible to manipulate. An isoelectronic carboxyl group was suggested to control bond-cleavage by the intramolecular hydrogen bond, [59] but the hydrogen bond strength is too weak to offer bistability. Our calculations have actually shown that the closed-bond configuration is a saddle point rather than a local minimum.

10.2 Switching Mechanisms in CNTs

10.2.1 Experimental and Theoretical Background

Previous studies[58, 59] have shown that simple carbenes or nitrenes bearing only saturated moieties give an open sidewall bond, leaving very little room to maneuver for switching purposes. Here we focus mostly on the interaction between unsaturated addends and the CNT surface, and more specifically on the closed-bond stabilization offered by unsaturated addends in the perpendicular orientation. When considering only the cyclopropane moiety on the functionalized armchair carbon nanotubes, the simplest rationalization for closed-bond stabilization comes from the *through-bond σ-π interaction* [87] (see figure 10-1) which withdraws electron density form the HOMO of cyclopropane to the LUMO of acetylenes, causing a decrease in the bond length of the cyclopropane base but an increase for the lateral bond lengths. However, the bond variation in substituted cyclopropane is typically smaller than 0.05 Å, suggesting this might be a less significant effect. On the other hand, a CNT is certainly a lot more complex than a cyclopropane. The sidewall bond-breaking chemistry of functionalized armchair carbon nanotubes is reminiscent of the valence tautomerism of 1,6-methano[10]annulene and the even more relevant methanofullerenes. During the search for improved electron-accepting organofullerens for photovoltaic applications, it was found that the quinone-type methanofullerenes or fluorenefullerenes that contain unsaturated moieties perpendicular to the surface of the fullerenes have a less negative first reduction potential than the parent C_{60} or other type of methanofullerenes by as much as 70 mV. The peak positions can be further tuned by electron donating or withdrawing groups attached to the addend. This phenomenon was ascribed to a *through space π-π interaction* which was called "periconjugation" [24, 49, 51]. The intramolecular electronic interaction between the π orbitals of quinone and nearby carbon atoms of C_{60}, separated by a spiro carbon atom, results in more extended conjugation, which possibly improves its electron accepting ability. As shown by the X-ray crystal structure, fullerene has an essentially [5]radialene-type electronic structure, *i.e.* the [6,6] bonds possess more double bond character while the [5,6] bonds are more single-bond like. The fact that the isolated fluorenefullerenes
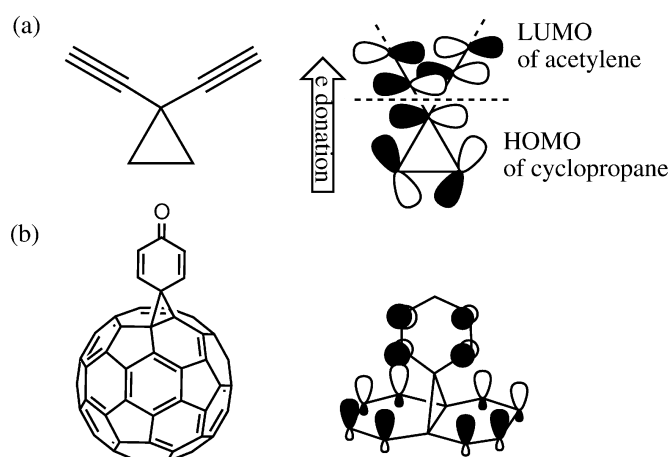
Figure 10-1: Two possible contributing effects for closed CNT sidewall bond stabilization: (a) through bond σ-π interaction in cyclopropane and (b) through space π-π periconjugation in quinone-type methanofullerene

were exclusively [6,6] fullerenes rather than [5,6] fulleroids also implies the existence of this stablizing interaction. As a relevant digression here, it is worth mentioning that unlike carbon nanotubes in which the sidewall bond lengths are largely affected by the addend identity, CNT curvature and chirality, in organofullerenes it is well known that experimentally only the kinetic [5,6]-open and the thermodynamic [6,6]-closed adducts were observed in most cases, whereas [5,6]-closed and [6,6]-open counterparts are almost never found due to the unfavorable endocyclic pentagon double bond conjugation [19]. This might be the reason why dinitrocarbene substituent does not have the same rotational bond-cleavage effect in fullerenes as in CNTs [59].

10.2.2 Periconjugation Effects for Closed Bond Stabilization

We demonstrated [26] that through-space $\pi - \pi$ conjugation effects are some of the most powerful electronic means by which to close the carbon-carbon bond on the sidewall of carbon nanotubes. As a demonstration, we plot the potential energy curve as a function of the carbon-carbon backbone distance (d_{16}) for different addends that functionalize some model naphtalene and pyrene backbone. Results are shown on figure 10-2. As can be seen clearly from the figure, for both naphtalene and pyrene backbones, the closed configuration is stabilized more when the addend π system spans longer and leans toward the aromatic plane. The stabilization strength goes in the order of $C_5O_2H_4$-diketone > $C_5O_2H_4$-dienol > C_5H_4 diene, much greater than the saturated C_5H_8 or CH_2. For the case of pyrene, the closed configuration is stabilized significantly more when X=$C_5O_2H_4$-diketone than when X=C_5H_4 by as much as 0.7 eV. This is a strong validation for the existence of the periconjugation effect as these substitutional groups are highly similar in structural motif and electronegativities. Any kind of inductive effect, if it exists, shall only differ slightly and shall not cause
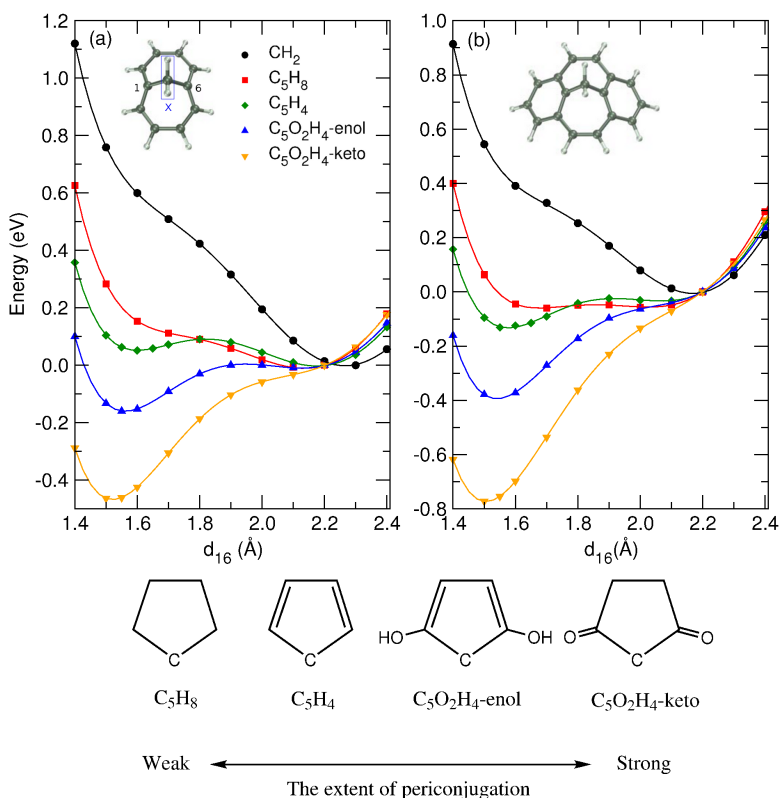
Figure 10-2: Potential energy surface as a function of d_{16} for (a) naphtalene (b) pyrene functionalized with ring substituents. The zeros are set at d_{16}=2.20 Å to mark the release of roughly the same strain energy. We observe a switch for the equilibrium carbon-carbon distance (d_{16}) from about 2.20 Å for saturated addends like CH_2 or C_5H_8 to about 1.55 Å for strong periconjugated groups like $C_5O_2H_4$ di-ketone group. In the case of a CNT backbone, the "open" conformation (for which $d_{16} \approx$ 2.20 Å) leads to a high conductance while the "closed" conformation (for which $d_{16} \approx$ 1.55 Å) leads to a much reduced conductance as we will see in section 10.3

such a dramatic stabilization. For more ample details of why periconjugation works so well we refer to [26].

10.2.3   Curvature Effects for Open Bond Stabilization

Periconjugation is a through space effect that exists between two separated π systems. As a consequence, organic addends with strong periconjugation effects will obviously interact much with the delocalized π network of carbon nanotubes, leading to a favored closed-bond configuration of the sidewall carbon atoms directly connected to the addend. However, the tautomerization between the bond-open and bond-closed form depends largely on the chirality and curvature of the CNT. In the case of zigzag and armchair CNTs, two types of C-C bonds exist: the "axial" bonds A (the C-C bonds

that are more "parallel" to the tube axis) and the "orthogonal" bonds O (the C-C bonds that lie more "perpendicular" to the tube axis). For both types of tubes, computational studies have shown that the A bond forms a closed-bond three-membered ring upon carbene cycloaddition and the O bond undergoes a sidewall bond opening upon carbene insertion [58, 8]. The reaction energy defined as $\Delta E = E_{\text{CNT-func}} - E_{\text{CNT}} - E_{\text{func}}$ increases linearly as the curvature increases for both modes of addition reactions, but the functionalized CNT is always more stable in the open O bond configuation due to release of strain than the closed A bond configuration. For the orthogonal bonds in armchair CNTs, the open bond configuration is favored in high curvature CNTs so as to release the strain, but as the tube grows larger and the curvature decreases, the closed bond configuration is gradually lowered in energy and eventually becomes more stable (for tubes larger than an (18,18)-CNT [58]). For small tubes, however, the competition between periconjugation and the curvature effect will determine the equilibrium between a closed and an open sidewall bond. We managed to identify some classes of addend pairs for which a bistability exists even on small tubes with large curvatures. In the next section we choose to study the influence of a diketone–diol pair on the conductance properties of (6,6) CNTs. We illustrate the equilibrium geometries of a (6,6) CNT functionalized with the two different addends in figure 10-3.
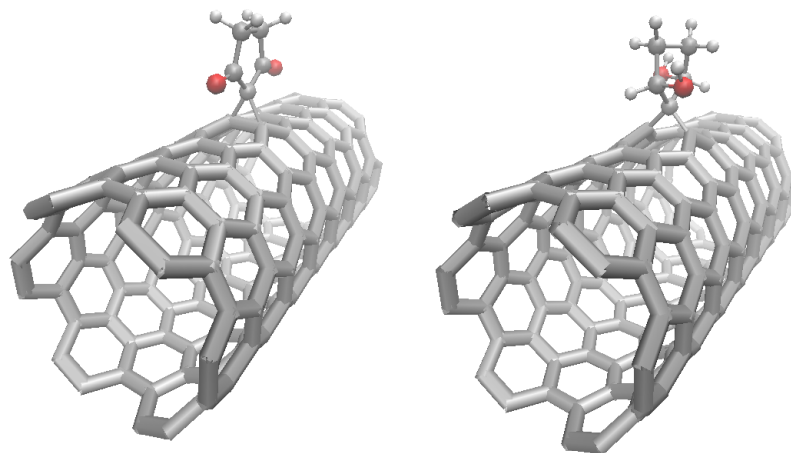
Figure 10-3: Relaxed structures of a functionalized (6,6) CNT with diketone and diol addends. The CNT sidewall bond is closed with the diketone substituent, and is open with the diol addend.

10.3 Details of Large Scale Transport Calculations

The conductivity of the proposed switches in the on and off states on CNTs were investigated. All calculations are performed using density-functional theory in the Perdew-Burke-Ernzerhof generalized-gradient approximation (PBE-GGA, see appendix A) with a plane wave basis set, periodic boundary conditions and Vanderbilt ultrasoft

pseudopotentials as implemented in the Quantum-ESPRESSO package. 30 Ry and 240 Ry cutoffs or higher were chosen for the wave functions and the charge density, respectively. The supercell included 5 principal layers of carbon atoms for a given (6,6) CNT plus the functional group(s). Each principal layer was an agglomeration of three (6,6) CNT unit cells. Structural optimization of the conductor part of the LCR supercell was performed at Γ with a cold smearing of 0.03 Ry. All components of all forces were converged to within 10^{-3} Ry/bohr, corresponding to less than 0.026 eV/Å in a relaxation calculation. The quantum conductance (or more precisely the transmission function $T(E)$) at zero bias is calculated using the Landauer formalism discussed in chapter 8 which has been fully implemented into Wannier90[1]. Figure
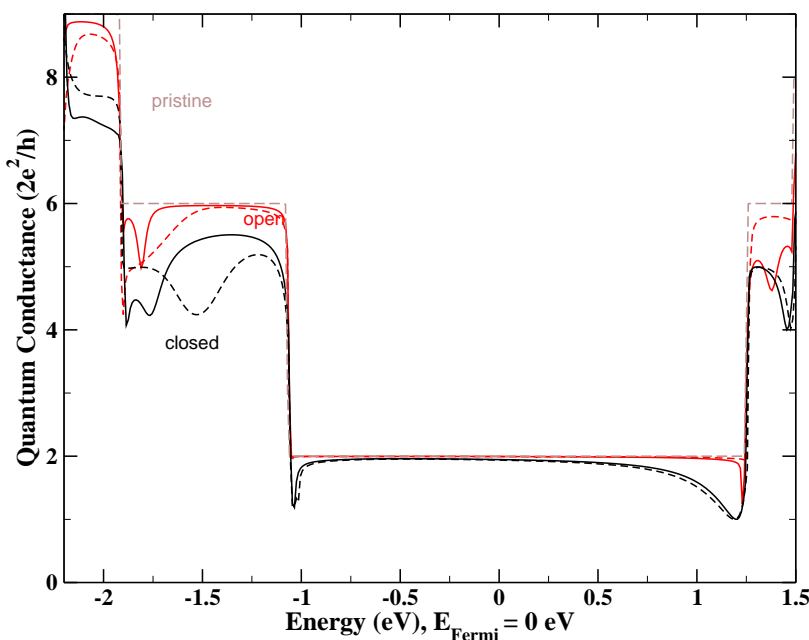
Figure 10-4: Quantum conductance of an infinitely long (6,6) carbon nanotube with a single functionalization. The solid lines correspond to the lactone/acid addend pair and the dashed lines correspond to the diketone/diol addend pair. The red color are for the "open" (acid and diol) and the black color are for the "closed" (lactone and diketone) conformations. The quantum conductance of the pristine tube is also given in the brown dashed line as a reference. The quantum conductance depends more on the "open" or "closed" conformations rather than on the specific addend identity in a wide energy range around the Fermi energy.

