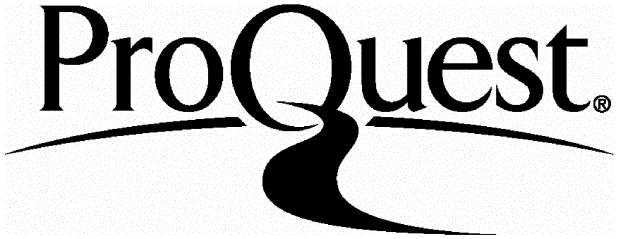
Quantum Mechanical Models of Point Defects in Gate Dielectrics

Francisco López Gejo

PhD Thesis

First supervisor: A. L. Shluger
Second supervisor: A.M. Stoneham

University College London



Abstract of the Thesis.

This Thesis studies the limits of applicability of state of the art electronic structure calculations, for predictive modelling and simulation of properties of dielectric materials. These materials play a fundamental role in microelectronics technology, as gate insulators in MOS devices.

Current devices are largely based on the properties of the Si/SiO_2 system. Large attention has been paid to the change in the energy barrier experienced by carriers at the junction, as the oxide layer becomes thinner. We have calculated the valence band offset at the Si/SiO_2 interface, directly from first-principles calculations of different models of Si/SiO_2 junction. We studied the dependence of the results on the choice of Hamiltonian and basis set, and found that the best results are obtained when using the B3LYP scheme and basis sets containing polarisation functions. We have shown that the interface states are confined in a region of the oxide whose dimensions do not depend on the thickness of the dielectric layer.

Hydrogen is believed to play a fundamental role in the processes that lead to the breakdown of the dielectric. We have calculated the stable sites for atomic hydrogen inside silica using a Density Functional Theory scheme. We found two shallow but stable minima in the same channel of the quartz structure, indicating that some type of interaction exists between the impurity and the host. We found that Hydrogen actually becomes polarised. The results, however, are critically affected by two of the approximations of the method: the approximative treatment of the exchange-correlation interaction, and the classic treatment of hydrogen nuclei. We have also calculated the isotropic hyperfine interaction parameters for H inside quartz, that can be compared directly with data from EPR experiments. The results, however, appeared to be very sensitive to the quality of the basis set.

Materials with higher dielectric constant than that of SiO_2 are being considered as alternative gate dielectrics. Their reliability of the material, however, depends critically on how easily defects can be generated, and on the ability of these defects to trap charge carriers that can then tunnel through the material. We have studied the energies of formation and the ionization potential and electron affinities of cation and anion vacancies, as well as substitutional Zr inside HfO_2. All these properties are defined as ground-state properties, and corrected for the underestimation of the band gap, typical of the Density Functional Theory scheme employed. Energy of formation of O vacancies is much lower than that of cation vacancies. The O vacancies are shown to be able to trap electrons when positively charged, and holes when neutral.

First-principle techniques are very computationally demanding methods. Alternatively, semi-empirical methods can be employed to explore the properties of systems whose sizes are out of the range of the ab initio schemes. We have updated an implementation of the INDO method in order to allow the study of systems containing several hundreds of atoms in single processor machines. A set of parameters for studying compounds containing Si, O, N and H using this technique has also been developed and tested. It allows the description of bulk structures of Si/SiO_2, Si_3N_4 and SiON with differences of 0.1 Å and 5° with respect to experimental bond-lengths and angles, respectively.

Contents

List of Figures

List of Tables

Acknowledgements.

I am very thankful towards whoever it was that invented the rule by which only my name appears in golden letters in the side of this book. This text is actually the final product of a wide collective effort. My merit has been, at the most, trying to survive my PhD studies keeping as much mental health as possible (not that I have succeeded). I will try to make some justice here, but be aware that lists of names are in random order, and most probably incomplete. Name repetition implies contribution in several battlefronts.

Many people of the CMMP group at UCL have guided me along the dark path that took me to discover that electrons are not little green balls. All of them stood stoically for many questions and discussions, and spent part of their time and patience on providing me with answers. The list includes Professors M. Stoneham and M. Gillan, and Doctors G. Gavartin, V. Sulimov, P. Shusko, and D. Alfè. I keep a special regard for A. Livshitz and L. Kantorovich who, apart from many scientific discussions, unveiled me the secrets of the MOSYM (nicknamed The Code). Some contributions external to the CMMP are those of Doctors C. Kaneta (Fujitsu Labs. Japan) A. Sokol (Royal Institution, UK), F. Corà (Royal Institution, UK) and A. Edwards (Air Force Research Lab. USA) and those of Professors P. Ugliengo (Universit di Torino, Italy, N. Harrison (Daresbury Labs. UK), J. Weil (University of Saskatchewan, Canada).

All other students at CMMP have been nice travel mates, some of whom heroically risked their lives while speaking to me when my social aptitudes were at an absolute minimum in a Monday morning.

This study has benefited from several computing resources (CMMP computer network, HIPerSPACE at the University College London, and CSAR service at the University of Manchester). System Administrators from all of them have displayed ability in keeping these systems awake and working, and have also been very kind in reneewing my very-often-forgotten passwords.

If I think of how my first supervisor should have been feeling during my studies, the only sensation that comes to my mind is that of a free fall. As a student, I have felt despair while undergoing the trial-and-error process in which true research is based. I guess my supervisor has felt despair while undergoing the you-have-done-the-same-mistake-five-times process in which I sometime based my research. All I can say is that the PhD studies were for me a long way, in which sometimes I found myself very lost, very unconfident, and very tired. A. Shluger has been able to handle the situation and take it to a happy end. This, to my mind, qualifies him not only as an excellent PhD supervisor, but probably as a jet fighter pilot as well.

My original concept for this book was to write a text on "Quantum models of point defects in dielectrics, the philosophy of life and science, why are we here and many other non-trivial questions". Redirecting the effort towards the writing of the technical report that it should be has required a non-negligible amount of effort, that has benefited from very useful comments of J. Gavartin, Adam (Dr. A. Foster), Roby (R. Rodriquez) and, of course, A. Shluger.

When I was ten years old, my teachers used to criticise my minimalist way of constructing too short sentences. Now, they have criticised my heavy Spanish Baroque style when constructing half-page wide sentences. Letting alone the cultural schizophrenia that this has caused in me, the elaboration of the final version of the Thesis required from external help. This book does not look like a treat on quantum cryptography thanks to the invaluable corrections by Adam and Roby. I assume all responsibility of all remaining misspelled words

and incoherent expressions. Che (C. Gannarelli) has given a hand when I needed it the most.

The writing of this book took me a period of time which seems has been quite long since people started asking me "Have you finished your Thesis?" rather that saying "Hi! How is it going?". I apologise for every inconvenience, as I have not been able to avoid making some collateral damage (which is actually quite a trendy thing, nowadays). Professor C. Pisani has been the first boss and victim after my PhD studies, but had the patience to wait until the text was written, and even contributed with very useful comments. I will always be thankful.

Moral support has been a key factor. My mother, father and brothers lead the classification there, since they made the miracle of being close while staying thousand kilometres away, same as the rest of my family. Roby and her words in a square of Sevilla changed the direction of this PhD, and of my life, more than what she still thinks. This was just a small, but fundamental, piece of her support, for which I will always be in debt. Jorge tough me that one can be in exactly the same place looking at the world either from upwards, or from downward. I can't avoid including in many of my regards him together with Mado, Marta the Queen, Laura, Eddie, Quique, Sancho, Rita, Fon and some more. It is an honour (and a pleasant surprise) that Cami and Xanthos count me among their friends. Jaime and Inma were very nice and very wise company. Last but not least, Adam has been an ispiring brother in arms and cold-bloded thermostat with cleaver advices when my Spaniard hot temper was about to melt.

I need to stop at certain point, since I am supposed to expose my research as well. I am conscious that I leave many names unmentioned. I hope they will understand.

To my parents.
For the goals we have achieved.
For the projects we share.

I pass as rapidly as possible
over three years of research,
of interest to specialist alone,
and over the development of a method
akin to controlled delirium,
of interest, probably,
to none but madmen.
And yet this term *delirium*
smacks too much of romanticism;
lets us say,
rather a constant participation,
as intensely aware as possible,
in *that which has been.*

Notes on the writing of Memoirs of Hadrian.
Marguerite Yourcenar.

Aim of this Thesis.

Many problems that are believed will interfere with the progress of miniaturization of microelectronic planar devices have been linked to the loss of reliability of the gate dielectric in MOSFET transistors. Currently, silicon dioxide is almost exclusively employed as a gate dielectric, but most of the problems are actually linked to the use of this material.

The relatively low dielectric constant of silica requires using very thin layers in order to maintain device performance. At such a small thickness, the material loses its insulating properties, since the probability of tunnelling of carriers increases dramatically. Charge trapping caused by point defects may also trigger breakdown processes.

Quantum mechanical modelling allows one to relate the properties of the materials to their microstructure. There are, however, important limitations in the application of these techniques. The cost of the calculation imposes a strong constraint on the size of the models that can be used, therefore making the stage of modelling of the structure crucial.

The overarching aim of my studies has been testing the limits of applicability of several quantum mechanical methods, for predictive modelling and simulation of properties in dielectric materials. This research has been funded by Fujitsu Laboratories Japan, and benefited from close collaboration with Dr. C. Kaneta, and her colleagues.

The three topics selected for the study have been the Si/SiO_2 interface, atomic hydrogen inside α-quartz, and some point defects inside zirconia and hafnia.

The first study intends to find a simple way to monitor the change in the electronic structure across the interface. The aim is to understand how the height of the barrier found by the carriers at the interface changes as the thickness of the oxide layer is reduced (this is a main concern for the developers of MOSFET technology). We also have studied the dependence of the height of the barrier on computational parameters, like the Hamiltonian or the basis sets that determine the accuracy of the calculation.

The second study intends to contribute to the understanding of the behaviour of hydrogen inside silica (also an old problem in microelectronics). Special attention will be paid to understand the effect of some of the approximations that are assumed in the techniques, like the classical treatment of the hydrogen nuclei (within the Born-Oppenheimer approximation) or the approximate treatment of the exchange

and correlation interaction by the DFT techniques.

The third topic is the modelling of defects inside alternative dielectrics that could substitute silica as gate insulator. Before this could happen, the reliability of the materials should be carefully asserted. The trapping ability of the defects inside the materials is one of the factors to study. This requires one to consider further corrections for the error introduced while calculating the position of the unoccupied states. A correction of this type will be presented and discussed.

There is, finally, a part of the Thesis that has focused on the development of computational tools rather than on its application. The aim has been to develop a program for calculation of the electronic structure of solids that will allow very fast calculations, ideally concerning systems with many atoms. The intention was to use such a tool on a first analysis of potentially interesting systems, prior to the employment of more accurate, but also more computing demanding, techniques.

We considered that semi-empirical techniques are interesting candidates for this type of task.The work presented here consists on an updating of a code implementing an INDO technique. This included an optimisation of the algorithms, and the addition of new features, that improved the range of applicability and accuracy. The parameterization required by the INDO method was also developed and tested for systems containing silicon, oxygen and nitrogen.

Structure of the text.

The text is organised as follows. The first chapter provides an introduction to the MOS technology, and to the problems that interfere with its development. The aim is to give an idea of the relevance of the problems studied along this Thesis.

The second chapter provides theoretical background. Special attention is paid here to describe very specific issues of the techniques and codes employed in this research, which may not be of common knowledge.

Chapter 3 describes several issues concerning the data analysis. This chapter focuses of describing how the physical properties of interest here are calculated from the information extracted in the electronic structure calculations.

Chapters 4, 5, 6 and 7 are devoted to present and discuss the research. Each chapter covers one of the topics already listed, and contains an introduction to the problem, a review of the research done so far, a presentation of the result, and an analysis and discussion of the results.

Last chapter contains the conclusions, where the relevance of the results of the study is discussed.