10-4 shows the quantum conductance of a singly functionalized (6,6) carbon nanotube with either lactone/acid or the diketone/diol addend pairs mentioned in the previous section. While the conductances for the "open" configurations (the acid and the diol) are negligibly affected and can hardly be distinguished from the pristine conductance around the Fermi level, the conductances of the "closed" configurations (the lactone

---

[1]code url : http://www.wannier.org

and the diketone) are lower. Nevertheless, the conductances at the Fermi level for both cases remain close.

However, the difference is greatly magnified in multiply functionalized CNTs. Figure 10-5 shows the quantum conductance of a 22 nm, 47 nm or 67 nm-long (6,6) CNT functionalized with 10, 20, or 30 addends, respectively, for the diketone/diol switch pair. The strong scattering of the π electrons in a window of about 2 eV around the Fermi energy in the "closed" conformation is evidenced by the increasing drop in quantum conductance with the degree of functionalization. The quantum conductance is reduced by about 13% from 1 to 10 addends, another 23% from 10 to 20 and at last another 29% from 20 to 30 addends in the "closed" (diketone) case. On the contrary, the "open" (diol) case shows only a 2% reduction from 1 to 10 addends, another 3% from 10 to 20, and another 3% from 20 to 30 addends. The average quantum conductance at the Fermi level functionalized with 30 addends in "closed" conformation drops to about 0.87 whereas it remains at about 1.84 in the "open" case. The fundamental origin of this strikingly different behavior goes back to the earliest prediction: the sidewall bond breakage in the "open" configuration preserves the π network, while in the "closed" conformation only a σ-like Wannier function is left, destroying locally the π conjugation.
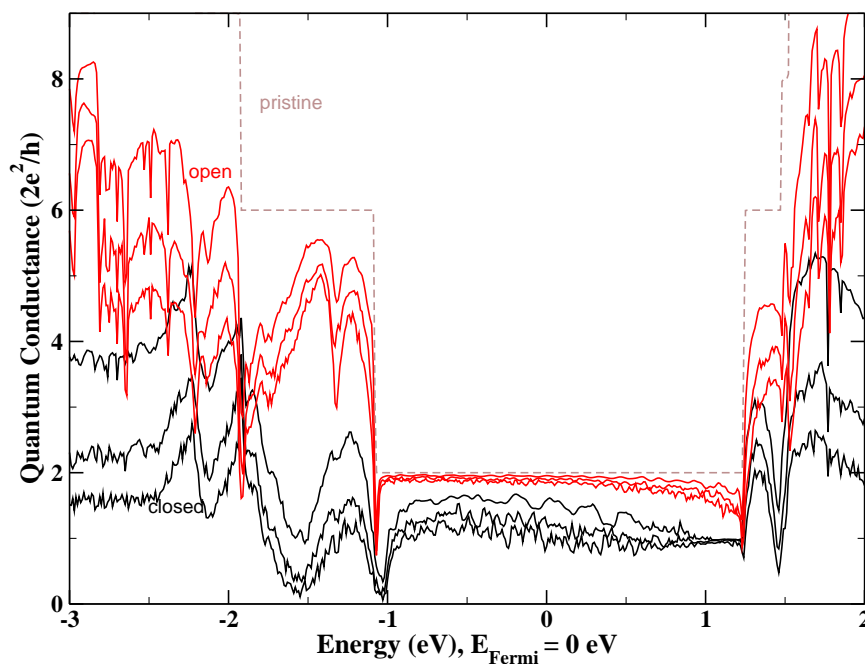
Figure 10-5: Average quantum conductance (in units of $2e^2/h$) of 20 randomly functionalized (6,6) carbon nanotubes with 10, 20 or 30 addends. Both types of addends are considered (ketone in black and diol in red). Inside a given set of conductance curves (black or red), the highest curve represents 10 addends and the lowest curve represents 30 addends. The quantum conductance of a pristine (6,6) carbon nanotube is given in dashed line as a reference.

11
Conclusions and Future Prospect

Taking the deliberate choice of presenting Density Functional Theory in a rather mathematically abstract way allowed us to naturally extend many of the known results of exact DFT to the case where the fundamental functional is orbital-dependent. This led to a clean formalism in which our novel correction scheme could find a natural place. We demonstrated both theoretically and numerically the much improved predictive power of the scheme on a wealth of physical properties, from molecular spectra and crystalline solids band structures to equilibrium conformations and chemical reaction barriers. A most important property of the correction has been to re-establish physical meaning into the Kohn-Sham orbital energies, without at the same time destroying the quality of the underlying functional on established properties. A new algorithm designed to extract ordered real-space Hamiltonian matrices allowed us for the first time to compute accurate full band structures of periodic solids from a self-consistent treatment of self-interaction correction. Although used for computing band structures and transmission functions, the obtained ordered real-space hamiltonian matrices could be used to compute Fermi energies, Fermi surfaces, or even more subtle properties like magnetic circular dichroism and more generally any physical property that one can obtain directly from the Hamiltonian matrix [103].

In a second part, we introduced novel algorithms and implementations leading to fully automated quantum conductance calculations within the Landauer formalism of coherent transport. We demonstrated the unique advantage of using a basis of Maximally Localized Wannier Functions for maximum transferability of hamiltonian matrices. This fact led to the development of a new package of computer codes allowing easy "virtual synthesis" of random or custom made model Hamiltonians of large scale complex systems. The goal here was to be able to break the fundamental "size barrier" of DFT inherent to its algorithmic cubic scaling in the number of electrons. Once again, despite the fact that those model Hamiltonians were only used for computing transmission functions, one should really see those as "exact tight-binding" mappings. Even more interesting is the prospect of merging the two approaches of self consistent Non-Koopman correction and large scale quantum transport to directly investigate the role of self-interaction correction on transport like in the case of molecular junctions [95]. The author has already started work in this exciting direction.

Appendices

A

Analytical Expression for LDA and PBE functionals

We give the analytical expression for the Exchange and Correlation energy of two representative functionals. The first is the Local Density Approximation (LDA) in its parametrization given by Perdew and Wang in 1992 [80]. The second is a Generalized Gradient Approximation (GGA) by Perdew, Burke and Ernzerhof from 1996 [78]. All of the energies will be given in *Hartree atomic units* ($\hbar = 1$, $e^2 = 1$, $m = 1$), which means that lengths are measured in *bohrs* and energies in *Hartrees*.

A.1 LDA functional in the Perdew-Wang parametrization

The exchange and correlation energy naturally splits into an exchange term only and a correlation term only. In the most general case, the "variables" entering the expression of the functionals are the **spin densities**, $n^\uparrow$ and $n^\downarrow$. We will then introduce two useful parameters, namely r_s and ζ, defined by

$$r_s = \left(\frac{3}{4\pi(n^\uparrow + n^\downarrow)} \right)^{1/3}, \quad \zeta = \frac{n^\uparrow - n^\downarrow}{n^\uparrow + n^\downarrow} \qquad \text{(A.1)}$$

Along with those parameters, we define two convenient functions

$$f(\zeta) = \frac{2g(\zeta) - 2}{2^{4/3} - 2}, \quad g(\zeta) = \frac{(1 + \zeta)^{4/3} + (1 - \zeta)^{4/3}}{2}$$

Now we can easily express the **exchange energy**

$$E_X^{\text{LDA}}[n^\uparrow, n^\downarrow] = \int_{\mathbb{T}^3} \left(n^\uparrow(\vec{x}) + n^\downarrow(\vec{x}) \right) \epsilon_x^{\text{LDA}}(n^\uparrow(\vec{x}), n^\downarrow(\vec{x})) d\vec{x}$$

with

$$\epsilon_x^{\text{LDA}}(n^\uparrow, n^\downarrow) = -\frac{3}{4\pi r_s} \left(\frac{9\pi}{4} \right)^{1/3} g(\zeta)$$

The **correlation energy** has a much more complicated expression. The energy reads

$$E_C^{\text{LDA}}[n^\uparrow, n^\downarrow] = \int_{\mathbb{T}^3} \left(n^\uparrow(\vec{x}) + n^\downarrow(\vec{x}) \right) \epsilon_c^{\text{LDA}}(n^\uparrow(\vec{x}), n^\downarrow(\vec{x})) d\vec{x}$$

with

$$\epsilon_c^{\text{LDA}}(n^\uparrow, n^\downarrow) = \epsilon_c^{\text{LDA}}(r_s, \zeta) = \epsilon_c^{\text{LDA}}(r_s, 0) + \alpha_c(r_s) \frac{f(\zeta)}{f''(0)} (1 - \zeta^4) + [\epsilon_c(r_s, 1) - \epsilon_c(r_s, 0)] f(\zeta) \zeta^4$$

In the above equation, $f''(0) \approx 1.709921$ while the functions $\epsilon_c^{\text{LDA}}(r_s, 0)$, $\epsilon_c^{\text{LDA}}(r_s, 1)$ and $\alpha_c(r_s)$ are all of the following functional form

$$h_{A,\alpha,\beta_1,\beta_2,\beta_3,\beta_4}(r_s) = -2A(1 + \alpha r_s) \log \left[1 + \frac{1}{2A \left(\beta_1 r_s^{1/2} + \beta_2 r_s + \beta_3 r_s^{3/2} + \beta_4 r_s^2 \right)} \right]$$

The values of the parameters $A, \alpha, \beta_1, \beta_2, \beta_3, \beta_4$ are given in table A.1.

	$\epsilon_c^{\text{LDA}}(r_s, 0)$	$\epsilon_c^{\text{LDA}}(r_s, 1)$	$-\alpha_c(r_s)$
A	0.031091	0.015545	0.016887
α	0.21370	0.20548	0.11125
β_1	7.5957	14.1189	10.3570
β_2	3.5876	6.1977	3.6231
β_3	1.6382	3.3662	0.88026
β_4	0.49294	0.62517	0.49671

Table A.1: Values of the parameters in the functional form for $\epsilon_c^{\text{LDA}}(r_s, 0)$, $\epsilon_c^{\text{LDA}}(r_s, 1)$ and $\alpha_c(r_s)$

A.2 Perdew-Burke-Ernzerhof (PBE) functional

The Perdew Burke Ernzerhof functional (known as PBE in the community) builds upon the LDA functional to include some known behavior of the interacting electron gas [78]. Fundamentally, the PBE functional introduces an explicit dependence of the exchange and correlation functional with respect to not only the spin densities but also their gradients in space. We then introduce the following supplementary "variables"

$$t = \frac{|\vec{\nabla}(n^\uparrow + n^\downarrow)|}{2\phi(\zeta) k_s (n^\uparrow + n^\downarrow)}, \quad s = \frac{|\vec{\nabla}(n^\uparrow + n^\downarrow)|}{2k_F(n^\uparrow + n^\downarrow)}$$

We also define

$$k_F = \left(3\pi^2 (n^\uparrow + n^\downarrow) \right)^{1/3}, \quad k_s = \left(\frac{4k_F}{\pi} \right)^{1/2}$$

and the function

$$\phi(\zeta) = \frac{(1+\zeta)^{2/3} + (1-\zeta)^{2/3}}{2}$$

Finally we have the constants

$$\beta = 0.066725, \quad \gamma = 0.0310907$$

With all of the above, the **correlation energy** reads

$$E_C^{\text{PBE}}[n^\uparrow, n^\downarrow] = \int_{\mathbb{T}^3} \left(n^\uparrow(\vec{x}) + n^\downarrow(\vec{x}) \right) \left[\epsilon_c^{\text{LDA}}(r_s, \zeta) + H(r_s, \zeta, t) \right] d\vec{x}$$

The function H has the following analytical form

$$H(r_s, \zeta, t) = \gamma \phi(\zeta)^3 \log \left[1 + \frac{\beta}{\gamma} t^2 \left(\frac{1 + At^2}{1 + At^2 + A^2 t^4} \right) \right]$$

where

$$A = \frac{\beta}{\gamma} \frac{1}{e^{-\frac{\epsilon_c^{\text{LDA}}(r_s, \zeta)}{\gamma \phi(\zeta)^3}} - 1}$$

Now we move on to the **exchange energy** which has the following form

$$E_X^{\text{PBE}}[n^\uparrow, n^\downarrow] = \int_{\mathbb{T}^3} \left(n^\uparrow(\vec{x}) + n^\downarrow(\vec{x}) \right) \epsilon_x^{\text{LDA}}(r_s, \zeta) F_x(s) d\vec{x}$$

where $F_x(s)$ is

$$F_x(s) = 1 + \kappa - \frac{\kappa}{1 + \mu s^2/\kappa}$$

along with the constants $\kappa = 0.804$ and $\mu = 0.21951$.

B

Converged Parameters for all DFT calculations

In this appendix, we will present all of the convergence tests that have been run in order to come up with a set of optimal parameters in practical calculations.

B.1 Convergence with Cutoff Parameters

In this section, we show the result of some benchmark calculations on 3 molecules, namely hydrogen H_2, methane CH_4 and ammonia NH_3 and a bulk solid, Gold. The pseudopotentials used in the molecular calculations were standard and already tested **Norm-Conserving** pseudopotentials from the *Quantum-Espresso* distribution : `H.pz-vbc.UPF`, `C.pz-vbc.UPF` and `N.pz-vbc.UPF`. pz stands for the Perdew-Zunger LDA parametrization while `vbc` stands for the type of Norm-Conserving pseudopotential, here `VonBarth-Car` type. UPF is a generic common format for storage of all pseudopotentials in Quantum-Espresso. In this thesis, except for quantum conductance calculations on functionalized carbon nanotubes where we used Ultrasoft pseudopotentials in `rrkj` format, all other calculations were performed with either the `vbc` type or the `fhi` type of Norm-Conserving pseudopotentials (`fhi` stands for Fritz Haber Institute). For all convergence analysis, we systematically tried to take each molecule in two different conformations in order to test the convergence in *energy differences* (which is what we care about in practice). Also we looked at orbital energies and sometimes lattice parameters for Gold. Molecules were inserted in a cubic supercell with a lateral size of 6 Å and we used full electrostatic corrections for both the energy and the potential [17] to simulate open boundary conditions. We used Γ sampling in this case. In the Bulk Gold case, we used a $12 \times 12 \times 12 \; \vec{k}$ mesh to converge all the energies.

B.1.1 Hydrogen Molecule

Table B.1 summarizes the result for the Hydrogen molecule H_2. As we can see, energy differences are converged to within 1 meV for a kinetic energy cutoff of about 60 Ryd (Rydbergs). For that same cutoff, the orbital energy is converged to 10 meV. The

Cutoff (Ryd)	Orbital energy ϵ_i (eV)	Energy difference ΔE (eV)
10	-16.60	-0.038
15	-16.80	-0.004
20	-16.93	-0.003
25	-17.01	-0.002
30	-17.06	0.003
35	-17.09	0.010
40	-17.12	0.016
45	-17.13	0.020
50	-17.14	0.022
55	-17.15	0.024
60	-17.15	0.025
65	-17.16	0.026
70	-17.16	0.026

Table B.1: Convergence of energy differences and orbital energies with the kinetic energy cutoff for H_2

converged kinetic and charge density cutoffs for hydrogen are then

$$\text{For hydrogen :}$$
$$E_{\text{kinetic}} = 60 \text{ Ryd}$$
$$E_{\text{charge}} = 4 \times 60 \text{ Ryd} = 240 \text{ Ryd}$$

(B.1)

The reason for choosing a charge density cutoff 4 times greater than the kinetic energy cutoff is based on the fact that for a given number of plane waves used to expand the wavefunctions $\phi_i(\vec{x})$ (that number being given by the kinetic energy cutoff), the charge density needs twice as many plane waves. Indeed if one writes a wavefunction in its plane wave expansion

$$\phi_j(\vec{x}) = \sum_{\vec{G}} c_{\vec{G}}^j e^{i\vec{G}\vec{x}}$$

then the charge density becomes

$$n(\vec{x}) = \sum_j |\phi_j(\vec{x})|^2 = \sum_j \sum_{\vec{G}\vec{G}'} \left(c_{\vec{G}}^j\right)^* c_{\vec{G}'} e^{i(\vec{G}'-\vec{G})\vec{x}}$$

Now because of the term $\vec{G}' - \vec{G}$ appearing in the charge density expression, we see that we need twice as many plane waves in its expansion. Going back to the definition of the kinetic energy cutoff

$$|\vec{G}|^2 \le E_{\text{kinetic}}$$

we see that if we double the number of plane waves for the charge density $\vec{G} \longmapsto 2\vec{G}$, then the charge density cutoff have to be 4 times the kinetic energy cutoff (because of the square on $\vec{G}$ in the above equation).

B.1.2 Methane Molecule

Now that we have a good idea of a decent kinetic energy cutoff for hydrogen, let us see what happens when we also have some carbon in the system. For this we benchmarked the same physical quantities as before (energy difference and orbital energies) for a methane molecule. The results are shown in table B.2. Similarly to

Cutoff (Ryd)	Orbital energies ϵ_i (eV)	Energy difference ΔE (eV)
10	-25.41 -16.34 -16.32 -16.31	0.044
15	-24.95 -16.55 -16.54 -16.54	0.032
20	-24.70 -16.72 -16.71 -16.71	0.031
25	-24.60 -16.82 -16.81 -16.80	0.026
30	-24.56 -16.86 -16.85 -16.85	0.016
35	-24.54 -16.88 -16.87 -16.87	0.009
40	-24.52 -16.89 -16.89 -16.88	0.006
45	-24.51 -16.91 -16.90 -16.90	0.005
50	-24.49 -16.92 -16.91 -16.91	0.006
55	-24.48 -16.93 -16.92 -16.92	0.006
60	-24.47 -16.93 -16.93 -16.92	0.006
65	-24.47 -16.94 -16.93 -16.93	0.005
70	-24.47 -16.94 -16.93 -16.93	0.005

Table B.2: Convergence of energy differences and orbital energies with the kinetic energy cutoff for CH_4

the case of hydrogen, we realize that total energy differences are converged to within 1 meV at 60 Ryd and orbital energies are converged to within 10 meV. We deduce the following for the converged parameters of carbon

$$\text{For carbon :}$$
$$E_{\text{kinetic}} = 60 \text{ Ryd}$$
$$E_{\text{charge}} = 4 \times 60 \text{ Ryd} = 240 \text{ Ryd}$$
(B.2)

B.1.3 Ammonia Molecule

In order to close this little convergence analysis, we also looked at another type of atom, namely nitrogen. We then used the ammonia molecule for benchmark. The results are shown in table B.3 The conclusions are here similar to the case of hydrogen

Cutoff (Ryd)	Orbital energies ϵ_i (eV)	Energy difference ΔE (eV)
10	-33.70 -18.74 -18.73 -13.41	-0.123
15	-32.13 -19.01 -19.01 -13.57	-0.069
20	-31.17 -19.25 -19.23 -13.76	-0.069
25	-30.64 -19.40 -19.39 -13.86	-0.056
30	-30.30 -19.49 -19.49 -13.92	-0.035
35	-30.03 -19.55 -19.55 -13.96	-0.010
40	-29.83 -19.58 -19.58 -14.00	0.007
45	-29.69 -19.60 -19.60 -14.03	0.012
50	-29.58 -19.62 -19.61 -14.05	0.013
55	-29.50 -19.63 -19.63 -14.07	0.013
60	-29.45 -19.64 -19.64 -14.08	0.013
65	-29.41 -19.65 -19.65 -14.09	0.013
70	-29.39 -19.65 -19.65 -14.09	0.014

Table B.3: Convergence of energy differences and orbital energies with the kinetic energy cutoff for NH_3

and carbon. Once again, energy differences are within 1 meV and orbital energies to within 10 meV (except for the deepest state at about 50 meV) for a kinetic energy cutoff of 60 Ryd. The parameters for nitrogen are again

For nitrogen :
$$E_{\text{kinetic}} = 60 \text{ Ryd}$$
$$E_{\text{charge}} = 4 \times 60 \text{ Ryd} = 240 \text{ Ryd}$$
(B.3)

B.1.4   Gold Convergence

Since we used Gold for one of our transport calculation we decided to compare the accuracy of our Norm-Conserving pseudopotential `Au.LDA.fhi.UPF` to a well tested Ultrasoft pseudopotential `Au.pz-d-rrkjus.UPF`. We looked at the convergence of the Fermi level, total energy and orbital energies for the two pseudopotentials. The results are gathered in tables B.4 and B.5.   The Ultrasoft pseudopotential converges energies and orbital energies to within 6 meV for a kinetic energy cutoff of 35 Ryd. We mention at this point that we used a charge density cutoff of 9 times the kinetic energy cutoff in the Ultrasoft case. The Norm-Conserving pseudopotential leads us to use a kinetic energy cutoff of about 45 Ryd to obtain the same accuracy

For Gold :
$$E_{\text{kinetic}} = 45 \text{ Ryd}$$
$$E_{\text{charge}} = 4 \times 45 \text{ Ryd} = 180 \text{ Ryd}$$
(B.4)

Cutoff (Ryd)	Fermi Level (eV)	Total energy (eV)	Orbital energies (eV)
20	-5.0629	-907.41097320	-12.225 -7.201 -7.201 -6.919
25	-5.0624	-907.43464671	-12.223 -7.203 -7.203 -6.916
30	-5.0622	-907.46617628	-12.221 -7.204 -7.204 -6.914
35	-5.0624	-907.47146958	-12.221 -7.204 -7.204 -6.915
40	-5.0624	-907.47299179	-12.221 -7.204 -7.204 -6.915
45	-5.0624	-907.47534788	-12.221 -7.204 -7.204 -6.914
50	-5.0624	-907.47624708	-12.221 -7.204 -7.204 -6.914
55	-5.0624	-907.47635076	-12.221 -7.204 -7.204 -6.914
60	-5.0624	-907.47666859	-12.221 -7.204 -7.204 -6.914
65	-5.0624	-907.47699281	-12.221 -7.204 -7.204 -6.914

Table B.4: Convergence of Fermi level, total energy and orbital energies with the kinetic energy cutoff for a Gold chain with our reference Ultrasoft pseudopotential

Cutoff (Ryd)	Fermi Level (eV)	Total energy (eV)	Orbital energies (eV)
15	-5.1905	-844.64252248	-13.593 -7.886 -6.918 -6.498
20	-5.0169	-883.04302239	-12.424 -7.160 -7.114 -7.073
25	-5.0557	-895.43530307	-12.376 -7.279 -7.275 -6.977
30	-5.0587	-898.46992749	-12.242 -7.248 -7.243 -6.891
35	-5.0513	-899.02505821	-12.224 -7.247 -7.247 -6.879
40	-5.0502	-899.09104693	-12.223 -7.246 -7.246 -6.878
45	-5.0502	-899.09335174	-12.223 -7.246 -7.246 -6.878
50	-5.0502	-899.09596362	-12.223 -7.246 -7.246 -6.878
55	-5.0502	-899.10057378	-12.223 -7.246 -7.246 -6.878
60	-5.0502	-899.10463576	-12.223 -7.246 -7.246 -6.878
65	-5.0502	-899.10745690	-12.223 -7.246 -7.246 -6.877
70	-5.0502	-899.10920414	-12.222 -7.246 -7.246 -6.877
75	-5.0502	-899.11021546	-12.222 -7.246 -7.246 -6.877
80	-5.0501	-899.11074213	-12.222 -7.246 -7.246 -6.877

Table B.5: Convergence of Fermi level, total energy and orbital energies with the kinetic energy cutoff for a Gold chain with our Norm-Conserving pseudopotential

We also observe an excellent match between the two pseudopotentials with comparable Fermi levels and orbital energies to within 40 meV. In a last test, we compared the prediction in the equilibrium lattice parameter of Bulk Gold (fcc bravais lattice) with experiment. The results are shown on figure B-1.
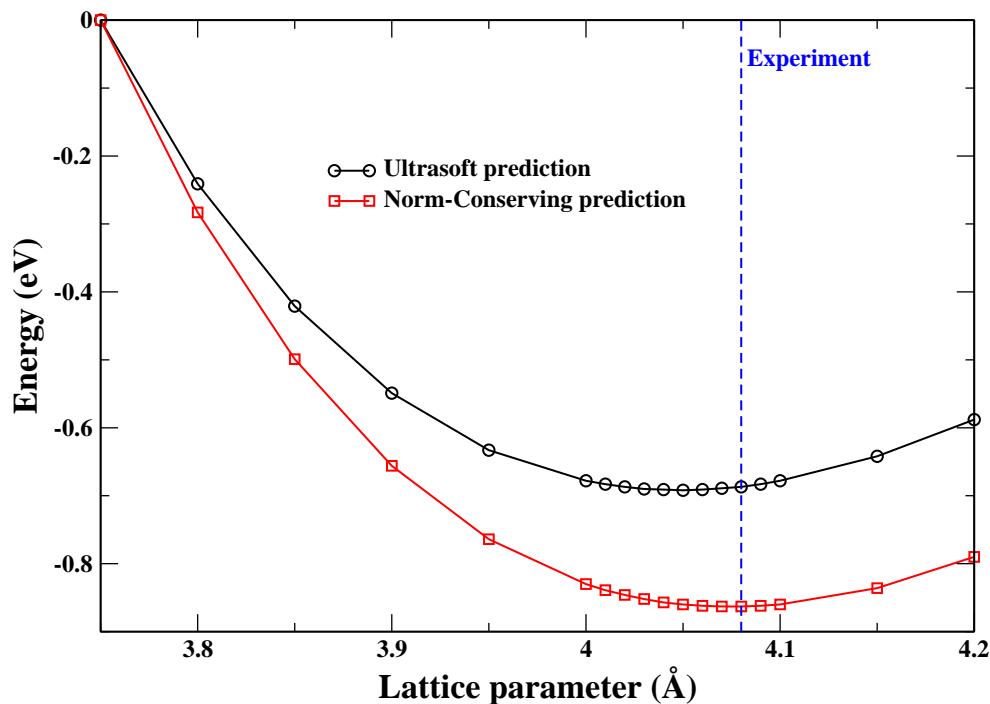
Figure B-1: Comparing the pseudopotential's equilibrium lattice parameter for fcc Gold to experiment. We observe an excellent agreement to within 0.01 Å for both pseudopotentials.