Part I

Motivation and Theoretical Background

Chapter 1

Motivation and technological context

1.1 Semiconductor technology. Reaching the limits.

Micro-integrated circuits (micro-chips) are among the most transcendental human inventions of all times. These devices are able to perform logical and arithmetic operations at much faster rate than any other known device. In addition, its production reports very attractive benefits. This has triggered the development of a powerful industry and required a very active research in order to keep a sustained rate of development; the performance of the devices has doubled in periods of two or three years [1]. As impressive as the rate of evolution and the rate of benefits of microelectronics industry is the fact that it is entirely based in a single, and simple strategy: reducing the size of the circuits.

The reason why miniaturization has such deep impact is that by reducing the size of the circuits a reduction both of the amount of material employed and of the energy consumed by the device when operating is achieved, all this without essentially modifying the nature of the process of manufacturing.

The research teams have found several difficult problems during the forty years of activity, and have many times managed to overcome them. Despite this, the oracles of the industry predict that the rate of development will suffer a drastic slowdown in approximately ten years time, since (yet another) fundamental problem will be found. Are these truly fundamental limits?

[1] This empirical rule was actually established by Moore [1] in the early days of the industrial production of micro-chips. The law, that has been so far fulfilled, describes the exponential growth of the complexity of the circuits due to the decrease of the size of the devices. This influences circuit speed, memory capacity and cost per unit.

1.2 The importance of the Si/SiO_2 system.

The reliance, almost exclusively, on the miniaturization process for improving the technology is due to the nature of the process of manufacturing[2], and this in turn is possible because of the properties of the Si/SiO_2 system.

The basic building block of the logical circuits is the transistor, which can be operated as a controlled switch and therefore process binary information. The most efficient design of transistor so far is the Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET). A schematic representation of the cross-section of a MOSFET is shown in Figure 1.1. In this device, the flow of current (main current) between the source region and the drain region is controlled by means of a bias applied to the gate. The source and drain are built as to have a different density of charge carriers (i.e. holes and/or electrons) than the channel region that separates them. This density of carriers can be, however, modified by applying the gate bias. All this process is possible thanks to the existence of a dielectric layer separating the gate from the channel, and preventing carriers from escaping through the gate. This dielectric is silicon dioxide. The reason is that silica displays high resistivity and dielectric strength, a large band gap, and a very low density of defects at the interface with silicon. However, some other features of the material present many problems that seem difficult to solve, but should be present soon in the microelectronics landscape.

1.3 The problem of the ultimate width. Exploring the interface properties.

Silicon dioxide has a relatively low dielectric constant ($\kappa = 3.9$). This forces one to use thin dielectric layers in the transistor in order to achieve capacitance enough as to produce the required main currents. If the present design of a transistor is kept, then the oxide gate layer is by far the thinnest in the whole structure (as it can be seen in figure 1.1). Predictions are that this thickness should be less than 1.0 nm for the year 2012 [2].

The main problems detected when constructing ultra-thin layers of silica are impurity penetration through the oxide, enhanced scattering of the carriers, reliability degradation, high gate leakage current and the need to grow very controlled uniform layers [3, 2]. In fact, most of the works that try to predict the ultimate width of the dielectric layer center their attention on the tunneling problem [4, 5, 6].

The wide band gap of bulk silica prevents the appearance of a carrier current between the gate and the channel regions. However, basic quantum mechanics predicts

[2]The structure of the micro-integrated circuits is built by overlapping layers of materials with different electrical properties. These properties are accurately controlled by means of doping and oxidation processes. The layers have different patterns that are implanted following a photo-lithography process. A reduction of the size of the circuits can be achieved, in principle, by reducing the size of the patterns implanted. The basis of the lithographic process is the use of masks to protect specific regions of the circuit. Silicon oxide is often used as masking material because it can easily be grown over the silicon substrate as native oxide, and it is chemically and mechanically stable over it.

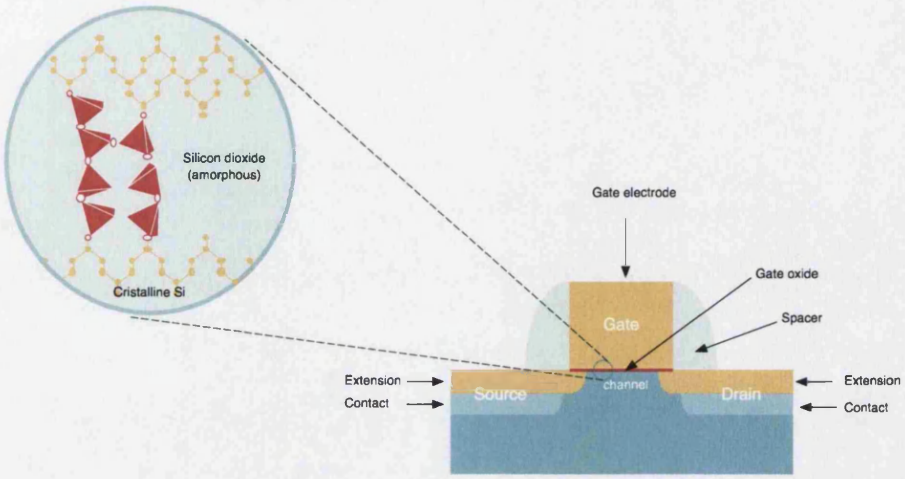


Figure 1.1: **Schematic representation of a MOS system.**
The dielectric layer (in red) is usually made of silicon dioxide, and is the thinnest structure inside this type of device. Predictions are that the thickness of the silica layer will be as small as a few atomic layers. In the magnified region, a schematic representation of the microstructure of the system is shown. The red tetrahedra are the fundamental SiO_4 structural units of the oxide, whose structure lacks of long-range order.

that the tunneling probability increases exponentially with the decrease of the oxide layer width [7]. Furthermore, layers with a thickness of around 1 nm contain around five atomic layers, from which two have been reported by experiments to belong to the interface [5]. The materials in the interface region do not display bulk properties, and therefore predictions based on bulk data are probably inaccurate. This problem is very representative: the devices are becoming so small that the quantum nature of the materials starts to manifest (in this case, as tunneling of carriers) and also structures that at the beginning constituted a small part of the MOS system (like the interface), are now an important part of it, and determine its properties. Both reasons justify that this time the researchers may be right in saying that the problems are truly fundamental.

It is becoming essential to obtain a clear picture of the evolution of the electronic structure at the interface, in order to be able to predict with accuracy the behavior of devices based on ultra-thin oxide layers, and understand which mechanisms are involved, and how parameters like the roughness of the interface, the structure of the oxide or the density of defects affect to it.

The study presented in **chapter 4** gives an overview of the capacity of state of the art first principles techniques to describe the electronic structure at the interface. A fundamental parameter, the valence band offset, has been chosen as a case study, and a comparative performance of different techniques, basis sets and models of the interface is shown.

1.4 Solutions for the tunneling problem. High-κ dielectrics.

The need to scale down the gate dielectric thickness at the same rate as the rest of the device is due to the low dielectric constant of silica. If the present structure of MOSFETs is kept, then the most easy solution for allowing further scaling down is the substitution of the gate dielectric.

One of the fundamental physical limits that the scaling down process is going to reach soon is the tunnelling of electrons between the channel and the gate electrode [7]. Experimental [5] and theoretical [8] studies predict a limit to that thickness (for the case of SiO_2) in a range between 7 and 14 Å. It is also known that the tunnelling current will contribute substantially to the overall leakage current at these scales.

The tunnel effect has explanation only in the framework provided by quantum mechanics. Two main factors determining the size of the effect are the thickness of the layer, and the barrier heights between the SiO_2 and the Si. In fact, both factors are interrelated as a thin layer presents a lower barrier, so it should be expected that the ultimate thickness should be a compromise between the thickness giving both acceptable tunnelling current and barriers [8].

The MOSFET device acts as a capacitor, and therefore the capacitance depends directly in the thickness of the dielectric. Capacitance is given by:

$$C = \frac{\kappa A}{t} \tag{1.1}$$

where κ is the dielectric constant, A is the area, and t the thickness. It is straightforward to conclude from this expression that a material with higher dielectric constant will allow to employ thicker layers without decreasing the capacitance. The increase in the thickness will, on the other hand, avoid the undesired tunnelling. In the microelectronics community the concept of equivalent thickness is used frequently:

$$t_{eqv} = \frac{\kappa_{SiO_2}}{\kappa_x} t_x \tag{1.2}$$

here t_x and κ_x are the thickness of the layer and the dielectric constant of the alternative material. The equivalent thickness t_{eqv} is the thickness of a layer of SiO_2 that will be required to obtain the same capacitance as that given by the layer of alternative material. If the new material has a higher dielectric constant κ than that of silicon dioxide, then a thicker gate layer can be used without losing the required capacitance. The bigger thickness will prevent the appearance of direct tunneling current between the gate and the channel.

The search for alternative dielectrics has been active for many years. The task is not so simple, because the materials have to fulfill many other characteristics apart from displaying a high κ[9]. The new dielectric should be mechanically stable over the silicon substrate, and the growth of too thick interlayers of SiO_2 should be avoided. These requirements constrain the group of materials to those shown in colour over the periodic table in figure 1.2. Additional conditions are the possibility of controlling the quality of the interface with Si, and the resistance to diffusion of oxygen, dopants

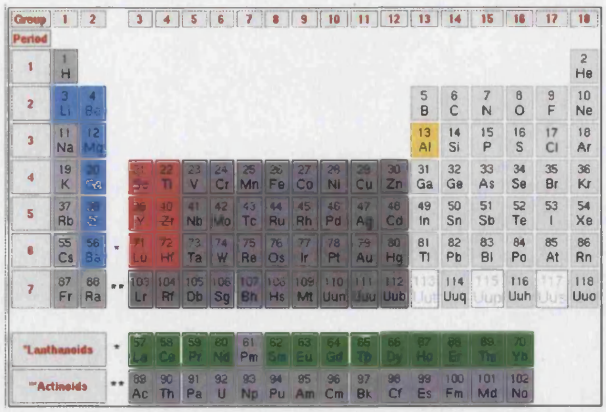


Figure 1.2: **High-κ dielectrics.**
The oxides of the elements that appear coloured are being considered as possible substitutes for silica as gate dielectrics. These oxides are the only ones that are stable over silicon substrate, and do not get reduced to form SiO_2 interlayer.

and impurities, and the reliability of the dielectric. The degradation of the dielecric it is directly related to the accumulation of charge inside the material [3]. It is therefore important to understand the mechanisms by which the materials become charged.

Historically, the search for a new dielectric started by considering materials which were familiar for the MOS community. Silicon nitride was the first material in the list, because it can be easily grown in a process similar to silicon oxide. Silicon nitride has a dielectric constant κ=7.0, while silicon oxide has κ=3.9. Silicon nitride presents, however serious disadvantages: lower barriers for the carriers, and a density of defects several orders of magnitude higher than the oxide [10]. Surprisingly, this is not the case of the oxynitride. Incorporation of nitrogen into the oxide network by annealing in NH_3 reduced the defect generation rate, while allowing one to increase the thickness of the layer. The effect of controlling defect formation is believed to be due to the ability of nitrogen of controlling hydrogen diffusion[3] [9]. Nitrogen has also been found to be efficient in controlling the diffusion of Boron ions from the doped Silicon layers [9], but for this effect to happen efficiently a high concentration of Nitrogen is required. A Si-N-O compound with a high concentration of Nitrogen is difficult to achieve without causing a decrease in device performance, because threshold voltage shifts appear. This is due to a positive charging of the layer as Nitrogen incorporates to the SiO_2 network, and to a non-homogeneous distribution of Nitrogen while this incorporation takes place.

The experience with Nitrogen pointed out the need to impose more selection rules on the search for the dielectric. Certainly the density of charge traps is another important parameter, since they are precursors for the dielectric breakdown process. The next materials tested, Ta_2O_5, TiO_2 and $(Ba_xSr_{x-1})TiO_3$ (short-named BST), were also familiar to the MOS community from their use in some related devices, like

[3]The relationship between hydrogen and the formation of defects in silicon oxide is explored in more depth in chapter 6.

bulk-capacitors, and DRAM memories. All of them presented problems of stability when grown over the Silicon substrate, outlining yet another important parameter to consider.

Despite its low dielectric constant, silicon dioxide displays almost unique structural characteristics [4], which are not exhibited by any other candidate, so far. Potential candidates like Ta_2O_5, TiO_2 and BST show tendency to react with Si at the interface. Despite these problems, heavy work has been done in developing Ta_2O_5-based technology and it seems that at least two generations of memory devices can be based on it [11]. Predictions from the industry [2] suggest that Ta_2O_5 could also be used in the development of a generation of transistors, and the next can make use of BST materials. Calculations of the barriers for carriers at the interface of each one of these materials with Silicon done by Robertson [12] suggest that the height of the barriers is too small, making them poor gate oxides. This is especially true if the density of dopants on the Silicon layers is increased.

Also with the status of known temporary solution is Al_2O_3, a well-characterized and familiar oxide (Aluminium has been widely used as gate electrode material), which presents wide gap (8.8 eV [12]) and high barriers, but not very big dielectric constant. Such a material could also be used in next generation, providing a thickness big enough as to avoid tunnelling, and with a capacitance similar to a layer of SiO_2 around 10-15 Å wide. Further scaling would present similar problems to silicon dioxide.

One of the characteristics that make silicon dioxide so attractive is its flexible amorphous structure. Amorphous structures show no grain boundaries or facets, which favour leakage current and cause fluctuations of the electrical properties. Often they also present very low density of defects at the interfaces, because of their capacity to adapt their structure, as to saturate most of the dangling bonds that are present at the junction[5].

In **chapter 5** I study the electrical properties of some point defects inside two materials that are being considered as possible alternative dielectrics: Hafnia and Zirconia. The ability of defects to act like point charges and substitutional impurities is discussed in terms of their ionization potentials and electron affinities. Some work presented in **chapter 7** concerns the electronic structure calculations (at semi-empirical level) of silicon nitride and silicon oxynitride, which have been considered for long time as possible alternative dielectrics[3]. Nitrogen is indeed used in current MOS systems to improve reliability of the gate dielectric, and stop diffusion of boron [9], and therefore there is interest in obtaining a clear understanding, at the microscopic level, of the role played by this species when introduced inside silicon dioxide.

[4] A more detailed description of the properties of SiO_2 can be found in section 4.2.1 .

[5] The dangling bonds appear due to the change in the density of species with a given coordination on each material forming the junction. In the particular case of the Si/SiO_2 interface, the density of four-fold coordinated Si atoms is approximately three times bigger in the crystalline Si than in the oxide.

1.5 Dielectric reliability. The role of hydrogen.

The electrical stress caused by operation of the devices degrades the microstructure of the oxide, modifying its properties. Breakdown processes are triggered if the charge accumulates excessively, causing the device to fail [3].

The mechanisms of breakdown are complex and still further research is required for full understanding. In the initial step of charge trapping, one possible mechanism involves the presence of Hydrogen and its diffusion through the oxide.

Hydrogen is always present in the MOS systems, due to its use as a gas (H_2) in annealing processes during the manufacturing of the device. Hydrogen has been reported to passivate defects at the interfaces and therefore reduce structural strain. This interfacial hydrogen can be freed by electrons tunneling through the oxide, and then diffuse through the dielectric, generating electron traps.

Many different computational methods have been employed in the study of these processes. The success of theoretical modelling on justifying Hydrogen's behavior inside silica very much depends on the charge state of the ion. Both H^+ and H^- interact strongly with the host, creating chemical bonds that limit the mobility of the ions, and allow simulating the system even using cluster models [13, 14, 15, 16]. Neutral Hydrogen (H^0) constitutes a much more difficult target, given its weak interaction with the host. Hydrogen is also the lightest element, making it necessary to account for its quantum nature.

Current DFT implementations make use of approximate functionals that introduce a certain error, especially when describing the exchange and correlation interactions. Hydrogen constitutes once more the extreme case, given that it only contains a single electron. Despite this limitation, there are already several works that have used DFT techniques for studying neutral hydrogen inside silica [17, 18, 19]. In my opinion, all these works are assuming that neutral Hydrogen can be described using DFT with the same accuracy as any other element of the periodic table, without considering the points just outlined above. The study presented in **chapter 6** tries to fill this gap, analyzing in detail the results and also trying to contrast them with experimental data.

The paramagnetic character of neutral hydrogen makes it ideal for Electron Paramagnetic Resonance (EPR) experiments, which provide useful structural information. Fortunately, several EPR measurements have been done for neutral Hydrogen in α-quartz [20]. The choice of a crystalline host has many advantages for the characterization of the system, since many of the sites where Hydrogen can stay inside the crystal become equivalent by symmetry. In the amorphous systems, on the other hand, the disorder of the structure makes any possible site virtually different from all the others, and therefore the study has to include statistical sampling. From the point of view of computational methods, translational symmetry can also be employed to reduce the cost of the calculation to within affordable limits (see next section). Both because of economy of the calculations and possibility to compare with the experimental data, we have also considered α-quartz as the host for hydrogen.

1.6 Computational methods as a tool for research.

In the struggle for achieving faster and cheaper devices, computational methods are gaining weight as research tools, for the simulation of new materials or the interpretation of complex experimental data concerning the action of defects or the relationship of microstructural features with electrical properties.

The three chapters referred to up to now in this introduction (4, 5 and 6) make use of first principles Hartree-Fock and/or Density Functional theory techniques, which are among the most popular for electronic structure calculations. They can be successfully applied to the study of point defects (see chapter 5 on point defects in hafnia and zirconia), although there are some intrinsic limitations due to the approximations implemented in the methods (see chapter 6 on diffusion of hydrogen in quartz), and therefore results should always be analyzed critically.

Another important limitation of the first principles techniques is the computational cost associated with them. On its application to solid state physics and chemistry, translational symmetry has been exploited for the simulation of crystalline systems. It is well known, however, that gate dielectrics display an amorphous structure, and therefore modelling these systems using periodic models is an approximation. There has been a long discussion about the effect of long range disorder in the properties of materials (see, for instance [21, 22, 23, 24]), and the discussion still remains open.

The use of semi-empirical techniques allows a remarkable reduction of the computational cost, although the accuracy and transferability is lower than in the first-principle methods. This makes semi-empirical methods ideal for carrying out extensive exploratory calculations. The remaining chapter of results (**chapter 7**) follows the latter approach. The INDO semi-empirical technique is basically an implementation of the Hartree-Fock scheme, where further approximations have been made for achieving computational efficiency. These approximations are based on neglecting selectively some integrals representing electron-electron interaction. These integrals are replaced by parameterized values. The price paid is that the results become parameter-dependent and are also affected by the new approximations, but in exchange systems with many more atoms can be simulated. The work presented in chapter 7 concerns the set up of a code for performing INDO calculations, and also the fitting of a set of parameters for studying some dielectric materials: silicon dioxide, silicon nitride and silicon oxynitrides.

These targets were fixed as part of a collaboration project with Fujitsu Japan, who showed interest in developing a tool that will allow a quick simulation of new materials, previous to more detailed analysis. The character of this chapter focuses in the development and improvement of the INDO scheme in the program. Several calculations of systems of interest are presented to prove the validity of the technique, and a possible application for determining the presence of localized states in amorphous networks is shown at the end.

Chapter 2

Methods of calculation

2.1 First-Principles techniques and codes.

All the studies concerning gate oxides that will be presented in this Thesis are based on results from first principles electronic structure calculations performed using either Hartree-Fock (HF) or Density Functional Theory (DFT). Both theories are widely employed in both computational chemistry and solid state physics, they have been extensively described elsewhere, and their capabilities and limitations have been explored for tens of years.

Since solids can be seem as "big molecules", the distinction between chemistry and solid state may seem arbitrary, but it is actually necessary and defines two big areas of research, with remarkably different approaches. The use of periodic models is more popular in the solid state disciplines, because this allows taking into account the long-range interactions and to avoid the surface effects of the cluster models. This approach has been widely employed here, and its advantages and disadvantages will be discussed in 2.2.

There are several texts devoted to the introduction of electronic structure calculation techniques, containing brief introductions to each one of the methods, and often present comparative studies. I have found particularly interesting and useful the texts by Springborg [25] and Ohno *et al.* [26]. It is also worth mentioning the book by Pisani, that focuses attention on the difficulties found while studying the electronic structure of extended systems [27]. An interesting comparative study of the accuracy of the different techniques in the specific case of (semiconductor) solid systems can be found in the work by Muscat *et al.* [28].

In what concerns the Hartree-Fock method, a very exhaustive derivation can be found, for instance, in the book by Szabo and Ostlund [29]. The implementation of the HF theory for the case of periodic systems turns out to be particularly complicated, mainly because of the non-local character of the exchange operator. This is also the reason why there are very few codes based on the HF technique for studying periodic systems. A detailed discussion on the evaluation of the exchange interaction at HF level will be given in 2.3.

All the HF calculations in this work have been carried out using the CRYS-

TAL 98 package, developed at Torino University (Italy) and Daresbury Laboratory (UK) [30]. This program is also unique in allowing switching between many different Hamiltonians, therefore opening the possibility to evaluate the accuracy of different techniques. An example of this type of comparative study is given in chapter 4. The CRYSTAL 98 package employs Gaussian basis sets for generating Bloch functions, that can be used as basis sets of systems with translational symmetry. There are several differences between the Gaussian basis sets used in molecular systems and those used in periodic systems. This topic will be addressed in detail in 2.4.1. Unfortunately, the CRYSTAL 98 package does not allow one to perform efficient geometry optimization of the atomic structures. This has constituted the biggest drawback when using this software.

The Density Functional Theory has been one of the main tools for the studying of extended systems, due to very efficient implementations based on plane wave basis sets. A relevant text describing this theory is that by Parr and Yang [31]. The denomination DFT is, however, too general, since there are many different schemes, that differ essentially in the way exchange and correlation interactions are calculated. Probably the most relevant techniques used in solid state have been the Local Density Approximation (LDA), and the Generalized Gradient Approximation (GGA). A brief description and further references can be found, for instance, in [25] and [27].The LDA has been the first and simplest approach. Both exchange and correlation of a given electron distribution are assumed to have the same value as that of a homogeneous gas of electrons with the same density. The GGA approach is able to account for the inhomogeneity of the electron distribution, through the information contained in the gradients of the density. The concept underlaying the GGA approach is to expand the expressions that describe the exchange and correlation as a truncated series, including terms containing the gradients of the density. Despite the gain in accuracy with respect to the LDA in many applications, it is worth noting that further improvements based in expansions that consider terms of higher order have not lead to any significant improvement.

I have employed the VASP 4.4 code [32], for performing first principles electronic structure calculations at DFT-GGA level, using plane wave basis sets, whose main characteristics are discussed in 2.4.2. This program becomes particularly efficient due to the use of ultra-soft pseudo-potentials to simulate the presence of core electrons. A detailed description of this is given in 2.4.3. The VASP package allows performing very efficient geometry relaxation.

The HF and DFT schemes are somehow complementary in what concerns some of their limitations. The HF technique does not account for electron correlation effects, but provides a way for evaluating exactly the exchange interaction. The DFT schemes, on the other hand, evaluate both exchange and correlation interactions, but in an approximate way. Recently, several hybrid methods have been presented, that aim to exploit the advantages of both HF and DFT schemes. One of the hybrid methods that has provided best results is the B3LYP scheme [33], which basically

allows one to evaluate the exchange and correlation interactions through a weighed sum of the exact HF exchange and the DFT functionals. This scheme has been employed here (as implemented in the CRYSTAL 98 code) for part of the studies, mainly due to its accuracy on the prediction of band gaps of semiconductors and oxides [28].