The prediction is very robust in both cases (4.070 Å in the Ultrasoft case and 4.075 Å in the Norm-Conserving case compared to 4.080 Å in experiment). This validates the use of our Norm-Conserving pseudopotential in all our calculations.

B.1.5 Conclusion

In conclusion, we would like to mention that we ran some more tests on CH_2S, SO_2 and CO_2. Those tests confirmed the validity of a 60 Ryd cutoff for the kinetic energy in converging energy differences to within 5 meV and orbital energies to within 40 meV. We can then believe that those uncertainties in the energies correspond to our numerical uncertainty in all subsequent calculations.

B.2 Convergence with Supercell Size

Satisfied with our findings on cutoff's convergence, we investigated the importance of using electrostatic corrections, for a good prediction of orbital energies. To that end, we used a simple linear chain of hydrogen atoms as a test system. The results are shown on figure B-2. The conclusion is that, whenever one wants to compute
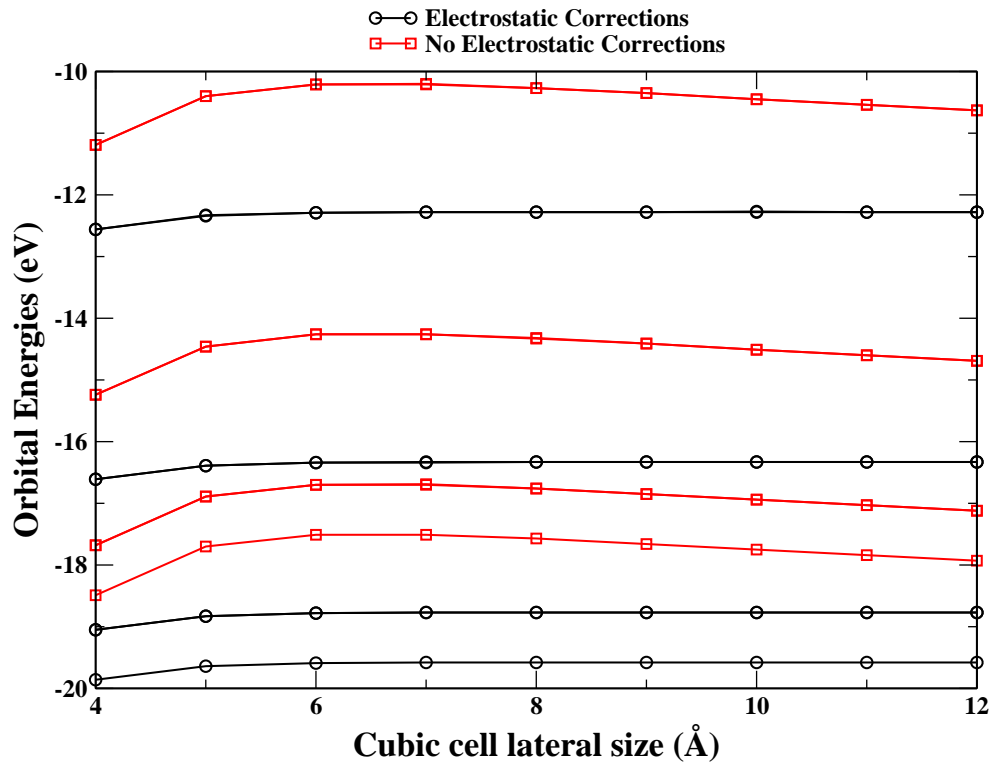
Figure B-2: Convergence of the orbital energies with the size of the cubic supercell. We observe a strikingly fast convergence when one uses electrostatic corrections, whereas in the alternative case, the energies are still not converged for a cubic cell of 12 Å.

some accurate orbitals energies, and so in particular some band structures with the supercell method of section 5.3, one needs to use electrostatic corrections. All band structures of quasi-one dimensional systems computed in this thesis made use of electrostatic corrections.

C
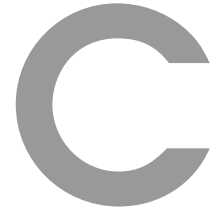
A pedestrian view of the Landauer formula

Behind its apparent simplicity, the Landauer formula [56] hides a number of amazing properties. In this appendix, we would like to explore these properties by first rediscovering the formula on simple exactly solvable models of quasi-one dimensional systems. Secondly we will generalize the formula to a generic quasi-one dimensional periodic system. In the last part we will introduce a central region with a disordered potential to model defects and see how the formula is transformed.

C.1 The fundamental approximations in Landauer theory

The Landauer formula has been established in the framework of "independant electron" theory or "mean-field" theory. This means that one describes the electron-electron interactions (and any two-body interactions for that matter) by the introduction of a mean-field one-body potential in the Schrödinger equation. Many "mean-field" theories have emerged over the span of the 20th century like Hartree theory [1], Hartree-Fock theory [92], band theory [50] and later Density Functional Theory [41]. A generalization to the Landauer formula that takes into account the full two-body electron-electron interaction and more has been established by Y. Meir and N. S. Wingreen [66].

The fundamental interpretation of Landauer conduction goes as follows. A quasi-one dimensional system consists of two semi-infinite periodic sub-systems called "leads" and a central region to which the two leads are connected. This central region will be called "conductor". The other end of the leads (the one not connected to the conductor) is connected "at infinity" to a macroscale "reservoir" at thermodynamic equilibrium with a given chemical potential μ and temperature T. The basic setup for this "lead-conductor-lead" system
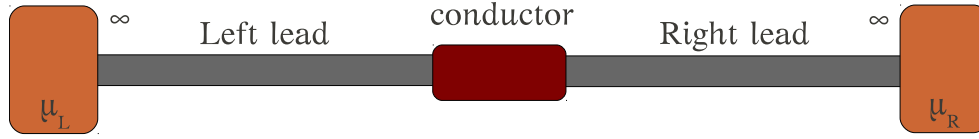
Rolf Landauer (1927-1999)

Figure C-1: The basic setup for a typical "lead-conductor-lead" quasi-one dimensional system.

is shown on figure C-1. Given this setup, electrons are supposed to be injected at one end of the system by one of the reservoir with a given energy E. The probability for an electron to be injected with energy E into the quasi-one dimensional system by the left lead is given by equation C.1.

$$\text{probability of injection at energy E by left lead} \; = \frac{1}{1 + e^{\frac{E - \mu_L}{kT}}} \qquad \text{(C.1)}$$

where μ_L is the chemical potential of the left reservoir. A similar equation holds for the right lead. Once injected, electrons propagate without any inelastic scattering event, which means that they keep their energy unchanged. As a consequence, the wavefunction of the electron is an eigenstate of the whole system for energy E. The central region introduces elastic scattering in the sense that an electron may be partially reflected and partially transmitted through that region, but nevertheless keeps its energy E unchanged. The net current flowing through the system is then just the algebraic sum of the currents carried by electrons moving from one reservoir to the other.

We will show in the next sections, how one can compute this net current and then move on to generalize the formula to a generic quasi-one dimensional system as shown on figure C-1.

C.2 A simple model of quasi-one dimensional lead

C.2.1 Description of the model

Let us here consider the simplest possible lead we can imagine. It will consist of a waveguide with rectangular cross section in the (x,y) plane of dimensions a and b respectively, and a long dimension L ($a \ll L, b \ll L$) in the z direction as shown on figure C-2. We impose a confinement potential in the x and y directions, such that the electron is free to move inside the waveguide, and the wavefunction vanishes outside it. In the extended direction (z), the electron is free to move and the overall wavefunction will be normalized over the volume of the waveguide. We will choose the origin of the coordinate system at the center of the (x,y) cross section. The Schrödinger

---

[1]Earliest work by Douglas R. Hartree on his now famous Hartree equations dates back to 1927.
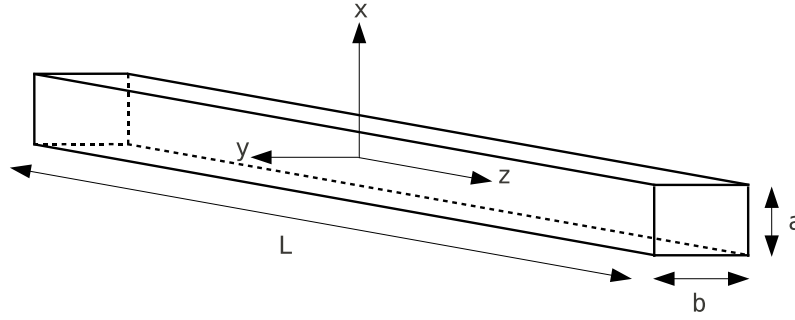
Figure C-2: A simple quasi-one dimensional waveguide. The electron's wavefunction is supposed to vanish outside the waveguide (in the x and y directions), and is unconstrained in the z direction.

equation to solve is then :

$$-\frac{\hbar^2}{2m_e}\vec{\nabla}\psi(\vec{x}) = E\psi(\vec{x})$$
<div style="text-align: right">(C.2)</div>

In choosing our eigensolutions, we will be guided by our physical intuition. In particular, actual physical solutions should represent propagating electron waves. This in turn, directs us to choose eigensolutions that are also eigenfunctions of the momentum operator in the z direction : $\hat{p}_z$.

C.2.2   Mathematical solution to the model

With the above mentioned hypotheses, one can easily solve for the mathematics of the problem. The boundary conditions introduce a set of 3 quantum numbers (n, m and k_z), where n and m are strictly positive integers, and k_z is a real number. The eigen-energies and eigen-functions are given by :

<div style="text-align: right">Eigen-energies   (C.3)</div>

$$E_{n,m,k_z} = \frac{\hbar^2}{2m_e}\left(n^2\frac{\pi^2}{a^2} + m^2\frac{\pi^2}{b^2} + k_z^2\right) \text{ with } (n,m) \in \mathbb{N}^2, k_z \in \mathbb{R}$$

<div style="text-align: right">Eigen-functions   (C.4)</div>

$$\psi_{n,m,k_z}(x,y,z) = \frac{1}{\sqrt{L}}f_{n,m}(x,y)e^{ik_z z}$$

<div style="text-align: right">with :</div>

$$f_{n,m}(x,y) = \begin{cases} \frac{2}{\sqrt{ab}}\cos(n\pi\frac{x}{a})\cos(m\pi\frac{y}{b}) & \text{if } n \text{ and } m \text{ are even} \\ \frac{2}{\sqrt{ab}}\cos(n\pi\frac{x}{a})\sin(m\pi\frac{y}{b}) & \text{if } n \text{ is even and } m \text{ odd} \\ \frac{2}{\sqrt{ab}}\sin(n\pi\frac{x}{a})\cos(m\pi\frac{y}{b}) & \text{if } n \text{ is odd and } m \text{ even} \\ \frac{2}{\sqrt{ab}}\sin(n\pi\frac{x}{a})\sin(m\pi\frac{y}{b}) & \text{if } n \text{ and } m \text{ are odd} \end{cases}$$
<div style="text-align: right">(C.5)</div>

The above eigenfunctions clearly represent normalized wavefunctions of right moving and left moving electrons. The solutions for which $k_z > 0$ represent right moving electrons while the corresponding $k_z < 0$ solutions represent left moving electrons. This can easily be verified by computing the action of the z component of the momentum operator onto those eigensolutions. We have :

$$\hat{p}_z \psi_{n,m,k_z}(x,y,z) = \hbar k_z \psi_{n,m,k_z}(x,y,z)$$

C.2.3 Elementary current carried by eigenfunctions

Now that we have the analytical expression for the eigenfunctions, it is an easy matter to determine the current density carried by electrons. Using the basic quantum mechanics formula for the probability current operator [2], the current density is :

$$\vec{j}\psi_{n,m,k_z}(x,y,z) = -e\frac{\hbar}{2im_e}\Big(\psi^*_{n,m,k_z}(x,y,z)\vec{\nabla}\psi_{n,m,k_z}(x,y,z)$$
$$-\psi_{n,m,k_z}(x,y,z)\vec{\nabla}\psi^*_{n,m,k_z}(x,y,z)\Big) \qquad \text{(C.6)}$$

Inserting the expression for the eigensolution into equation C.6 leads us to :

$$\vec{j}\psi_{n,m,k_z}(x,y,z) = -\frac{e\hbar k_z}{m_e}\frac{1}{L}\begin{pmatrix} 0 \\ 0 \\ |f_{n,m}(x,y)|^2 \end{pmatrix}$$

which, when integrated in the x and y directions, gives us the total current :

$$\vec{j}\psi_{n,m,k_z}(x,y,z) = \frac{-e}{L}\frac{\hbar k_z}{m_e}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad \text{(C.7)}$$

It is quite interesting at this point to realize that absolutely no approximations have been made to arrive at that result. Moreover the formula can easily be interpreted in terms of classical observables like velocity and charge density. Indeed equation C.7 basically says that the elementary quantum mechanical total current is nothing but the linear charge density $\frac{-e}{L}$ multiplied by the z-component of the velocity $\frac{\hbar k_z}{m_e}$.