2.2 Models used to calculate properties of solids.

Solids constitute a challenge for computational science, since they require one to describe the contribution due to the long-range interaction. *Brute force* modelling, where a huge number of atoms is considered, is currently out of the range of state of the art methods [1]. Very briefly, the different methodologies that have been applied, so far, for studying properties of solids are:

- **Molecular cluster model.** A specific region of the solid (like, for instance, a point defect) is modelled as a molecule. The broken bonds that connect the cluster with the rest of the solid are saturated with hydrogen atoms. Long-range interaction is obviously not reproduced, but the mechanical constraint caused by the presence of the lattice can be partially reproduced if the saturating hydrogen atoms are kept in fixed positions. Examples of this approach can be found in the work by Pacchioni and co-workers [34].

- **Embedded cluster model.** A region of the solid is modelled as a molecular cluster, and the rest of the solid is modelled at a lower level of theory (classic model). The broken bonds are saturated using a more or less sophisticated scheme. Interesting examples of embedding technique can be found in the works by Shluger and Sushko[35].

- **Periodic model.** Following the concepts developed for the description of crystalline systems, a solid is constructed by periodic translation of a cell. The model, therefore, displays translational symmetry, but the unit cell does not need to be necessarily a basic crystallographic cell. This opens the possibility to model complex structures, like defects or amorphous systems, using the periodic approach, although several factors need to be considered (see below). Several examples of periodic models will be shown here.

- **Perturbed cluster model.** This technique is based in the Green Function Theory. The electronic structure of a "perfect" solid is calculated first (in practice, using a periodic model). Then a region of the solid is selected and its structure modified (creation of the defect). The electronic structure of this region is re-calculated, considering the interaction due to the rest of the solid as a perturbation. Exchange of charge density between the host and the cluster

[1] The only exception to this statement will be the so called order-N DFT schemes. Practical implementations of this schemes were just starting to be available while I was performing the work presented in this Thesis.

is allowed. An example of this type of approach can be seen in the work by Pisani and co-workers [36].

The use of periodic models allows one to obtain a full quantum mechanical description of the system, which is homogeneously consistent , i.e. the electron density is described with equal accuracy in the whole volume of the cell. Cluster models, on the contrary, are always affected by the propagation of the perturbation caused by surface bonds, and reliable descriptions can be obtained only in the most internal region of the cluster. Perturbed cluster models, which are the only other method that offers a consistent quantum mechanical description of an isolated defect inside a solid, have in practice very drastic limitations, that constrain the quality of the basis sets that can be used. In this work, periodic models have been used almost exclusively. There are several difficulties and problems with these models, which will now be discussed in detail in the rest of this section.

2.2.1 Modelling defects in periodic systems. Interaction between images.

When a periodic model of a point defect is built, there is a contribution to the total energy of the system due to the interaction between the defect and its images in neighboring cells. This defect-defect interaction, mainly of electrostatic nature, has full physical sense. However, it is often not possible to reproduce a system with a realistic density of defects (i.e. number of defects per unit of volume), since the cost of the calculations constrains us to employ too small cells. Since the electrostatic interaction has long-range character, there is a slow convergence of the total energy with the size of the cell. It is then necessary to substract the contribution due to the defect-defect interaction, in order to obtain properties of the system that will not depend on the size of the simulation cell. This is equivalent to obtain a description of an isolated defect.

Care should then be paid to understand how strong is the interaction between a defect and its images, and how this interaction affects to the total energy and the wavefunction. Several studies have focussed their attention to this problem (see, for instance, [37, 38] and references therein).

In order to estimate the contribution of the electrostatic energy, we consider the system to be an assembly of point charges (the set of nuclei), together with a continuous charge distribution (electron density). Since it is periodic, the system should display translational symmetry, an therefore it is possible to find a representative unit cell (in principle of arbitrary shape). The charge density of the system is then

$$\rho(\mathbf{r}) = \sum z_i \delta(\mathbf{r} - \mathbf{r}_i) + n(\mathbf{r}) \tag{2.1}$$

and verifies $\rho(\mathbf{r} + \mathbf{L}) = \rho(\mathbf{r})$ where $\mathbf{L}$ is a lattice vector, and $\mathbf{r}$ is a vector in the three-dimensional space. The electrostatic potential is then given by

$$\phi(\mathbf{r}) = \sum_{\mathbf{L}} \int_{cell} d^3\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{2.2}$$

where the sum over the set of lattice vectors is equivalent to an integration over the whole volume occupied by the system. If the system is infinite, so is the sum (this is equivalalent to introduce the periodic boundary conditions). In the asymptotic limit $\mathbf{L} \rightarrow \infty$, the terms of the sum in the right hand side of equation 2.2 have the form $q_n P_n \cos(\theta) |\mathbf{L}|^{-(n+1)}$ where n is the order of the lowest non-zero multipole q_n of the charge distribution, and P_n are the Legendre functions. It is then possible to proof [37] that the exact summation of the potential in an infinite lattice is:

$$\phi(\mathbf{r}) = \sum_{\mathbf{L}} \int_{cell} d^3\mathbf{r}' \rho(\mathbf{r}') \left[\psi(\mathbf{r}, \mathbf{r}') - \frac{2\pi}{3V} r'^2 \right] \tag{2.3}$$

$$\psi(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{L}} erfc(\eta |\mathbf{r} - \mathbf{r}' + \mathbf{L}|) + \frac{4\pi}{V} \sum_{\mathbf{G} \neq 0} \frac{e^{-\frac{\mathbf{G}^2}{4\eta^2}} e^{i\mathbf{G}(\mathbf{r}-\mathbf{r}')}}{\mathbf{G}^2} \tag{2.4}$$

where V is the volume of the simulation cell, and is a $\mathbf{G}$ reciprocal lattice vector.

The result just shown assumes that the unit cell does not contain any significant non-vanishing multipole. If, for instance, it contains a non-vanishing dipole moment, an additional term appears in the potential. Such term corresponds to a constant electric field, whose magnitude and shape depend in the shape of the boundary conditions and the way the unit cell is defined. This causes ambiguity in the results, and has been subject of extensive discussion.

In the approach followed here, we have subdivided the density inside a unit cell containing a defect in two terms:

$$\rho(\mathbf{r}) = \rho_{perfect}(\mathbf{r}) + \rho_{defcut}(\mathbf{r}) \tag{2.5}$$

where the first density $\rho_{perfect}(\mathbf{r})$ is that of an unpertured solid (i.e. the solid without the defect) and the second one $\rho_{defcut}(\mathbf{r})$ is that of the defect itself. We can asume, without loss of generality, that the density of the perfect solid displays translational symmetry ($\rho_{perfect}(\mathbf{r})$=$\rho_{perfect}(\mathbf{r} + \mathbf{L})$), and does not have any non-vanishing multipole. Such splitting of the density can be achieved in practice by performing two different calculations employing the same simulation cell, but with and without the defect inside. $\rho_{perfect}(\mathbf{r})$ corresponds to the density of the supercell without defect, while $\rho_{defcut}(\mathbf{r})$ can be found by substracting the density of the cell containing the defect from that of the "perfect" cell (the density of the defect is "cutted out" of the solid).

There are three different contributions to the electrostatic energy. The first one is the interaction of the unperturbed structure with itself, and the second the interaction between the density of the defect with that of the unperturbed structure. Both contributions are independent of the size of the unit cell.

The third contribution is due to the interaction between the density of the defect

and its images, which depends on the size of the cell as discussed. It is possible to estimate this contribution by an appropriate choice of the boundary conditions. In the work by Makov and Payne, it was shown that, for cubic unit cells, the total energy becomes

$$E_{Tot} = E_{def} + \frac{2\pi \mathbf{P}^2}{3V} \tag{2.6}$$

where $\mathbf{P}$ is the dipole moment of the defect, and E_{def} is the electrostatic energy without any correction, of the cell containing the defect. For a detailed derivation of the expressions, where the dipole moment is expressed in terms of the point charges and the uniform density as decomposed in Equation 2.1, we refer to the original paper and appendix by Makov and Payne. A generalization of the equations for any Bravais lattice has been presented recently by Kantorovich [38].

Only dipole-dipole interaction has been considered here. For the sizes of cells considered in our studies, the contribution due to higher order multipoles (quadrupole-quadrupole) showed to be independent of the size of the cell.

In this expression, the problem of the ill-definition of the dipole moment remains latent: different choices of supercell will give different dipoles, and therefore different values of the energy. However, if the supercell chosen contains the defect density in the same configuration as in the bulk limit (that is, for an infinitely large unit cell), then the dipole becomes invariant to the choice of supercell. In practice, we have chosen supercells where the density of the defect was entirely contained within the box, i.e. the density was not splitted by any of the surfaces limiting the cell.

The expressions just presented account for the changes in the charge distribution induced by interaction of the defect with its images. There is, however, another source of cell-size-dependence of the energy, which is the dielectric response of the medium to the presence of the defect. In this case there is a non-electrostatic contribution to the energy, which prevents us to be able to estimate the contribution by means of the methods just presented. Instead, a phenomenological approach is often employed (and we have used it here), consisting on scalling the potential by the dielectric constant of the material, κ.

$$E_{Tot} = E_{def} - \frac{1}{\kappa} E_{dip} \tag{2.7}$$

In principle, the dielectric constant that should be employed in the calculation should be that deduced from first-principles calculations. When this value is not available, we have made use of the experimental one.

2.2.2 Charged systems in periodic models.

The total energy of a periodic system where the cell is not neutral is divergent. The reason is to be found in the contribution of the long-range electrostatic interaction between the different cells forming the system. Similarly to the case described in previous section, it is possible to substract such contribution from the total energy

of the system.

Studying a charged defect using a periodic model involves dealing with unit cells whose charge state is not neutral. Then it is convenient to consider the subdivision of the density in the density of the unperturbed lattice, plus that associated to the defect, as showed in Equation 2.5. In this case, we assume that the unperturbed density corresponds to that of a neutral system, and that the unbalanced charge is due to the defect only.

In order to cancel the effect of the charge q of the defect, one can add a charged background of equal magnitude but opposite sign $(-q)$, uniformly distributed over the whole volume V of the cell. The density of the defect plus the charged background is then

$$\rho'(\mathbf{r}) = \frac{-q}{V} + \rho_{defcut}(\mathbf{r}) \tag{2.8}$$

here the prime indicates that this density does not correspond to the total density of the system, since the density of the unperturbed cell (i.e. the cell without the defect) is not being considered. If a point charge q placed at a point of the cell $\mathbf{r}_0$ is added and substracted, then the density is subdivided in two different contributions

$$\rho'(\mathbf{r}) = \rho_1 + \rho_2 = \left[q\delta(\mathbf{r}-\mathbf{r}_0) + \frac{-q}{V}\right] + [\rho_{defcut}(\mathbf{r}) - q\delta(\mathbf{r}-\mathbf{r}_0)] \tag{2.9}$$

by convenience, $\mathbf{r}_0$ is chosen so that ρ_2 has no dipole, and we assume that the origin of coordinates corresponds with the center of the cell. Note that ρ_1 actually is size dependent, since the density of the neutralizing background is inversely proportional to the volume of the cell.