C.2.4 Counting the number of eigenstates at energy E

If one wants to compute the total current that can be carried through the waveguide around a given energy E, then basically two ingredients are required. One is the elementary current carried by one eigenstate (determined in the previous section), and the second is the number of eigenstates that the waveguide can support between energies E and E+dE. In order to determine that latter, one needs to know how many possible triplets of quantum numbers (n,m,k_z) have an energy E (given by

---

[2]Look for example at http://en.wikipedia.org/wiki/Probability_current

$\frac{\hbar^2}{2m_e}\left(n^2\frac{\pi^2}{a^2} + m^2\frac{\pi^2}{b^2} + k_z^2\right)$) that sits between E and E+dE. Given the lateral confinement of the waveguide, only the last quantum number, k_z, can be continuous. Using boundary conditions compatible with propagating electron waves (here periodic boundary conditions), one can determine the allowed values for k_z.

$$\text{Allowed } k_z \text{ values are such that } e^{ik_zL} = 1 \qquad \text{(C.8)}$$

which shows that $k_z = 0, \pm\frac{2\pi}{L}, \pm\frac{4\pi}{L}, \pm\frac{6\pi}{L}, ...$ What this result teaches us is that neighboring k_z quantum numbers are separated by a distance $\frac{2\pi}{L}$ in k space. Now if we go back to equation C.3, we see that k_z and E are related through :

$$E = E_{n,m} + \frac{\hbar^2 k_z^2}{2m_e} \text{ with } E_{n,m} = \frac{\hbar^2}{2m_e}\left(n^2\frac{\pi^2}{a^2} + m^2\frac{\pi^2}{b^2}\right)$$

Taking E and k_z to be continuous variables leads us to :

$$dE = \frac{\hbar^2}{m_e}k_z dk_z \text{ for all allowed quantum numbers } (n,m)$$

We end up with a direct relationship between an infinitesimal amount of energy dE and an infinitesimal amount of wavevector dk_z. By "allowed" quantum numbers (n,m), we mean all (n,m) such that there exist a k_z for which $E = E_{n,m} + \frac{\hbar^2 k_z^2}{2m_e}$. Remembering that we established the "distance" between two k_z quantum numbers to be given by $\frac{2\pi}{L}$, we then arrive at the desired number of possible states in the energy window dE :

$$
\begin{aligned}
\text{number of states between E and E+dE, } N(E)dE \quad &= \quad \frac{|dk_z|}{\frac{2\pi}{L}} \\
&= \quad \frac{Lm_e}{2\pi\hbar^2}\frac{dE}{|k_z|}
\end{aligned}
$$
$$\text{for all allowed quantum numbers } (n,m) \qquad \text{(C.9)}$$

C.2.5   Total current between E and E+dE

Using the results of the two previous sections, it is quite easy to compute the total contribution to the current from electrons with an energy between E and E+dE. All one needs to do is to multiply the number of possible allowed states by the elementary contribution to the current by one eigenstate. All in all, the current is :

$$\vec{J}(E)dE = \sum_{\text{allowed } (n,m)} -\frac{e\hbar k_z}{m_e}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\frac{Lm_e}{2\pi\hbar^2}\frac{dE}{|k_z|}$$

After simple algebraic manipulations, the above expression reduces to :

$$\vec{J}(E)dE = \sum_{\text{allowed } (n,m)} -\frac{e}{h} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \text{sign}(k_z)dE \qquad \text{(C.10)}$$

What a remarkably simple formula! It says in substance, that for a given energy window dE, the total current is given by the number of allowed transverse quantum states (indexed by (n,m)) multiplied by the same current contribution $-\frac{e}{h}\text{sign}(k_z)dE$.

C.2.6 Landauer formula for the simple waveguide

With the result of the previous section, it is now an easy task to compute the total net current flowing through the system. We just have to algebraically add the total current carried by electrons with energy E and coming from the left reservoir with probability $f_L(E)$ and the total current carried by electrons with energy E and coming from the right reservoir with probability $f_R(E)$. The result reads

$$J = \int_{-\infty}^{+\infty} \left[f_L(E)\vec{J}_L(E) + f_R(E)\vec{J}_R(E) \right] dE = \frac{e}{h}\int_{-\infty}^{+\infty} [f_R(E) - f_L(E)]N(E)dE \qquad \text{(C.11)}$$

where $N(E)$ is the total number of "open channels" at energy E in the waveguide (i.e. the number of allowed transverse quantum states at energy E in the waveguide). This beautifully simple formula is called the Landauer formula, in honor of Rolf Landauer who first arrived at it back in the 1970s. The interpretation of that formula goes as follow. For an energy E, there is N(E) quantum states that the waveguide can carry. So the net current will be given by that number multiplied by the net probability for an electron to go through the system given solely by the difference in Fermi functions of the two macroscopic reservoirs.

C.3 A general quasi-one dimensional lead

In building up the complexity of the lead, we now introduce a general quasi-one dimensional lead instead of the simple waveguide of the previous section. To picture such a lead, just think of a general Silicon nanowire or a Carbon nanotube. In this case the system still possesses some translation symmetry in the z direction. But the confinement potential is not constant anymore. Instead we introduce a general potential function $V(\vec{x})$ that satisfies the z-periodicity condition

$$V(\vec{x} + Ma\hat{z}) = V(\vec{x}), \text{ with } M \in \mathbb{Z} \text{ and } a \text{ the periodicity in z}$$

Since we have a periodic potential function, we know from Part II that the eigenfunctions of the quasi-one dimensional lead can be written as

$$\psi_{nmpk_z}(\vec{x}) = \xi_{nmpk_z}(x,y,z)e^{ik_z z} \qquad \text{(C.12)}$$

where $\xi_{nmpk_z}(\vec{x})$ is a periodic function in z. n and m are the two quantum numbers associated with the transverse quantum states[3], while p is the band index labeling the Bloch states for a given value of k_z in the Brillouin Zone $]-\frac{\pi}{a},\frac{\pi}{a}]$. Associated with those eigenfunctions we have a set of eigenenergies $E_{nmp}(k_z)$ which satisfy the Schrödinger equation

$$\hat{H}\psi_{nmpk_z}(\vec{x}) = E_{nmp}(k_z)\psi_{nmpk_z}(\vec{x})$$

with $\hat{H} = -\frac{\hbar^2}{2m}\vec{\nabla}^2 + V(\vec{x})$ and the periodic part of the Bloch functions ξ_{nmpk_z} satisfy the k_z-dependent equation

$$\hat{H}_{k_z}\xi_{nmpk_z}(\vec{x}) = E_{nmp}(k_z)\xi_{nmpk_z}(\vec{x}) \qquad \boxed{\text{C.13}}$$

where $\hat{H}_{k_z}$ is the following hamiltonian

$$\hat{H}_{k_z} = -\frac{\hbar^2}{2m}\left(\vec{\nabla} + ik_z\hat{z}\right)^2 + V(\vec{x})$$

C.3.1 Conservation of the current

Since we are dealing with a system in steady state (i.e. independent of time), then a simple theorem of Quantum Mechanics tells us that the current $\vec{j}$ verifies the following identity

$$\vec{\nabla}.\vec{j} = 0$$

This equation is true for *any* current computed from an **eigenfunction** of the system with a definite (eigen)energy E. Using the divergence theorem[4], we can integrate the above differential equation over a volume V that represents a cylinder of radius R much greater than the diameter of the quasi-one dimensional system and of length Δz in the z direction.

$$\int_V \vec{\nabla}.\vec{j} = \int_S \vec{j}.d\vec{S} = 0$$

Knowing that any eigenfunction of the system will decay exponentially at large distance from the system's axis we see that the contribution to S that comes from the lateral surface of the cylinder will vanish. The consequence of that, is that we can prove the following identity

$$I_z = \int_{\mathbb{R}^2} j_z dx dy \text{ does not depend on } z$$

What have we proved here? We just proved that the current intensity[5] in the z direction is independent of z. This seems indeed quite intuitive since the system is stationary and quasi-one dimensional in nature.

---

[3]since the system is confined in the x and y directions, there exists two integer quantum numbers that characterize the localized transverse states. See Messiah for a proof [68].

[4]see for example http://en.wikipedia.org/wiki/Divergence_theorem

[5]integrated in the x and y directions, i.e. the transverse directions of the waveguide.

C.3.2 Current and group velocity

From that observation, we will first express the current intensity in z in terms of a matrix element of the z component of the momentum operator $\hat{p}$. For this we simply use the mathematical expression for the z component of the current and find

$$I_z = \frac{1}{2m} \int_{\mathbb{R}^2} \left[\psi^* \frac{\hbar}{i} \frac{d}{dz} \psi - \psi \frac{\hbar}{i} \frac{d}{dz} \psi^* \right] dx dy$$

where ψ is an eigenfunction of the lead with a definite energy E. Now we use the fact that I_z is independent of z to replace it by its mean value

$$I_z = \frac{1}{L} \int_0^L I_z dz = \frac{1}{L} \int_0^L \frac{1}{2m} \int_{\mathbb{R}^2} \left[\psi^* \frac{\hbar}{i} \frac{d}{dz} \psi + \left(\psi^* \frac{\hbar}{i} \frac{d}{dz} \psi \right)^* \right] dx dy dz$$

But the volume $[0,L] \times \mathbb{R}^2$ is nothing else than the normalization volume for our wavefunctions, so we see that since $\frac{\hbar}{i} \frac{d}{dz}$ is the z component of the momentum operator, we end up with

$$I_z = \frac{1}{mL} \langle \psi | \hat{p}_z | \psi \rangle \qquad \text{(C.14)}$$

Let us now turn to an *a priori* totally different quantity, namely the group velocity of the band electrons. By group velocity we simply mean the k_z derivative of the band energy $E_{nmp}(k_z)$. Using the Hellmann-Feynman theorem, it is straightforward to prove the following identity

$$\frac{1}{\hbar} \frac{dE_{nmp}}{dk_z} = \frac{1}{\hbar} \langle \xi_{nmpk_z} | \frac{\partial \hat{H}_{k_z}}{\partial k_x} | \xi_{nmpk_z} \rangle$$

Using equation C.13 we arrive at

$$\frac{\partial \hat{H}_{k_z}}{\partial k_x} = \frac{\hbar}{m} \left(\frac{\hbar}{i} \frac{d}{dz} + k_z \right)$$

Injecting the expression for ξ_{nmpk_z} in terms of ψ_{nmpk_z}, we can prove the following final expression for the group velocity

$$\frac{1}{\hbar} \frac{dE_{nmp}}{dk_z} = \frac{1}{m} \langle \psi_{nmpk_z} | \hat{p}_z | \psi_{nmpk_z} \rangle \qquad \text{(C.15)}$$

Equations C.14 and C.15 give finally

$$I_z^{nmpk_z} = \frac{1}{\hbar L} \frac{dE_{nmp}}{dk_z}$$

So the contribution to the total current in the z direction due to the eigenfunction ψ_{nmpk_z} is nothing else than the group velocity of band (nmp).

C.3.3 Total current between E and E+dE

At this point we are in a position to compute the total current between energy E and E+dE in the same way we did in the previous section. We simply have to multiply the contribution to the current at energy E by the density of states at that energy. For this we use a classic result of solid state physics that gives the expression of the density of states in a 1D system as a function of the band energies

$$D_{nmpk_z}(E)dE = \frac{L}{2\pi} \frac{1}{|\frac{dE_{nmp}}{dk_z}|} dE$$

With this expression, the total current between E and E+dE is

$$\vec{J}(E).\hat{z}dE = -e \sum_{\text{allowed } nmp} I_z^{nmpk_z} D_{nmpk_z}(E)dE = -\frac{e}{h} \sum_{\text{allowed } nmp} dE$$

This is quite unexpected and amazingly simple. Just like in the case of the simple model of waveguide seen before, the total current between E and E+dE is simply proportional to the number of "open channels" at that energy multiplied by a fundamental constant.

C.3.4 Landauer formula for the general waveguide

The Landauer formula is then left *unchanged* compared to the one we obtained using a very simple model

$$J = \int_{-\infty}^{+\infty} \left[f_L(E)\vec{J}_L(E) + f_R(E)\vec{J}_R(E) \right] dE = \frac{e}{h} \int_{-\infty}^{+\infty} [f_R(E) - f_L(E)]N(E)dE$$

where $N(E)$ is just the number of "open channels" at energy E in the quasi-one dimensional system. We need to pause a little bit to realize how general the result is. It says in substance that for *any* quasi-one dimensional lead, as long as the system is periodic in the direction of propagation, the total current flowing from one reservoir to the other is basically driven by the difference in Fermi functions of the two reservoirs. We see that the key step here is the fact that the elementary contribution to the current from states (nmp) is exactly compensated by the density of states of the corresponding band. The only direct influence of the waveguide is to tell how many channels (i.e. how many quantum transverse states and how many bands for each transverse state) are open at a given energy E.

C.4 Introducing a central disordered region

Having established the Landauer formula for a general quasi-one dimensional system, we now move up the scale of complexity and introduce a *finite* central region in the waveguide that will in general break the translation symmetry, just like in figure C-1. In this case, we separate the system in a left lead, the central region and a right

lead. The fundamental observation, which comes from basic mathematics on second order linear differential equations, is that one can expand any general eigenfunction of the whole system, in terms of two linearly independent solution for a given energy E. So we see that in the left lead and the right lead, the eigenfunction of the system (i.e. the leads plus the central region) can be written as

$$|\psi(E)\rangle = A_+|\psi_+(E)\rangle + A_-|\psi_-(E)\rangle$$

where $|\psi_+(E)\rangle$ and $|\psi_-(E)\rangle$ are respectively the right moving eigenfunction of the lead and left moving eigenfunction of the lead. We actually computed those functions in the previous section on the general waveguide. Those are nothing else than the eigenfunctions of equation C.12 with $k_z > 0$ and $k_z < 0$ respectively.

In order to choose the appropriate physical solution, we use our physical intuition and decide to study an incoming wave coming from the left reservoir. Physically that wave will be scattered by the conductor and so we expect the eigenfunction of the system in the left lead to look like this

$$|\psi(E)\rangle = A_+|\psi_+(E)\rangle + A_+r|\psi_-(E)\rangle \qquad \boxed{\text{C.16}}$$

where we introduced the coefficient of reflexion r and the incoming wave amplitude A_+, that will be set to 1 in the following as a convention [6]. In the right lead, no wave is coming from the right reservoir (by construction) and so only a right moving wave should be present

$$|\psi(E)\rangle = A_+t|\psi_+(E)\rangle \qquad \boxed{\text{C.17}}$$

with a transmission coefficient t. Symmetrically we can construct a similar solution for an incoming wave from the right reservoir. Since the mathematics will be similar we will not explicitly compute the current for that solution but just state what the answer is.

C.4.1 Conservation of the current

Just like in the previous section, the system is stationary, so this means that the result of equation C.14 is still perfectly valid for any eigenfunction $|\psi\rangle$ at a given energy E. What's more, for the perfect waveguide, equation C.14 is true for *any* eigenfunction of the waveguide. So in particular, the linear combination

$$|\psi(E)\rangle = a|\psi_+(E)\rangle + b|\psi_-(E)\rangle$$

verifies

$$I_z = \frac{1}{mL}\langle\psi(E)|\hat{p}_z|\psi(E)\rangle$$

---

[6]This convention is arbitrary of course and we could totally continue our derivations keeping A_+ everywhere. The net result will be that A_+ will factor out in front of every expression and will in particular give the unit of electron flux in the system. So this is why we set it to 1 here for convenience.

for every choice of (a, b) with I_z being the constant intensity of equation C.14. From this we can demonstrate the following identities (that will become useful later on)

$$\langle \psi_+(E)|\hat{p}_z|\psi_+(E)\rangle = \langle \psi_-(E)|\hat{p}_z|\psi_-(E)\rangle = mLI_z \qquad \boxed{\text{C.18}}$$

and

$$\langle \psi_+(E)|\hat{p}_z|\psi_-(E)\rangle = \langle \psi_-(E)|\hat{p}_z|\psi_+(E)\rangle = 0 \qquad \boxed{\text{C.19}}$$

C.4.2   Total current through the system

We are now ready to compute the total current carried by an eigenfunction of the system. We just have to plug the expressions for the general solution in equations C.16 and C.17 into the formula for the current C.14 and find

$$\begin{aligned} I_z^{nmp+} = \frac{1}{mL}\langle \psi(E)|\hat{p}_z|\psi(E)\rangle &= \frac{1}{mL}\left(\langle \psi_+(E)|\hat{p}_z|\psi_+(E)\rangle + |r|^2 \langle \psi_-(E)|\hat{p}_z|\psi_-(E)\rangle \right) \\ &= \frac{1}{mL}|t|^2 \langle \psi_+(E)|\hat{p}_z|\psi_+(E)\rangle \end{aligned}$$

from which we deduce the conservation of the current $1 + |r|^2 = |t|^2$. In the above use has been made of equations C.18 and C.19. Of course the reflexion and transmission coefficients are specific to the eigenfunction of energy E and so these are in general functions of the energy E. The number of states at that energy in each lead is still the same as for the perfect waveguide. All in all, we see that the total contribution to the current from states between E and E+dE is given by

$$\vec{J}(E).\hat{z}dE = -e \sum_{\text{allowed } nmp} I_z^{nmp+} D_{nmpk_z}(E)dE = -\frac{e}{h} \sum_{\text{allowed } nmp} |t(E)|^2 dE$$

Once again, the wonderful cancellation of the density of states and the current of an eigenfunction makes the result extremely simple and intuitive. The interpretation goes as follows. Because of the disordered conductor, an incoming wave $|\psi_+(E)\rangle$ that would normally be an eigenfunction of the system is not anymore. As a consequence, some part of that wave is being reflected off the conductor. This explains the need for a reflexion coefficient r. In the right lead, some of the wave can survive and so there is in general a transmission coefficient t. The net result is that instead of transmitting the full wave, the system only transmits $|t(E)|^2$ of it. We also wish to add that in general, the reflexion and transmission coefficients depend on the quantum numbers nmp.

C.4.3   General Landauer formula with a conductor

All the hard work has now been done. We just need to algebraically sum up the different currents, to arrive at the celebrated Landauer formula

$$J = \int_{-\infty}^{+\infty} \left[f_L(E)\vec{J}_L(E) + f_R(E)\vec{J}_R(E) \right] dE = \frac{e}{h} \int_{-\infty}^{+\infty} [f_R(E) - f_L(E)] \mathscr{T}(E)dE \qquad \boxed{\text{C.20}}$$

where $\mathcal{T}(E) = \sum_{\text{allowed } nmp} |t(E)|^2$ is called the *transmission function* of the conductor. This concludes our presentation of a "scattering" approach to the Landauer formula. We saw that the results given by a simple model of lead (i.e. the simple waveguide) actually carry through all the way to a general quasi-one dimensional lead with a general band structure and a general transverse lateral confinement potential. Then the introduction of a conductor that breaks the translation symmetry only results in a slight modification of the Landauer formula through the introduction of an energy dependent transmission function $T(E)$. The key results that allowed us to establish the formula are :

- The quasi-one dimensional nature of the leads that introduces a density of states inversely proportional to the gradient of the band energies

- The stationarity condition that allows us to prove the z independence of the z component of the current density integrated in the x and y directions

- The finite influence of the conductor on the leads, i.e. sufficiently far from the conductor, we recover the solutions of the perfect lead which allows us to expand the general eigenfunction of the full system in terms of eigenfunctions of the leads.

- Of course, a major approximation has been to describe the electronic structure of the full system by a mean-field theory, i.e. by an independent electron model.

D

Exact Solution for the Ground State of H_2^+

We present an exact solution to the calculation of the Ground State of the hydrogen molecular ion H_2^+. A detailed and modern introduction to the exact study of this fundamental molecular model can be found in ref [39].

D.1 Adimensional equation

The hydrogen molecular ion (H_2^+) is the probably the simplest molecule in the Universe ! Indeed it consists in two protons, H^+, separated by a distance R and bound together by a unique electron. The two protons are taken to be on the z-axis symmetrically positioned above and below the origin at positions $z = -\frac{R}{2}$ and $z = +\frac{R}{2}$. In the Born-Oppenheimer approximation, the Hamiltonian $\hat{H}$ of the system can be written as

$$\hat{H} = -\frac{\hbar^2}{2m}\vec{\nabla}^2 - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r_1} - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r_2} + \frac{e^2}{4\pi\epsilon_0}\frac{1}{R} \tag{D.1}$$

where r_1 is the distance between the first proton and the electron, and r_2 is the distance between the second proton and the electron. The first term on the right-hand side is the electron's kinetic energy. The next two terms are the Coulomb attraction from the two protons onto the electron. At last the final term is the Coulomb repulsion of the two protons. The stationary Schrödinger equation can be written for the electron wavefunction $\psi(\vec{x})$

$$\hat{H}\psi(\vec{x}) = E\psi(\vec{x})$$

To simplify the equation, we introduce the Bohr-Hartree atomic units. For this we write the total energy E as $E = E_0\tilde{E}$, where $E_0 = m\left(\frac{e^2}{4\pi\epsilon_0\hbar}\right)^2$ is the Hartree energy, and the position as $\vec{x} = a_0\vec{r}$, where $a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2}$ is the Bohr radius. Injecting the expressions for E and $\vec{x}$ in terms of $\tilde{E}$ and $\vec{r}$, we arrive at

$$\left[-\frac{1}{2}\vec{\nabla}^2 - \frac{1}{\tilde{r}_1} - \frac{1}{\tilde{r}_2}\right]\psi(\vec{r}) = \epsilon\psi(\vec{r})$$

where $\epsilon = \tilde{E} - \tilde{R}$ ($\tilde{R} = R/a_0$, $\tilde{r}_1 = r_1/a_0$ and $\tilde{r}_2 = r_2/a_0$).

D.2 Confocal Elliptical Coordinates

The "trick" to solve the equation is to use a coordinate system that will naturally "separate" the partial differential equation (equation over 3 variables at the same time, $\vec{r}$) into three ordinary differential equations (equation over a single variable). The "magic" coordinate system is given by the confocal Elliptical coordinates. This system is defined with respect to the Cartesian coordinates by

$$x = \frac{\tilde{R}}{2}\sqrt{(\xi^2 - 1)(1 - \eta^2)}\cos(\phi) \qquad \text{(D.2)}$$

$$y = \frac{\tilde{R}}{2}\sqrt{(\xi^2 - 1)(1 - \eta^2)}\sin(\phi)$$

$$z = \frac{\tilde{R}}{2}\xi\eta$$

In this coordinate system, the Lapalcian operator $\vec{\nabla}^2$ becomes

$$\begin{aligned}
\vec{\nabla}^2 =\ & \frac{1}{\left(\frac{\tilde{R}}{2}\right)(\xi^2 - \eta^2)}\left[\frac{\partial}{\partial\xi}\left((\xi^2 - 1)\frac{\partial}{\partial\xi}\right)\right. \\
& + \frac{\partial}{\partial\eta}\left((1 - \eta^2)\frac{\partial}{\partial\eta}\right) \\
& + \left.\frac{\xi^2 - \eta^2}{(\xi^2 - 1)(1 - \eta^2)}\frac{\partial^2}{\partial\phi^2}\right]
\end{aligned}$$

The Coulomb interaction of the two protons onto the electron then becomes

$$-\frac{1}{\tilde{r}_1} - \frac{1}{\tilde{r}_2} = -\frac{2}{\tilde{R}}\frac{2\xi}{\xi^2 - \eta^2}$$

The intervals in which the coordinates vary are

- $\xi \in [1, +\infty[$

- $\eta \in [-1, 1]$

- $\phi \in [0, 2\pi[$

D.3 Using the Symmetries

In order to find what the analytical expression for the wavefunction looks like we will use the basic rotational symmetry of the system. About the z-axis, the system possesses a rotational symmetry. This means that the Hamiltonian $\hat{H}$ commutes with the z-component of the angular momentum operator $\hat{L}_z = \frac{\hbar}{i}\frac{\partial}{\partial\phi}$. As a consequence, the

wavefunction can be chosen to be an eigenfunction of $\hat{L}_z$ as well. So we end up with

$$\psi_m(\xi, \eta, \phi) = f(\xi, \eta) \frac{e^{im\phi}}{\sqrt{2\pi}} \tag{D.3}$$

where m is an integer number $0, \pm 1, \pm 2, \dots$ Injecting the expression for the wavefunction of equation D.3 into the Schrödinger equation, we can naturally separate the equation over ξ and η into two separate equations over ξ only and η only by introducing a "separation constant" γ.

D.4 Solution

After splitting the wavefunction into a product of functions over ξ and η ($f(\xi, \eta) = g(\xi)h(\eta)$), the two separate equations are

$$\frac{\partial}{\partial \xi}\left((\xi^2 - 1)\frac{\partial g(\xi)}{\partial \xi}\right) - \left[\frac{m^2}{\xi^2 - 1} - \epsilon'\xi^2 - \gamma - 2\tilde{R}\xi\right]g(\xi) = 0 \tag{D.4}$$

$$\frac{\partial}{\partial \eta}\left((1 - \eta^2)\frac{\partial h(\eta)}{\partial \eta}\right) - \left[\frac{m^2}{1 - \eta^2} + \epsilon'\eta^2 + \gamma\right]h(\eta) = 0$$

where $\epsilon' = \frac{1}{2}\tilde{R}^2\epsilon$.