There are three contributions to the energy electrostatic energy, that we will denote E_{ij} with $i,j = 1,2$ refering to the interaction between the densities ρ_i and ρ_j. The interaction of ρ_1 with itself is equal to the Madelung energy of a lattice of point charges plus a the neutralizing background, and therefore $E_{11} = -\frac{q^2\alpha}{2L}$, where L is the linear dimension of the cell, and is the (lattice dependent) Madelung constant. E_{22} corresponds to the interaction between a neutral charge density in a lattice (since ρ_2 has no dipole), which displays fast convergence with the cell size. Finally, E_{12} is the interaction between the two densities ρ_1 and ρ_2. It is possible to difference two terms E_{12}^a and E_{12}^bin this contribution, which correspond to the interaction between ρ_2 and the point charge in ρ_1 and the interaction between ρ_2 and the uniformly charged background, respectively. In the case of E_{12}^a , the dominant interaction is that between the point charge and the quadrupole moment of ρ_2 (if the lattice is cubic, then this interaction vanishes by symmetry). On the other hand E_{12}^b is size dependent. The final expression for E_{12} is [37]:

$$E_{12} = \frac{2\pi q}{3V}\int_{cell} d^3\mathbf{r}\rho_2(\mathbf{r})\mathbf{r}^2 + O(L^{-5}) \tag{2.10}$$

Summing up all the contributions mentioned, the electrostatic energy turns to be

$$E = E_{def} - \frac{q^2\alpha}{2L} - \frac{2\pi q}{3V}\int_{cell} d^3\mathbf{r}\rho_2(\mathbf{r})\mathbf{r}^2 + O(L^{-5}) \tag{2.11}$$

Up to know, we have excluded considering the environment in which the defect is introduced. Similarly to the case presented in the previous section, it is necessary to account for the response of the medium to the presence of the defect. And once more, this can be done in first instance by scaling the electrostatic contribution by the dielectric constant of the material, ε.

$$E = E_{def} - \frac{q^2\alpha}{2L\kappa} - \frac{2\pi q}{3V\kappa}\int_{cell} d^3\mathbf{r}\rho_2(\mathbf{r})\mathbf{r}^2 + O(L^{-5}) \tag{2.12}$$

There has been some concern on the validity of the approach just taken. The arguments against this procedure are that the validity of the approximation is very dependent in the assumption that the density related to the point defect remains very localized. This may not be the case for certain type of defects, especially if they are negativelly charged and the extra electrons move to delocalized states in or near the conduction band.

2.2.3 Reference energy levels in infinite systems. The band alignment problem.

The use of periodic systems causes a lack of reference level in the energy scale. The infinite extension prevents simulation of the process of extraction or addition of any particle (electron or nuclei), since there is no place to take it from. Therefore, it is not possible to define a vacuum level that would be a universal reference and would allow one to establish an absolute energy scale. The problem of defining a reference level is far from being trivial.

The need of a reference level will become evident when studying how the band structures of two different materials align (as in chapter 4). In order to understand the dependence of the energy barriers on the width of the oxide layer it is necessary to observe in detail the evolution of the band structure of the system in a direction perpendicular to the plane of the junction. The electronic levels experience a shift that constitutes an energy barrier for the charge carriers. Although the junctions can be structurally abrupt, the evolution of the electronic structure extends over a few layers at each side of the interface (see figure 2.1). This fact has been proven both by theory [39] and experiments [5]. From the point of view of device engineering, however, the evolution of the electronic structure has been idealized by a step function (also shown in figure 2.1), neglecting the actual spatial extension of the change. A fundamental parameter in this model is the offset, defined as the change in energy of the edge of the band across the junction. Such idealization may become inadequate for the design of new devices, since the junctions implemented in working devices are soon going to be so small that the thickness of the layers may not allow the materials to recover their bulk electronic structure (let alone the action of other factors as defects commonly present at the interface or roughness of the junction).

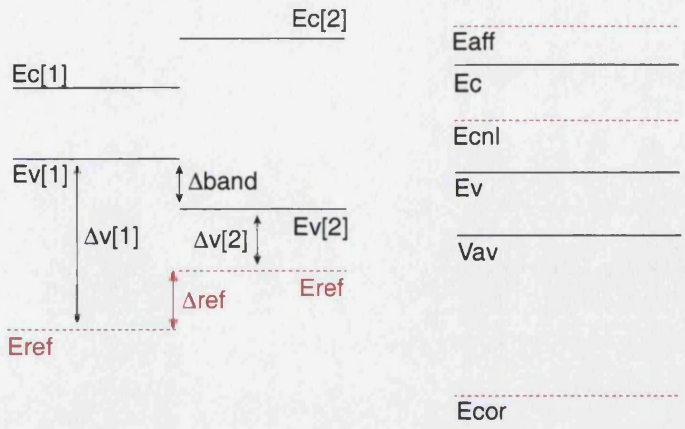


Figure 2.1: **The band alignment problem.**
The left side plot shows the most common approach taken to find how two band gaps align at an interface. The band offset Δband measures how energy of the edge of a given band (on the graph, the top of the valence band) changes at each side of the junction. Rather than determining it directly, it is possible to define a reference level, Eref, and then determine the change in the position of this level across the interface, Δref. Then the distance between the edge of the band and the reference level, at each side of the junction is found (Δv[1] and Δv[2]). This information suffices for determining the offset Δband=Δv[1] -Δv[2]- Δref. The graph on the left shows some examples of reference levels: the electron affinity, Eaff, the charge neutrality level, Ecnl, the averaged electrostatic potential, Vav, and the inner core levels, Ecor. See text for further comments on these choices.

Band alignment theories try to find the relationship between the essential features of the electronic structure at the junction, and those that characterize its microstructure [2]. The list of different techniques on band alignment, both experimental and theoretical is large, and the discussion concerning the different features of each one would be very large. I will rather list here the most important concepts and the theories that have been most relevant. An excellent and exhaustive review on the band alignment problem can be found in the text by Franciosi and Van de Walle [40].

The problem of band alignment is usually formulated in terms of the edges of the band, as illustrated in Figure 2.1: given the junction between two materials, we want to know what is the change in energy Δband experienced by an electron in a state on the edge of a given band, as its moves across the interface. The most common strategy for solving this problem follows two steps: a) determining the change in energy Δref of a reference level Eref, and b) determining the position of the edge of the bands with respect to the reference level at each side of the junction (in the Figure the quantities Ev[1] and Ev[2] refer to the valence band). This information is enough to find Δband.

Also in Figure 2.1, common choices for the reference level are shown. The electron affinity Eaff appears in the first band alignment theory, by Anderson [41]. Although

[2]Ideally, it will be possible to determine the height of the energy barriers for carriers through a careful selection of the composition of the materials and a control over the process of formation of the interface.

conceptually it is absolutely correct, the affinity in fact depends critically on properties of the surface, like orientation and relaxation. There have been several attempts to find surface-independent levels; Frensley and Kroemer tried to identify the vacuum level for each semiconductor [42], while Harrison used as a reference the electronic states of free atoms [43]. Since the mentioned theories rely on bulk-independent magnitudes, they are not able to account for the effect of the particular features of the interface.

A conceptually different approach has been followed by Tersoff, by defining a neutrality level [44]. Tersoff's study actually presents a very intuitive and interesting mechanism that determines the charge alignment, which is worth describing here in brief. When the interface forms, the electron density at the interface redistributes, filling some gap states that are intrinsic to the system (not due to the presence of any point defects). This redistribution generates dipoles, that tend to drive the system towards the configuration in which the dipoles become zero. Tersoff's main assumption is that the system will tend to be very close to this canonical situation, and finds the last gap level that should be filled in such state (charge neutrality level). Such level should be actually aligned in both materials at the interface, so both materials are in the zero dipole situation, and this allows determination of the relative position of their band structures. The gap states are mixtures of valence and conduction band states, and their spectral weight can be deduced from the Density of States near the gap region.

The effects of the charge redistribution at the interface can be considered through the information provided by electron structure calculations. In order to refer, once more, to an intrinsic property, Van de Walle and Martin [45] calculate the average electrostatic potential of the materials. This technique can only be applied to materials which display periodicity, so the electrostatic potential can be averaged over a distance equivalent to the lattice parameter, therefore displaying a constant value over all the material. A modification of this method, by Demkov and Sankey [46], uses as reference level the average spatial dependence of the expectation value of the Hamiltonian calculated for the valence states of a reference atom type. Obviously, this method can only be applied for studying interfaces of materials which have common atomic species in their chemical composition.

For the sake of completeness, common experimental techniques are listed here. Transport methods (either I-V or C-V measurements) were historically the first to be employed on the study of band alignment. These methods, however, give an average over the direction perpendicular to the interface, while the offset is spatially very localized at the junction. The presence of dopants and the performance of the electrical contacts have been also reported to affect the measurement. Optical methods include absorption, luminescence, light scattering and photo-emission spectroscopy. The latter technique has proven to give remarkably good results and has been widely applied to the study of the Si/SiO_2 system. However, there are extra effects that should be taken into account, like the fact that the system becomes charged and is

left in an excited state. A deeper discussion concerning this issue will be provided while discussing the results in chapter 4.

The method that has been used here differs from all the discussed ones, since it does not follow the strategy of identifying a reference level; instead, the band edges are compared directly. The possible advantages and disadvantages of this scheme will be discussed in section 3.1

2.3 Evaluation of the Coulomb and exchange interaction in extended systems.

Both the Coulomb and exchange interactions have long-range character, and therefore its evaluation requires considering all the nuclei of the system, and the whole of the electron density. In the case of infinite solids, it is obviously impossible to design any practical implementation without assuming any further approximations.

2.3.1 Treatment of the Coulomb series. Ewald summation.

The terms that represent the coulomb interaction constitute a series with conditional convergent character. The technique proposed by Ewald [47, 48] is routinely employed to sum this series ensuring its convergence. The idea behind this technique is to rearrange the terms, adding and substracting a screening charge distribution, to give two convergent series, whose sum gives the correct value of the potential. The charge density can be written as a sum over point charges plus a screening background:

$$\rho_{ewald}(\mathbf{r}) = \sum_{L} \delta(\mathbf{r} - \mathbf{r}_i - \mathbf{L}) - \rho_{back} \tag{2.13}$$

where $\mathbf{r}_i$ are the position of the point charges and $\mathbf{L}$ runs over the complete set of lattice vectors. This density is then divided in two contributions:

$$\rho_1(\mathbf{r}) = \frac{1}{\mu\sqrt{\pi}} \sum_{L} exp\left(-\frac{(\mathbf{r} - \mathbf{r}_i - \mathbf{L})^2}{\mu}\right) - \rho_{back} \tag{2.14}$$

$$\rho_2(\mathbf{r}) = \sum_{\mathbf{L}} \left[\delta(\mathbf{r} - \mathbf{r}_i - \mathbf{L}) - exp\left(-\frac{(\mathbf{r} - \mathbf{r}_i - \mathbf{L})^2}{\mu}\right)\right] \tag{2.15}$$

These two series employ arrays of gaussians. They cancel the background potential in one of the series, and are substracted from the delta function charge distribution in the other. The width of the gaussians μ can be tuned to optimise the rate of convergence.

These densities are employed to compute the Ewald potential (the potential due to $\rho_1(\mathbf{r})$ is actually more easily computed in the reciprocal space). The final shape of this potential is

$$\begin{aligned}\phi_{ewald} = & \sum_{\mathbf{G}\neq 0} \frac{4\pi}{V\mathbf{G}^2} exp\left(-\frac{\mu^2\mathbf{G}^2}{4} + i(\mathbf{r}-\mathbf{r}_i)\mathbf{G}\right) \\ & + \sum_{\mathbf{L}} \frac{erfc\left(\frac{|\mathbf{r}-\mathbf{r}_i-\mathbf{L}|}{\mu}\right)}{|\mathbf{r}-\mathbf{r}_i-\mathbf{L}|} - \frac{\pi\mu^2}{V}\end{aligned} \tag{2.16}$$

where $\mathbf{G}$ runs over the reciprocal lattice vectors, V is the volume of the cell, and

$$erfc(x) = 1 - erfc(x) = 1 - \frac{2}{\sqrt{\pi}}\int_0^x exp(-y^2)dy \tag{2.17}$$

The electrostatic energies are calculated multiplying the Ewald potential by the appropriate product of charges (q_iq_j). In principle, Ewald technique can be employed for the calculation of Coulomb interaction in both DFT and HF schemes. Eventually, differences are linked to the type of basis sets employed in the calculation.

In the case of local basis sets, the direct application of Ewald technique is still a very time-consuming task, because it requires considering a high number of electron-electron integrals [49]. In the CRYSTAL code, an alternative approach is followed. For each atomic shell λ, the system is subdivided into two zones, in which the electron-electron integrals are either calculated explicitly or approximated by means of a multipolar expansion.