The second equation in D.4 is quite well-known and is called the Spheroidal wave equation. This equation is completely equivalent to the θ part of the Schrödinger equation in the case of the Hydrogen atom. It is the generalization of the θ equation to the "angular" equation of the confocal elliptical coordinate system. Knowing that, we impose on physical grounds that $h(\eta)$ should be *finite* for all values of $\eta \in [-1, 1]$. This finiteness condition imposes some strict conditions onto the allowed values for the separation constants γ and ϵ'. For a given m, we introduce an infinite series of n values such that γ and ϵ' are related through

$$\gamma_{mn} = -\left(\epsilon'_{mn} + S[m, n, \pm\sqrt{\epsilon'_{mn}}]\right) \tag{D.5}$$

$S[m, n, z]$ being a function that gives the n-th eigenvalue of the Spheroidal equation for a given m and ϵ'. In Mathematica®, this function is called `SpheroidalEigenvalue`.

With this first relation between γ and ϵ', we now move on to the "radial" equation (i.e. the equation in ξ). From an analysis of the behavior of the differential equation for $\xi \approx 1$ and $\xi \longrightarrow +\infty$, we can write down the analytical expression for $g(\xi)$ as

$$g(\xi) = (\xi^2 - 1)^{\frac{m}{2}}(\xi + 1)^\sigma e^{-\sqrt{-\epsilon'}\xi} y\left(\frac{\xi - 1}{\xi + 1}\right)$$

where $\sigma = \frac{\tilde{R}}{\sqrt{-\epsilon'}} - m - 1$ and $y(u)$ is a smooth function of $u \in [0, 1[$. The advantage of introducing this *a priori* obscure change of function is that we can get rid of all "pathological" behaviors in $g(\xi)$ and end up with a differential equation for $y(u)$ which is

well-behaved (and so in particular much more amenable to numerical computations). The next step consists in solving for $y(u)$ by the method of power series. We write

$$y(u) = \sum_{t=0}^{+\infty} y_t u^t$$

Injecting the power series solution into the differential equation for $y(u)$ (that we don't give here), we can obtain a 3-term recurrence relation for the coefficients y_t

$$\alpha_t y_{t+1} - \beta_t y_t + \delta_t y_{t-1} = 0$$
with
$$\begin{aligned}
\alpha_t &= (t+1)(t+m+1) \\
\beta_t &= 2t^2 + (4\sqrt{-\epsilon'} - 2\sigma)t - \gamma - \epsilon' - 2\sigma\sqrt{-\epsilon'} - (m+1)(m+\sigma) \\
\delta_t &= (t-1-\sigma)(t-1-\sigma-m)
\end{aligned}$$

Introducing $F_t = \frac{y_{t+1}}{y_t}$, we can find a recurrence for F_t

$$F_{t-1} = \frac{\delta_t}{\beta_t - \alpha_t F_t}$$

which, given that $y_{-1} = 0$, gives us a continued fraction expression for $F_0 = \frac{y_1}{y_0} = \frac{\beta_0}{\alpha_0}$

$$F_0 = \frac{\beta_0}{\alpha_0} = \cfrac{\delta_1}{\beta_1 - \alpha_1 \cfrac{\delta_2}{\beta_2 - \alpha_2 \frac{\delta_3}{\beta_3 - \alpha_3 \cdots}}} \tag{D.6}$$

Let us finally reduce the problem to the search for the Ground State total energy. The Ground state being characterized by $m = 0 = n$, we can simplify the continued fraction and use equation D.5 to find an equation for $z = \sqrt{-\epsilon'}$ only. The continued fraction then gives us an equation that, once truncated to a certain given depth in the fraction tree, lets us numerically solve for ϵ' which directly gives us the Ground State energy. This is the technique that we used to find the exact Ground State potential energy curve used in section 7.3.6.

E

Definition of the MLWF Signatures

Here we detail the set of spatially-dependent integrals that we use to determine a signature for each MLWF. These signatures are used for two purposes. First, they enable a sorting algorithm to distinguish between MLWFs of different shapes with similar centers. Thus MLWFs may be ordered consistently between unit cells, a key requirement for our approach to automated transport. Secondly, signatures are used to determine the relative parity of MLWFs so that a consistent parity-pattern may also be enforced.

We begin with the integral

$$I_n(\vec{q}) = \frac{1}{V} \int_V w_n(\vec{x}) e^{i\vec{q}(\vec{x}-\vec{x}_c)} d\vec{x} \qquad \text{(E.1)}$$

where V is the volume of the supercell, $\vec{q}$ is a vector in reciprocal space and $\vec{x}_c$ is the center of Wannier function $w_n(\vec{x})$ (we assume sampling at Γ-point only). One may write $w_n(\vec{x}) = \sum_m U_{mn} u_m(\vec{x})$, where $u_m(\vec{x})$ is the periodic part of the Bloch wavefunction at band m. U_{mn} is the unitary matrix that links the Bloch basis to the Wannier basis that minimizes the spread functional. $u_m(\vec{x})$ can be written in terms of its discrete Fourier transform $\tilde{u}_m(\vec{G})$, $u_m(\vec{x}) = \sum_{\vec{G}} \tilde{u}_m(\vec{G}) e^{i\vec{G}\vec{x}}$. Thus, the integral in equation E.1 may be written as

$$I_n(\vec{q}) = e^{i\vec{q}\vec{x}_c} \sum_m U_{mn} \tilde{u}_m^*(\vec{q}), \qquad \text{(E.2)}$$

where $\vec{q}$ is a $\vec{G}$-vector of the form $l\vec{b}_1 + m\vec{b}_2 + n\vec{b}_3$, where $\{l,m,n\} \in \mathbb{Z}^3$ and $\{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$ are the reciprocal lattice vectors. Equating real and imaginary parts of equation E.1 and equation E.2, one may write

$$I_n^{\text{Re}}(\vec{q}) = \frac{1}{V} \int_V w_n(\vec{x}) \cos(\vec{q}(\vec{x}-\vec{x}_c)) d\vec{x}$$

$$= \text{Re}\left[e^{i\vec{q}\vec{x}_c} \sum_m U_{mn} \tilde{u}_m^*(\vec{q}) \right], \qquad \text{(E.3)}$$

and

$$I_n^{\text{Im}}(\vec{q}) = \frac{1}{V} \int_V w_n(\vec{x}) \sin(\vec{q}(\vec{x} - \vec{x}_c)) d\vec{x}$$

$$= \text{Im}\left[e^{i\vec{q}\vec{x}_c} \sum_m U_{mn} \tilde{u}_m^*(\vec{q}) \right]. \tag{E.4}$$

Since most DFT codes compute $\tilde{u}_m(\vec{G})$, obtaining any set of I_n incurs negligible computational expense.

The set of integrals that are used to determine a signature are given by

$$I_n = \frac{1}{V} \int_V w_n(\vec{x}) \sin^\alpha\left(\frac{2\pi}{L_x}(x - x_c)\right) \sin^\beta\left(\frac{2\pi}{L_y}(y - y_c)\right) \sin^\gamma\left(\frac{2\pi}{L_z}(z - z_c)\right) d\vec{x} \tag{E.5}$$

where $\vec{x}_c = (x_c, y_c, z_c)$, $V = L_x L_y L_z$, $\alpha, \beta, \gamma \in \{0, 1, 2, 3\}$ and $\alpha + \beta + \gamma \leq 3$. Each of the resulting 20 integrals may be written as linear combinations of those outlined in equations E.3 and E.4. The signature of the MLWF is thus given by the 20-element unit vector of these integrals. Dot products between two MLWFs' signatures reveal in a compact form their relative shape and parity.

References

[1] Instability of stable negative ions in the x[alpha] method or other local density functional schemes. *Chemical Physics Letters*, 57(4):605 – 607, 1978. 81

[2] Pseudopotential methods in condensed matter applications. *Computer Physics reports*, 9(3):115 – 197, 1989. 102

[3] Valence and core photoelectron spectroscopy of c60, buckminsterfullerene. *Chemical Physics Letters*, 176(2):203 – 208, 1991. 106

[4] Y.-S. Lee I. Souza D. Vanderbilt A. A. Mostofi, J. R. Yates and N. Marzari. Wannier90: A tool for obtaining maximally-localised wannier functions. *Comput. Phys. Commun.*, 178:685, 2008. 66

[5] C.-O. Almbladh and U. von Barth. Exact results for the charge and spin densities, exchange-correlation potentials, and density-functional eigenvalues. *Phys. Rev. B*, 31(6):3231–3244, Mar 1985. 78

[6] S. L. Altmann. *Band Theory of Solids : An Introduction from the Point of View of Symmetry*. Oxford Science Publications, 1994. 56

[7] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098–3100, Sep 1988. 82

[8] H. F. Bettinger. Addition of carbenes to the sidewalls of single-walled carbon nanotubes. *Chem. Eur. J.*, 12:4372–4379, 2006. 153

[9] E. I. Blount. Formalisms of band theory. *Solid State Physics*, 13:305–373, 1962. 51, 66

[10] Christian Brouder, Gianluca Panati, Matteo Calandra, Christophe Mourougane, and Nicola Marzari. Exponential localization of wannier functions in insulators. *Phys. Rev. Lett.*, 98(4):046402, Jan 2007. 67

[11] M. Burghard and K. Balasubramanian. Functionalization of carbon nanotubes. *Small*, 1:180–192, 2005. 149

[12] B. Diu C. Cohen-Tannoudji and F. Laloe. *Quantum Mechanics*. Wiley-Interscience, 2006. 30, 61, 124

[13] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55(22):2471–2474, Nov 1985. 86

[14] H. Chermette, I. Ciofini, F. Mariotti, and C. Daul. Correct dissociation behavior of radical ions such as h[sub 2][sup +] in density functional calculations. *The Journal of Chemical Physics*, 114(4):1447–1453, 2001. 117

[15] A. J. Coleman. Structure of fermion density matrices. *Rev. Mod. Phys.*, 35(3):668–686, Jul 1963. 39

[16] Ismaila Dabo, Andrea Ferretti, Nicolas Poilvert, Yanli Li, Nicola Marzari, and Matteo Cococcioni. Koopmans' condition for density-functional theory. *Phys. Rev. B*, 82(11):115121, Sep 2010. 85, 99

[17] Ismaila Dabo, Boris Kozinsky, Nicholas E. Singh-Miller, and Nicola Marzari. Electrostatics in periodic boundary conditions and real-space corrections. *Phys. Rev. B*, 77(11):115139, Mar 2008. 101, 161

[18] W. A. de Heer. Nanotubes and the pursuit of applications. *MRS Bull.*, 29:281–285, 2004. 149

[19] F. Diederich and C. Thilgen. Covalent fullerene chemistry. *Science*, 271:317–323, 1996. 151

[20] M. DiVentra. *Electrical Transport in Nanoscale Systems*. Cambridge University Press, 2008. 123

[21] Navit Dori, Mahesh Menon, Lennart Kilian, Moritz Sokolowski, Leeor Kronik, and Eberhard Umbach. Valence electronic structure of gas-phase 3,4,9,10-perylene tetracarboxylic acid dianhydride: Experiment and theory. *Phys. Rev. B*, 73(19):195208, May 2006. 80, 81, 107

[22] C. A. Dyke and J. M. Tour. Covalent functionalization of single-walled carbon nanotube for materials applications. *J. Phys. Chem. A*, 108:11151–11159, 2004. 149

[23] E.N. Economou. *Green's Functions in Quantum Physics, 3rd Edition*. Springer Series in Solid State Sciences, 2006. 124, 128

[24] M. Eiermann, R. C. Haddon, B. Knight, Q. C. Li, M. Maggini, N. Martin, T. Ohno, M. Prato, T. Suzuki, and F. Wudl. Electrochemical evidence for through-space orbital interactions in spiromethanofullerenes. *Angew. Chem. Int. Ed. Engl.*, 34:1591–1594, 1995. 150

[25] D. C. Elias, R. R. Nair, T. M. G. Mohiuddin, S. V. Morozov, P. Blake, M. P. Halsall, A. C. Ferrari, D. W. Boukhvalov, M. I. Katsnelson, A. K. Geim, and K. S. Novoselov. Control of graphene's properties by reversible hydrogenation: Evidence for graphane. *Science*, 323(5914):610–613, 2009. 109

[26] Nicolas Poilvert Elise Li and Nicola Marzari. Switchable conductance in functionalized carbon nanotubes via reversible sidewall bond cleavage. *ACS Nano*, 2011. 149, 151, 152

[27] Macdonald L. D. Engel E., Chevary J. A. and Vosko S. H. *Z. Phys. D*, 23(7), 1992. 82

[28] H. Englisch and R. Englisch. Exact density functionals for ground-state energies. *Phys. Stat. Sol.*, 123:711–721, 1984. 42

REFERENCES

[29] H. Eschrig. *The Fundamentals of Density Functional Theory, Second Edition*. EAG.LE, Leipzig, 2003. 29, 30, 36

[30] Helmut Eschrig. $t > 0$ ensemble-state density functional theory via legendre transform. *Phys. Rev. B*, 82(20):205120, Nov 2010. 42

[31] E. Cances et al. *Handbook of Numerical Analysis - Special Volume: Computational Chemistry*. North Holland, 2003. 40, 42

[32] Bradley P. Feuston, Wanda Andreoni, Michele Parrinello, and Enrico Clementi. Electronic and vibrational properties of c_{60} at finite temperature from ab initio molecular dynamics. *Phys. Rev. B*, 44(8):4056–4059, Aug 1991. 105, 106