The criterion followed to determine whether an electron-electron integral should be calculated explicitly or approximated, is an estimation of the charges associated to the overlap distributions. Since the basis sets are local, such charges should decrease exponentially. A single normalized s-type gaussian is automatically associated to each atomic shell[3], and employed in the estimation of the charges. The exact calculation of the integral is avoided if either of the overlapping charges is smaller than a threshold parameter S_C. The accuracy of the approximation can be controlled by means of S_C, and actually the accuracy of the calculations performed with CRYSTAL depend critically on this issue.

Outside the region where integrals are calculated explicitly, the multipoles[4] $Y_l^m(\mathbf{r}_c;\{\lambda\})$ associated to the charge distribution ρ_λ referred to a point $\mathbf{r}_c$, are used to estimate the integrals by means of the expression:

$$(\{12\mathbf{g}\} \mid \{\lambda,\mathbf{h}\}) = \sum_{lm} Y(ln;\lambda)\,(\{12\mathbf{g}\} \mid \{(lm)_s,\mathbf{h}\}) \tag{2.18}$$

[3]These gaussians (named "adjoined gaussians") are chosen so their exponent reproduces approximately the absolute value of the corresponding atomic orbitals at intermediate and long-range. In practice, the exponent is assumed to be the smallest of the exponents of the gaussian type orbitals. A single adjoined gaussian is used for all atomic orbitals in one shell.

[4]The multipole $Y_l^m(\mathbf{r};\{\lambda\})$ corresponds to the unnormalized complex solid harmonic

$$Y_l^m(\mathbf{r};\{\lambda\}) = x_c P_l^{|m|} cos(\theta) exp(im\phi)$$

with integers $l \geq 0$ and $|m| \leq l$. Here, x_c, θ and ϕ are spherical polar coordinates of the vector $\mathbf{x}_c = \mathbf{x} - \mathbf{r}_c$ and $P_l^{|m|}$are Legendre functions.

here the indices 1,2 account for atomic orbitals inside the unit cell, and both **g** and **h** are translation vectors. Such integrals correspond to the interaction between the distribution $\{12\mathbf{g}\}$ and the shell distribution $\{\lambda, h\}$ in terms of the multipoles of the later with respect to its center at $(\mathbf{h} + \mathbf{s}_\lambda)$. Ewald technique can then be employed efficiently for estimating the contribution from the non-analytical region.

2.3.2 Treatment of exchange and correlation.

There are deep differences between the HF and DFT schemes in what concerns the way exchange and correlation interactions are considered. This reflects directly over the design of the algorithms.

In the case of the DFT techniques, the contributions due to exchange and correlation are taken into account by means of a functional of the density. Such a functional has the shape of a multiplicative operator. The evaluation of such term, no matter how complicated is the form of the functional, is in any case relatively simple and unexpensive in terms of computational cost.

Correlation interaction is not treated explicitly in the HF equations[5], but the exchange terms have non-local character, and their evaluation is a complicated task. This is in fact one of the main reasons why there are few implementations of the HF scheme applied to periodic models. Once more, we will focus now in the methodology employed in the CRYSTAL package.

The exchange series does converge without displaying any significant problem. In infinite systems, it is still necessary to impose the truncation of the series (otherwise infinite).

The contribution to the total energy due to exchange interaction is

$$E_{ex} = \frac{1}{2} \sum_{1,2,\mathbf{g}} P_{12}^{\mathbf{g}} \left[-\frac{1}{2} \sum_{3,4,\mathbf{n}} P_{3,4}^{\mathbf{n}} \sum_{\mathbf{h}} (\{10; 3\mathbf{h}\} \mid \{2\mathbf{g}; 4\mathbf{h} + \mathbf{n}\}) \right] \tag{2.19}$$

where each numerical index is an atomic shell, and each vector indicates the translation vector associated to the cell to which the shell belongs. $P_{ij}^{\mathbf{k}}$ are elements of the density matrix

$$P_{ij}^{\mathbf{k}} = \sum_n \int_{BZ} d\mathbf{k} exp(i\mathbf{k} \cdot \mathbf{g}) a_{in}^\star(\mathbf{k}) a_{jn}(\mathbf{k}) \theta \left[\varepsilon_F - \varepsilon_n(\mathbf{k})\right] \tag{2.20}$$

where $a_{jn}(\mathbf{k})$ is the coefficient of the j-th bloch function in the n-th crystalline orbital at point k, and its associated eigenvalue $\varepsilon_n(\mathbf{k})$.

The evaluation of equation 2.19 requires selecting which of the integrals should be included. The convergence of the exchange energy with respect to the **h** vector is fast, since the overlap distribution decreases exponentially with increasing **h**. The

[5]The Restricted Hartree-Fock (RHF) theory does not account for correlation effects at all. In the Unrestricted (UHF) scheme, however, different sets of orbitals are considered for each spin population, to allow treating open shells. Because of this, rigurously, correlation effects are partially taken into account in the UHF scheme.

situation is more complicated for the **g** and **n** indexes. The reason is to be found when considering terms at long distance, since the number of terms per unit distance increases as $|\mathbf{g}|^{d-1}$ with d the dimensionality of the system.

In the scheme implemented in CRYSTAL the integrals involved in the evaluation of the exchange interaction are selected following a procedure similar to that employed in the case of the Coulomb series (described in previous subsection). The summations in both **g** and **n** are truncated when the overlaps $\int dr\chi_1^{\mathbf{0}}\chi_2^{\mathbf{g}}$ and $\int dr\chi_3^{\mathbf{0}}\chi_4^{\mathbf{n}}$ are smaller than a given threshold value, R_{ex} which is also a parameter to be defined at running time, and is also determinant of the accuracy of the calculation.

There is an important consequence of using this selection criterion, which affects directly the quality of the basis sets that can be employed in the calculations. There is no distinction between the indexes **g** and **n** when the criterion is applied (i.e. the truncation is symmetrical with respect to **g** and **n**). But in equation 2.19 the summation over **n** is performed for each value of **g**. For a given **g**, the truncation is such that all integrals within a sphere of radius R_g centered in **g** are considered, while at the same time all integrals within a sphere of radius R_n centered in the origin are considered. For small values of **g**, the whole sphere of radius R_g is contained within that of radius R_n. However, for **g** values close to the range of the density matrix, the center of the sphere of radius R_g is close to the boundary of that with radius R_n. This causes the exclusion of some exchange terms and the potential appears distorted. During the self-consistent cycle, this causes the density matrix of a given cycle to react in the variational direction against the distortion caused in the previous iteration, eventually taken the calculation to divergence (or non-convergence).

This problem becomes more evident when employing basis sets with big variational freedom, since the "reaction" of the density matrix becomes bigger. This limits the possibilities to use basis sets with many basis functions, in contrast with the case of molecular systems, where increasing the variational freedom of the basis set usually increases the quality of the results. Similar difficulties are found if the basis set contains diffuse functions (gaussians with small α exponents). In this case, the problem is that the contribution of the diagonal elements which are disrregarded is more significant.

2.4 Description of the electron density: Basis sets and pseudo-potentials.

The choice of basis set is one of the most fundamental factors determining the quality of the results of an electronic structure calculation. In practice, the computer codes are adapted as to consider specific types of basis functions, with the aim of achieving efficient computations. The programs that have been used in this study employ either Gaussian (CRYSTAL98) or plane wave (VASP) basis sets. Several issues concerning each type of functions will be discussed next.

2.4.1 Gaussian basis sets.

This type of functions are a popular choice in molecular applications, especially for convenient expression of multi-electron integrals, and despite the fact that some other types of functions, such as Slater type orbitals, display a better behavior in the vicinity of the nuclei. Gaussian type-orbitals are actually composed as a product of an angular function and a radial function

$$g(\mathbf{r}) = Y_{lm}(\theta,\phi)R(r) = Y_{lm}r^l \exp(-\alpha r^2) \tag{2.21}$$

where l, m are integers $l \geq 0$ and $|m| \leq l$, r, θ and ϕ are spherical polar coordinates. The parameter α in the exponent of the radial functional is a variational parameter. Generally, the construction of a basis set implies the optimisation of the α parameters for all the gaussian functions forming the basis functions.

The orbital is usually constructed as a linear combination (contraction) of several Gaussians

$$\varphi_\mu(\mathbf{r}) = \sum_j d_{j\mu} g_j(\mathbf{r}) \tag{2.22}$$

this allows preparation of basis sets with well behaved orbitals. The coefficients d_j are also optimised during the construction of the basis sets. All the gaussians of a contraction are centered at the same point, typically the position of a given nuclei.

Gaussian orbitals depend directly on both the nature and the position of nuclei. Given their local character, further manipulation is needed in order to be able to use them as basis sets for periodic systems. It is possible to generate sets of Bloch functions[6] using gaussian orbitals

$$\phi_\mu^\kappa(\mathbf{r}) = \sum_T \varphi_\mu(\mathbf{r} - \mathbf{A}_\mu - \mathbf{T}) \exp(i\kappa \cdot \mathbf{T}) \tag{2.24}$$

here A_μ notes the position of the atom to which the gaussian orbital belongs. Once the basis set of the atoms contained in the unit cell is defined, it is reproduced by translation over the equivalent atoms in the different cells.

When using gaussian basis sets, the cost of the calculations is independent of the nature of the orbital, and therefore core electrons can be described as easily as valence electrons (this contrasts with the difficulty on describing the core electrons when using plane wave basis sets). Still, pseudo-potentials can be employed to simulate

[6] It is possible to construct functions that intrinsically reflect the periodicity of the system (Bloch functions)[50], that is:

$$\psi_n^k(\mathbf{r} + \mathbf{T}) = exp(ik \cdot \mathbf{T})\psi_n^k(\mathbf{r}) \tag{2.23}$$

where $\mathbf{T}$ is a translation vector. The index n labels the different, discretely spaced, eigenvalues, similarly to the molecular case. The new index k represents a vector in the Brillouin zone (the region of the reciprocal space that contains the set of non-equivalent points which are closest to the origin). Each vector k is therefore equivalent to (k+$\mathbf{K}$), where $\mathbf{K}$ is a translation vector of the reciprocal lattice. Therefore, there is as many solutions as there is non-equivalent k points. This number is strictly equivalent to the number of cells contained in the solid. If the crystal is assumed to be infinite (as is the case), then k can taken as all the possible values inside the Brillouin zone, and the eigenvalues ε_n^k become continuous functions of k. The spectra of ε_n^k forms the band structure of the solid.

atomic cores and reduce the cost of the calculations.

Gaussian orbitals are not orthogonal to each other, and this complicates the algorithms for evaluation of the integrals. A brief description of how this is done can be found both in [27] and [30].

The construction and optimisation of the gaussian basis set requires, as shown, bigger effort than the definition of a plane wave basis set. In general terms, for at least each atomic species contained in the unit cell, it is required to define the number of orbitals, their symmetry (s, sp, p,...) and the set of coefficients and exponents of each contraction.

In general terms, gaussian sets employed in molecular systems cannot be employed straight away on the study of periodic systems. The reason is that in molecular systems very diffuse functions are required to allow a correct description of the decay of the wave function, since the molecule is surrounded by vacuum. In periodic systems, this situation does not exist, but rather the use of diffuse functions rapidly increases the number of integrals that need to be evaluated, and the tails run into regions where there is already large variational freedom due to the presence of the basis sets of atoms in adjacent cells. The situation just described above, where two different orbitals overlap, leads to an overestimation of the binding energies (Basis Set Superposition Error, BSSE). This problem is well known in the case of molecular systems, and several methods have been developed to estimate corrections (see for instance [51]). Much less work has been done in the case of extended systems. Additionally, in the particular case of the HF calculations (like those performed with CRYSTAL98), diffuse functions cause numerical instability in the algorithms that evaluate the exchange interaction (see 2.3).