[33] Daniel S. Fisher and Patrick A. Lee. Relation between conductivity and transmission matrix. *Phys. Rev. B*, 23(12):6851–6854, Jun 1981. 122, 128

[34] J. M. Foster and S. F. Boys. Canonical configurational interaction procedure. *Rev. Mod. Phys.*, 32(2):300–302, Apr 1960. 108

[35] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, 2009. 101

[36] R. W. Godby, M. Schlüter, and L. J. Sham. Self-energy operators and exchange-correlation potentials in semiconductors. *Phys. Rev. B*, 37(17):10159–10175, Jun 1988. 83

[37] S. Goedecker and C. J. Umrigar. Critical assessment of the self-interaction-corrected–local-density-functional method and its algorithmic implementation. *Phys. Rev. A*, 55(3):1765–1771, Mar 1997. 88, 101, 102

[38] Jurgen Grafenstein, Elfi Kraka, and Dieter Cremer. The impact of the self-interaction error on the density functional theory description of dissociating radical cations: Ionic and covalent dissociation limits. *The Journal of Chemical Physics*, 120(2):524–539, 2004. 88, 116

[39] Jean-Philippe Grivet. The hydrogen molecular ion revisited. *Journal of Chemical Education*, 79(1):127–132, 2002. 180

[40] O. Gunnarsson and B. I. Lundqvist. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B*, 13(10):4274–4298, May 1976. 73

[41] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, Nov 1964. 37, 168

[42] Mark S. Hybertsen and Steven G. Louie. First-principles theory of quasi-particles: Calculation of band gaps in semiconductors and insulators. *Phys. Rev. Lett.*, 55(13):1418–1421, Sep 1985. 115

[43] A. Grossman J. E. Avron and R. Rodriguez. Spectral properties of reduced bloch hamiltonians. *Annals of Physics*, 103:47–63, 1977. 59

[44] I. Jiménez, L. J. Terminello, D. G. J. Sutherland, J. A. Carlisle, E. L. Shirley, and F. J. Himpsel. Accurate valence band width of diamond. *Phys. Rev. B*, 56(12):7215–7221, Sep 1997. 114, 115, 116

[45] Benny G. Johnson, Carlos A. Gonzales, Peter M. W. Gill, and John A. Pople. A density functional study of the simplest hydrogen abstraction reaction. effect of self-interaction correction. *Chemical Physics Letters*, 221(1-2):100–108, 1994. 88

[46] W. Jones and N. H. March. *Theoretical Solid State Physics, Volume 1: Perfect Lattices in Equilibrium*. Dover Publications Inc., 1985. 52

[47] Charles W. Bauschlicher Jr., Harry Partridge, and David Ceperley. The dissociation energy of he2+. *Chemical Physics Letters*, 160(2):183 – 188, 1989. 116

[48] J. Katriel and E. R. Davidson. Asymptotic behavior of atomic and molecular wave functions. *Proc. Natl. Acad. Sci.*, 77:4403–4406, 1980. 30

[49] M. Keshavarz-K, B. Knight, R. C. Haddon, and F. Wudl. Linear free energy relation of methanofullerene c61-substituents with cyclic voltammetry: Strong electron withdrawal anomaly. *Tetrahedron*, 52:5149–5159, 1996. 150

[50] C. Kittel. *Introduction to Solid State Physics, Eighth Edition*. John Wiley and Sons, Inc, 2005. 168

[51] B. Knight, N. Martin, T. Ohno, E. Orti, C. Rovira, J. Veciana, J. Vidal-Gancedo, P. Viruela, R. Viruela, and F. Wudl. Synthesis and electrochemistry of electronegative spiroannelated methanofullerenes: Theoretical underpinning of the electronic effect of addends and a redutive cyclopropane ring-opening reaction. *J. Am. Chem. Soc.*, 119:9871–9882, 1997. 150

[52] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, Nov 1965. 38, 40, 72

[53] T. Korzdorfer. On the relation between orbital-localization and self-interaction errors in the density functional theory treatment of organic semiconductors. *The Journal of Chemical Physics*, 134(9):094111, 2011. 79

[54] B. Kramer and J. Masek. Influence of the phase coherence length on ballistic transport. *ZEITSCHRIFT FUR PHYSIK B CONDENSED MATTER*, 76(4):457–462, Apr. 1989. 122

[55] Kari Laasonen, Alfredo Pasquarello, Roberto Car, Changyol Lee, and David Vanderbilt. Car-parrinello molecular dynamics with vanderbilt ultrasoft pseudopotentials. *Phys. Rev. B*, 47(16):10142–10153, Apr 1993. 102

[56] R. Landauer. Electrical resistance of disordered one-dimensional lattices. *Phil. Mag.*, 21:863–867, 1970. 168

[57] S. Lebègue, M. Klintenberg, O. Eriksson, and M. I. Katsnelson. Accurate electronic band gap of pure and functionalized graphane from gw calculations. *Phys. Rev. B*, 79(24):245117, Jun 2009. 113, 115

[58] Y.-S. Lee and N. Marzari. Cycloaddition functionalizations to preserve or control the conductance of carbon nanotubes. *Phys. Rev. Lett.*, 97:116801, 2006. 150, 153

[59] Y.-S. Lee and N. Marzari. Cycloadditions to control bond breaking in naphthalenes, fullerenes, and carbon nanoutbes: A first-principles study. *J. Phys. Chem. C*, 112:4480–4485, 2008. 149, 150, 151

[60] Young-Su Lee. *Electronic Structure and Quantum Conductance of Nanostructures*. PhD thesis, MIT, 2006. 139

[61] Young-Su Lee, Marco Buongiorno Nardelli, and Nicola Marzari. Band structure and quantum conductance of nanostructures from maximally localized wannier functions: The case of functionalized carbon nanotubes. *Phys. Rev. Lett.*, 95(7):076804, Aug 2005. 67

[62] E. Lieb. Density functionals for coulomb systems. *International Journal of Quantum Chemistry*, 24:243–277, 1983. 29, 34

[63] B. A. Lippmann and Julian Schwinger. Variational principles for scattering processes. i. *Phys. Rev.*, 79(3):469–480, Aug 1950. 128

[64] Truhlar D.G. Lynch B.J., Zhao Y. Database of chemical reaction barriers. 120

[65] Nicola Marzari and David Vanderbilt. Maximally localized generalized wannier functions for composite energy bands. *Phys. Rev. B*, 56(20):12847–12865, Nov 1997. 64, 66

[66] Yigal Meir and Ned S. Wingreen. Landauer formula for the current through an interacting electron region. *Phys. Rev. Lett.*, 68(16):2512–2515, Apr 1992. 168

[67] H. Mera and Y. M. Niquet. Are kohn-sham conductances accurate? *Phys. Rev. Lett.*, 105(21):216408, Nov 2010. 123

[68] A. Messiah. *Quantum Mechanics*. Dover Publications, 1999. 174

[69] F. Miao, S. Wijeratne, Y. Zhang, U. C. Coskun, W. Bao, and C. N. Lau. Phase-coherent transport in graphene quantum billiards. *Science*, 317(5844):1530–1533, 2007. 122

[70] I. Souza N. Marzari and D. Vanderbilt. An introduction to maximally-localized wannier functions. *Psi-K Scientific Highlight of the Month*, 57, 2003. 66

[71] Nicola Marzari Nicolas Poilvert, Matthew Shelley and Arash Mostofi. Automated quantum conductance calculations using maximally-localized wannier functions. *Computer Physics Communications*, 2011. 138

[72] R. G. Parr and W. Yang. *Density Functional Theory of atoms and molecules*. Oxford University Press, 1994. 31

[73] Serguei Patchkovskii and Tom Ziegler. Improving "difficult" reaction barriers with self-interaction corrected density functional theory. *The Journal of Chemical Physics*, 116(18):7806–7813, 2002. 88

[74] 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(4):1045–1097, Oct 1992. 101, 138

[75] J. Perdew. Size-consistency, self-interaction correction, and derivative discontinuities in density functional theory. *Advances in Quantum Chemistry*, 21:113–134, 1990. 84, 86

[76] J. Perdew and K. Burke. Comparison shopping for a gradient-corrected density functional. *International Journal of Quantum Chemistry*, 57:309–319, 1996. 73

[77] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048–5079, May 1981. 85, 86, 101

[78] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, Oct 1996. 158, 159

[79] John P. Perdew, Robert G. Parr, Mel Levy, and Jose L. Balduz. Density-functional theory for fractional particle number: Derivative discontinuities of the energy. *Phys. Rev. Lett.*, 49(23):1691–1694, Dec 1982. 74, 78

[80] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45(23):13244–13249, Jun 1992. 73, 158

REFERENCES

[81] M. N. Piancastelli, M. K. Kelly, Y. Chang, J. T. McKinley, and G. Margaritondo. Benzene adsorption on low-temperature silicon: A synchrotron-radiation photoemission study of valence and core states. *Phys. Rev. B*, 35(17):9218–9221, Jun 1987. 105

[82] M. Prato. Controlled nanotube reactions. *Nature*, 465:172–173, 2010. 149

[83] Raffaele Resta. Quantum-mechanical position operator in extended systems. *Phys. Rev. Lett.*, 80(9):1800–1803, Mar 1998. 66

[84] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen. Fully self-consistent gw calculations for molecules. *Phys. Rev. B*, 81(8):085103, Feb 2010. 103

[85] Frank J. Ruess, Lars Oberbeck, Michelle Y. Simmons, Kuan Eng J. Goh, Alex R. Hamilton, Toby Hallam, Steven R. Schofield, Neil J. Curson, and Robert G. Clark. Toward atomic-scale device fabrication in silicon using scanning probe microscopy. *Nano Letters*, 4(10):1969–1973, 2004. 122

[86] M P Lopez Sancho, J M Lopez Sancho, and J Rubio. Quick iterative scheme for the calculation of transfer matrices: application to mo (100). *Journal of Physics F: Metal Physics*, 14(5):1205, 1984. 134

[87] L. T. Scott, M. J. Cooney, C. Otte, C. Puls, T. Haumann, R. Boese, P. J. Carroll, A. B. Smith, and A. de Meijere. Enhancement of through-space and through-bond π-orbital interactions - syntheses and properties of permethylated and perspirocyclopropanated cyclotetradeca-1,3,6,9,12-pentayne. *J. Am. Chem. Soc.*, 116:10275–10283, 1994. 150

[88] Kazuhiko Seki, Nobuo Ueno, Ulf O. Karlsson, Rainer Engelhardt, and Ernst-Eckhard Koch. Valence bands of oriented finite linear chain molecular solids as model compounds of polyethylene studied by angle-resolved photoemission. *Chemical Physics*, 105(1-2):247 – 265, 1986. 110, 111

[89] L. J. Sham and M. Schlüter. Density-functional theory of the band gap. *Phys. Rev. B*, 32(6):3883–3889, Sep 1985. 83, 84

[90] Matthew Shelley. *Theory and Simulation of Semiconducting Nanowires for Thermoelectric Applications*. PhD thesis, Imperial College, London, 2011. 122

[91] Y.-P. Sun, K. Fu, Y. Lin, and W. Huang. Functionalized carbon nanotubes: Properties and applications. *Acc. Chem. Res.*, 35:1096–1104, 2002. 149

[92] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry*. Dover, 1996. 168

[93] D. Tasis, N. Tagmatarchis, A. Bianco, and M. Prato. Chemistry of carbon nanotubes. *Chem. Rev.*, 106:1105–1136, 2006. 149

[94] Murilo L. Tiago, P. R. C. Kent, Randolph Q. Hood, and Fernando A. Reboredo. Neutral and charged excitations in carbon fullerenes from first-principles many-body theories. *The Journal of Chemical Physics*, 129(8):084311, 2008. 106

[95] C. Toher and S. Sanvito. Effects of self-interaction corrections on the transport properties of phenyl-based molecular junctions. *Phys. Rev. B*, 77(15):155402, Apr 2008. 156

[96] G. P. Tolstov. *Fourier Series*. Dover Publications, 1976. 53

[97] R. van Leeuwen and E. J. Baerends. Exchange-correlation potential with correct asymptotic behavior. *Phys. Rev. A*, 49(4):2421–2431, Apr 1994. 78

[98] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon. Quantized conductance of point contacts in a two-dimensional electron gas. *Phys. Rev. Lett.*, 60(9):848–850, Feb 1988. 122

[99] David Vanderbilt and R. D. King-Smith. Electric polarization as a bulk quantity and its relation to surface charge. *Phys. Rev. B*, 48(7):4442–4455, Aug 1993. 67

[100] Janne Viljas. Molecular electronics, a brief introduction. Technical report, Aalto University School of Science, 2011. 132

[101] Oleg A. Vydrov. *Correcting the Self-Interaction Error of Approximate Density Functionals*. PhD thesis, Rice University, 2007. 87, 108

[102] Gregory H. Wannier. The structure of electronic excitation levels in insulating crystals. *Phys. Rev.*, 52(3):191–197, Aug 1937. 63

[103] Jonathan R. Yates, Xinjie Wang, David Vanderbilt, and Ivo Souza. Spectral and fermi surface properties from wannier interpolation. *Phys. Rev. B*, 75(19):195121, May 2007. 156

[104] E. Zeidler. *Nonlinear Functional Analysis and Its Applications: Part 2 B: Nonlinear Monotone Operators*. Springer Verlag, 1989. 25