There is, finally, a little comment on the notation adopted here to describe Gaussian basis sets. In general, the standard notation has been employed[7]. The name of each set has the form $X - Y_1Y_2 \ldots Y_nG\star$ where the number X indicates how many Gaussians are contained in the core shells; there is as many Y_i numbers as valence shells, each one indicating the number of Gaussians contained in the respective shell; and the $\star$ symbol, if present, indicates that extra diffuse shell has been included. As an example, the 6-31G$\star$ set contains 6 Gaussians in the core shells, two valence shells with 3 and 1 Gaussians, respectively, and a diffuse shell.

2.4.2 Plane wave basis sets.

Plane waves are natural candidates as basis sets in periodic models, since they automatically satisfy the Bloch condition. The formal definition of a plane wave is the following:

$$\phi^k_{\mathbf{K}}(\mathbf{r}) = \Omega^{-\frac{1}{2}} \exp[i\mathbf{r} \cdot (k+\mathbf{K})] \tag{2.25}$$

where $\Omega^{-\frac{1}{2}}$ is a normalization factor, κ is a point within the Brillouin zone, and $\mathbf{K}$ is a vector of the reciprocal lattice.

[7]The only exception to this notation will be the STO-3G basis set, which is an adaptation of the Slater type basis functions that employs Gaussians per primitive.

The set of plane waves (PW) is complete and orthonormal, and therefore any continuous, normalizable function can be expanded with arbitrary precision in the PW basis set. Obviously, practical applications require to use finite sets, and therefore it is necessary to impose a selection criteria. The number of PW contained in a given basis is usually selected by defining a *cut-off* energy, E_{ctff}: for each k in the Brillouin zone, such that all the PW that satisfy

$$(k+\mathbf{K})^2 < E_{ctff} = R^2 \tag{2.26}$$

are included. Here R corresponds to the radius of the sphere, in the reciprocal space, centered at the origin. The finest detail that can be revealed for a given E_{ctff} is in the order of $1/R$. In practice, the number of plane waves required for describing correctly the rapid changes of the orbitals of core electrons becomes prohibitive. Instead, pseudo-potentials can be used for simulating the core electrons (this issue will be discussed in subsection 2.4.3).

A main characteristic of the PW set is that the functions are not associated in any way with the either the position, or the nature of the atoms contained in the unit cell. This situation is especially desirable when employing methods where the position of the ions varies constantly (for instance, molecular dynamics or structural relaxation). However, the number of plane waves contained in a cell depends both on the shape and volume of the unit cell. The other main advantage, as already has been mentioned, is the convenience and simplicity of most of the mathematical expressions, especially if the momentum representation is used.

2.4.3 Pseudo-potentials.

Most of the fundamental properties of matter can be justified in terms of the chemical bonding that links ions together in molecules and solids. Only the electrons of the outer shells of those ions participate actively in this bonding, and therefore a detailed description of the inner (core) electrons is not necessary in many cases. On the other hand, a significant part of the cost of the calculation is linked to the description of core electrons (this is especially true when plane wave basis sets are employed). One possible solution, that allows avoidance of the description of the core electrons, while simulating their presence (so the wave function in the region outside the core is correct), is to introduce pseudo-potentials.

The procedure for the introduction of the pseudo-potentials is well established. We consider here the DFT scheme to illustrate it. The implementation in the HF is, however, strictly parallel.

The one-electron Hamiltonian is partially modified, by introducing a new term V_{ps} that accounts for the nuclear attractions, already screened by the core electrons. For a system containing A nuclei and n electrons, from which n' are valence electrons, the one-electron equations become

$$\left[-\frac{\nabla^2}{2}+\int\frac{\rho'(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'+\mu_{exc-corr}(\mathbf{r};[\rho'])+\sum_A V_{ps(A)}\right]\psi_i'(\mathbf{r})=\varepsilon_i'\psi_i'(\mathbf{r}) \qquad (2.27)$$

This equation is identical to the all-electron DFT equations, apart from the substitution of the potential by the pseudo-potential operator. The primes symbols indicate that the solution of the equation is limited to valence electrons only. Therefore, only n'/2 orbitals are considered, and the valence density $\rho'(\mathbf{r}')$ is used to define the coulomb, exchange and correlation contributions to the total energy.

The new terms $V_{ps(A)}$ are also in charge of keeping the orthogonality between core and valence electrons (Exclusion Principle). The general expression of a pseudo-potential under these conditions is:

$$V_{ps(A)}=-\frac{Z_A-n_A^{core}}{|\mathbf{r}-\mathbf{A}|}+\left\{\sum_l U_l^{s-r}(|r-A|)\right\}_A+W^{s-r}(|r-A|) \qquad (2.28)$$

here Z_A represents the charge of nuclei A, n_A^{core} the number of electrons contained in the core of atom A. Both $U_l^{s-r}(|r-A|)$ and $W^{s-r}(|r-A|)$ are functions with short-range character, that can be defined in different ways.

The pseudo-potentials are usually fitted to a basis of atomic data obtained using all electron first-principles calculations of isolated atoms or ions. Several conditions are applied, and fulfilled as best as possible: a) Pseudo-valence eigenvalues have to coincide with true eigenvalues, b) the pseudo-orbitals should be as similar as possible to the true ones, in the region outside the core. This is usually achieved by requiring the pseudo-wavefunction to be continuously differentiable at least twice at a cut-off radius that defines the core region:

$$\phi_{l\epsilon}^{ps}(r)^{(n)}\,|_{r=R_c}=\phi_{l\epsilon}^{AE}(r)^{(n)}\,|_{r=R_c} \qquad n=0,1,2,\ldots \qquad (2.29)$$

where $\phi_{l\epsilon}^{AE}(r)^{(n)}$ is the solution of the radial Schrödinger equation for an specific energy ϵ:

$$\left[\frac{\hbar^2}{2m}\left(\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}\right)+V(r)-\epsilon\right]\phi_{l\epsilon}^{AE}(r)^{(n)}=0 \qquad (2.30)$$

Often, it is also required that the charge enclosed inside the core region should be the same for the pseudo wavefunctions and the all-electron wavefunctions (norm-conserving condition).

$$\int_0^{R_c}\phi_{l\epsilon}^{ps}(r)^2dr=\int_0^{R_c}\phi_{l\epsilon}^{AE}(r)^2dr \qquad (2.31)$$

Obviously, a minimum set of four adjustable parameters is required to fulfill these conditions, but practical implementations require additional parameters used for improving convergence of the pseudo-wavefunctions expanded in the basis of plane

waves [52].

In the calculations presented here, pseudo-potentials have been used only for the calculations in which plane wave basis sets have been employed (chapters 5 and 6). The VASP 4.4. code allows one to use ultra-soft pseudo-potentials, allowing a significant reduction in the number of plane waves in the basis set. The original concept of this type of pseudo-potentials can be found in the work by Vanderbilt [53], and details on how they have been employed inside VASP can be found in [52]. The high efficiency of this type of pseudo-potential is achieved by relaxing the norm-conserving constraint, but then a generalized eigenvalue problem must be faced. The lack of norm-conserving constraint leads to a deficit between the charge described by the pseudo-potential and that of the all-electron wavefunction, that should be compensated by means of an augmentation function. In the implementation in VASP, these augmented functions are constructed employing information from norm-conserving pseudo-potentials. This procedure (which differs from the original proposal by Vanderbilt) allows one to generate ultrasoft-pseudopotentials with very similar behavior to their norm-conserving counterparts. There is an issue on the transferability of the pseudo-potentials, i.e. if they can be used for target systems that have not been used to fit the pseudo-potentials. Kresse and Hafner show in their original work [52] how good transferability can be achieved if the reference norm-conserving pseudo-potentials are of good quality. In any case, the validity of the pseudo-potentials will be tested, when used, by checking how accurate host systems can be described (i.e. by checking the structural and electronic properties of perfect systems, free of defects).

In general terms, the use of pseudo-potentials leads to satisfactory results. Particular cases in which this method cannot be applied are the study at very high pressures, or other processes in which core electrons play a key role, like the hyperfine interaction (which is the source of structural information in Electron Paramagnetic Resonance experiments), or high-energy (X-ray) spectroscopy, which involves excitation of core electrons.

2.5 A semi-empirical approach. The INDO method.

In contrast to the first-principles techniques, the semi-empirical schemes implement approximations with the aim of reducing the cost of the calculations. Generally, the approximations imply a simplification of the mathematical method, by applying some criteria with a physical meaning. In the words of Pople [54], "semi-empirical methods no longer attempt to derive the properties of the system directly from the principles of quantum mechanics, but rather seek to interpret correlations with experimental data". There are many semi-empirical methods with many different conceptual backgrounds, and providing an overview of all of them is beyond the scope of this introduction. Instead, the semi-empirical method known as Intermediate Neglect of Differential Overlap (INDO) will be reviewed in closer detail, since it is

the only one employed in part of the research of the Thesis.

The INDO method is actually part of a wider family of semi-empirical approaches, based on the zero differential overlap approximation, and developed mainly by Pople [8] and co-workers. The main derivations can be found in the series of publications [56, 57, 58], and a very good review of the methods, with examples of its implementation into computer code, can be found in the book by Pople and Beveridge [54].

The basis for any of these semi-empirical schemes is the Hartree-Fock theory, which is used to find the best single-determinant approximation to the ground state wave-function. In order to save computing resources, the semi-empirical schemes consider valence electrons only. The molecular orbitals ψ_i^γ are defined as a linear combination of atomic orbitals, ϕ_μ

$$\psi_i^\gamma = \sum_\mu c_{\mu i}^\gamma \phi_\mu \tag{2.32}$$

where i runs over the number of molecular orbital, μ over the number of atomic orbitals, and γ accounts for the spin ($\gamma = \alpha, \beta$). The basis of atomic orbitals in this particular case is composed of Slater functions[9]. The matrices, $\mathbf{C}^\gamma$, of linear coefficients, $c_{\mu i}^\gamma$, can be determined by solving the (matrix) equations

$$\mathbf{F}^\gamma \mathbf{C}^\gamma = \mathbf{C}\mathbf{S}^\gamma \mathbf{E}^\gamma \tag{2.33}$$

were $\mathbf{E}^\gamma$ is the matrix of orbital energies, $\mathbf{S}^\gamma$ is the overlap matrix ($S_{\mu\nu} = \int \phi_\mu^\star \phi_\nu$) and F^γ is the Fock matrix, whose elements are defined as

$$F_{\mu\nu}^\gamma = H_{\mu\nu}^{core} + \sum_{\lambda\sigma} [P_{\lambda\sigma}(\mu\nu|\lambda\sigma) - P_{\lambda\sigma}^\gamma(\mu\lambda|\nu\sigma)] \tag{2.34}$$

were $H_{\mu\nu}^{core}$ represents the kinetic energy of electrons and their interaction with the core:

$$H_{\mu\nu}^{core} = \int \phi_\mu^\star(1) \left[-\frac{\nabla^2}{2} + \sum_A^{atoms} \frac{Z_A}{r_{1A}} \right] \phi_\nu(1) d\tau_1 \tag{2.35}$$

note that this quantity depends on the coordinates of one electron. The integration is performed over the spatial coordinates. Z_A corresponds to the charge of the ion A and r_{1A} is the distance of the electron (1) to the nucleus A.

The rest of the terms on the right hand side of equation 2.34 correspond to the interaction between electrons. The $P_{\lambda\sigma}$ and $P_{\lambda\sigma}^\gamma$ are elements of the total charge

[8] John A. Pople has been awarded the Nobel Prize in Chemistry in 1998 "for his development of computational methods in quantum chemistry". He is also co-author of the GAUSSIAN code, which is among the most advanced and widely used computer programs for electronic structure calculation [55].

[9] Slater orbitals, similarly to Gaussian orbitals, are composed as a product of an angular and a radial function. The functional form of the radial function employed in the INDO method will be shown further in the text. The Slater orbitals are not orthonormal, although an orthogonalized set can be constructed following a simple procedure [54]. The use of Slater orbitals in modern *ab initio* routines is not popular due to the difficulty to evaluate the electronic integrals, when using this type of basis set.

density and the charge density associated to the electrons with spin γ

$$P^{\gamma}_{\lambda\sigma} = \sum_{i}^{p} c^{\gamma}_{\lambda i} c^{\gamma}_{\sigma i} \tag{2.36}$$

$$P_{\lambda\sigma} = \sum_{\gamma} P^{\gamma}_{\lambda\sigma} \tag{2.37}$$

and $(\mu\nu|\lambda\sigma)$ are the integrals

$$(\mu\nu|\lambda\sigma) = \int\int \phi^{*}_{\mu}(1)\phi^{*}_{\lambda}(2)\frac{1}{r_{12}}\phi_{\nu}(1)\phi_{\sigma}(2)d\tau_1 d\tau_2 \tag{2.38}$$

here τ_i correspond to the spatial coordinates of electron i.

The evaluation of the $(\mu\nu|\lambda\sigma)$ integrals is one of the most time consuming parts of the calculation, since it requires one to consider all one-, two-, three- and four-center integrals. Most of these integrals, however, have values close to zero, since they represent the interaction due to orbitals with very small overlap between them. The Zero Differential Overlap Approach proposes a method for systematically neglecting those integrals with very small contribution. In fact, the criteria of selection can be applied at different levels, that lead to different techniques:

- **Complete Neglect of Differential Overlap (CNDO).** All the integrals that depend on the overlapping charge density of different basis orbitals are neglected (i.e. $(\mu\nu|\lambda\sigma) = 0$ unless $\mu = \nu$ and $\lambda = \sigma$).

- **Intermediate Neglect of Differential Overlap (INDO).** The one center integrals representing exchange interaction are preserved (although parametrised), allowing to differenciate states of different spin multiplicity. All other integrals are set to zero unless $\mu = \nu$ and $\lambda = \sigma$.

- **Neglect of Diatomic Differential Overlap (NDDO).** The integrals are only neglected if the μ and ν belong to different atoms. Therefore, all two-electron two-center integrals are retained. This technique allows a better reproduction of the dipole-dipole interactions.

The INDO scheme, which has been employed here, is the simplest of these techniques that allows resolution of electronic states that belong to the same electron configuration. This is obviously due to the fact that the exchange interaction is not neglected.

The application of the Zero Differential Overlap Approximation requires introducing several extra approximations to guarantee that the results are invariant to rotations of the basis set and hybridization. In the particular case of the INDO method the approximations assumed are the following:

- Overlap integrals $S_{\mu\nu}$ are neglected unless $\mu = \nu$ (that is, overlap matrix is

assumed to be diagonal). This simplifies the set of equations to

$$\mathbf{F}^{\gamma}\mathbf{C}^{\gamma} = \mathbf{C}^{\gamma}\mathbf{E}^{\gamma} \tag{2.39}$$

This approximation is not applied, however, in the evaluation of the core matrix elements.

- Two-, three- and four- center integrals of the type $(\mu\nu|\lambda\sigma)$ are assumed to be zero unless $\mu = \nu$ and $\lambda = \sigma$. Remaining integrals are approximated by the expression:

$$(\mu\nu|\lambda\sigma) = \gamma_{AB} = (s_A s_A|s_B s_B) \tag{2.40}$$

where $\mu \in A$ and $\nu \in B$, and s_i is the Slater orbital (s-type) in atom i. The integral $(s_A s_A|s_B s_B)$ is therefore calculated explicitly.

- The interactions involving one orbital only (diagonal core matrix elements, $H^{core}_{\mu\mu}$) are evaluated by separating the part of the interaction of ϕ_μ with the nucleus in which the orbital is centered (A), and the interactions with the rest of nuclei.

$$H^{core}_{\mu\mu} = U_{\mu\mu} - \sum_{B \neq A} Z_B \gamma_{AB} \tag{2.41}$$

Note that the elements $U_{\mu\lambda}$, with $(\mu \neq \lambda;\ \mu, \lambda \in A)$ vanish in the cases where pure s and p orbitals are used as a basis, but are non-zero if hybrid orbitals are considered. The $U_{\mu\mu}$ elements were calculated in the original derivation by subtracting the electron interaction terms from an average between the electron affinity and the ionization potential [59]. In the version implemented in the MOSYM code (used in our calculations) the $U_{\mu\mu}$ elements are calculated explicitly.

- The two-center core matrix elements ($H^{core}_{\mu\nu}$) are approximated by the expression

$$H^{core}_{\mu\nu=} \frac{\beta_A + \beta_B}{2} S_{\mu\nu} \tag{2.42}$$

where $\mu \in A$ and $\nu \in B$. The β_A and β_B parameters depend only on the nature of the atoms A and B, respectively.

Once these approximations have been assumed, and considering a basis set formed with s and p orbitals (rather than hybrids) the elements of the Fock matrix reduce to the form:

$$\begin{aligned} F^{\gamma}_{\mu\mu} &= U_{\mu\mu} + \sum_{\lambda \in A} \left[P_{\lambda\lambda}(\mu\mu|\lambda\lambda) - P^{\gamma}_{\lambda\lambda}(\mu\lambda|\mu\lambda) \right] \\ &+ \sum_{B \neq A} (P_{BB} - Z_B)\gamma_{AB}\ ;\ \mu \in A \\ F^{\gamma}_{\mu\nu} &= (2P_{\mu\nu} - P^{\gamma}_{\mu\nu})(\mu\nu|\mu\nu) - P^{\gamma}_{\mu\nu}(\mu\mu|\nu\nu); \\ &\quad \mu, \nu \in A\ \mu \neq \nu \end{aligned} \tag{2.43}$$

$$F^{\gamma}_{\mu\nu} = \frac{\beta_A + \beta_B}{2} S_{\mu\nu} - P^{\gamma}_{\mu\nu}\gamma_{AB} \; ; \; \mu \in A, \nu \in B$$

where $P_{BB} = \sum_{\mu}^{A} P_{\mu\mu}$ represents the total valence electron density on atom B. Once these elements have been calculated, the system of equations can be solved by means of standard diagonalization techniques. Similar to the first-principles HF scheme, the equations need to be solved in a self-consistent way, since the elements of the Fock matrix depend directly on the molecular orbitals.

The last summation in the right-hand side of Equation 2.43 for the elements $F^{\gamma}_{\mu\mu}$ with $\mu \in A$, are the penetration terms, that represent the attractive interaction between electrons in the orbital ϕ_μ and the nuclei B, screened by all the electrons in B. The neglect of the differential overlap matrix, however, causes an overestimation of the penetration terms. In the original derivation by Pople and co-workers, this was compensated by assuming that the attraction with the nucleus cancels exactly the penetration terms ($V_{AB} = Z_B\gamma_{AB}$). Instead, in the modification proposed by Shluger [60], which is implemented in the code used in our work, it is proposed that the underestimation of the attraction between the nucleus B and the electrons in ϕ_μ is caused by the assumption that the cores have zero-radius (point charge approximation). In order to account for the extended nature of cores, the interaction between electrons in ϕ_μ and nucleus B is calculated by means of the expression:

$$V^{B}_{\mu\mu} = Z_B \left[\frac{1}{R_{AB}} - \left(\frac{1}{R_{AB}} - (\mu\mu|s_B s_B) \right) e^{-\alpha_B R_{AB}} \right] \tag{2.44}$$

the α_B parameters depend on the nature of atom B and on the symmetry (s, p, ...) of the orbital ϕ_μ.

In the implementation of the modified INDO scheme that we have used, some extra information, apart from the INDO parameters β_B and α_B, is required. The basis set is defined as single-exponent Slater functions, whose radial part is defined as

$$R_{nl}(r) = \frac{1}{\sqrt{(2n)!}} (2\xi)^{n+\frac{1}{2}} r^{n-1} e^{-\xi r} \xi \tag{2.45}$$

It is possible to define in the input the values of the exponents ξ for each different type of shell, allowing one to fit the valence basis set. On the other hand, in the initial step of the self-consistent calculation the code generates a Fock matrix by means of the (more approximated) Hückel method, where two parameters, the electronegativity E_{neg} and an initial population of the orbitals $P^{(0)}_{i,A}$, are used to estimate the value of the diagonal matrix elements. The complete set of parameters required as input for an INDO calculation is summarized in Table 2.1.

Notation	Meaning	Quantity
ζ	Exponent of the Slater orbital.	One for each type of orbital (s, p, d) of each atomic species.
E_{neg}	Electronegativity.	One for each type of orbital (s, p, d) of each atomic species.
$P^0_{i(A)}$	Initial population of orbital $i$in atom A.	One for each type of orbital (s, p, d) of each atomic species.
$\beta_{[i(A),j(B)]}$	Interaction between electrons in orbital of type $i$in atom $A$and electrons in orbital of type $j$in atom B.	One for each $[i(A), j(B)]$pair.
$\alpha_{i(A),B}$	Factor scaling the interaction of electrons in orbital of type $i$in atom $A$and nuclei N, (due to extension of nuclei).	One for each $[i(A), B]$pair.

Table 2.1: **The INDO parameters.**
All types of parameters required for performing an INDO calculation are listed here. For each type, there are indicated (from left to right) its usual notation, its meaning, and how many parameters of this type are required. Detailed explanation about how this parameters are used can be found in the text.

Chapter 3

Data analysis

Probably the most important, and difficult, part of a computational study is to determine an estimate of the accuracy obtained with the methods that have been employed. Electronic structure calculations give as primary information the total energy of the system (in both HF and DFT schemes) and either an approximation to the wave function (in the HF case) or the electron density (in the DFT case). All the other properties of the system are then derived from this information.

The rest of the section will indicate how different properties can, and have, been calculated in the work presented here. The subjects of study chosen are actually different in nature, and the properties analyzed in each case are also considerably different. In brief, this is the list of properties that have been analyzed in each one of the studies:

- **Study of the band alignment at the Si/SiO_2 interface (chapter 4).** The valence-band offset (change in energy of the edge of the valence band across the interface) is often used as a representative parameter of the barrier of energy found by the carriers at the junction. Given the technological interest of this issue, great attention has been devoted to develop models that will allow one to determine the band-offset, as has been shown in 2.2.3.

  The method employed here, described in 3.1, removes the need of a reference level for the energy, and provides enough spatial resolution to observe the change of the offset as a function of the distance in a direction perpendicular to the interface. Several factors condition the accuracy of the method. This issue will be discussed in depth in chapter 4.

- **Study of point defects in HfO_2 and ZrO_2.** The reliability of gate dielectrics depends critically on the ability for structural defects to act as charge traps. This issue is addressed here by characterizing a) the energy required for the defect formation (that determines their abundance in the material) and b) the energy required to charge defects (that determines their charge trapping ability).

  Formation of point defects often requires adding or removing an atom. The

calculation of atomic energies requires some comments, and are discussed in 3.2.1.

The process of charging is understood as the capture of an electron from the conduction band of the materials. The problem, in this case, is that the electronic structure techniques employed do not predict correct position of the unoccupied states of the system. The methods proposed here for the calculation of the electron affinities and ionization potentials include an empirical correction to calculated band gaps.

- **Study of stability and diffusion of hydrogen inside Silica.** Atomic hydrogen is a paramagnetic center, and therefore Electronic Paramagnetic Resonance (EPR) experiments can be employed for obtaining structural information, in this case possible stable sites for the atom inside silica. Apart from comparing directly the results of structural relaxations with the models deduced from EPR data, it is possible to estimate the value of several parameters that characterize the hyperfine interaction between the unpaired spin and the neighboring nuclei. The basis of the EPR technique and the how the hyperfine parameters can be calculated are discussed in 3.3.

- **Development of an INDO code.** The work described in chapter 7 is mainly focused on describing several modifications done to a code designed to perform INDO calculations, and very few physical results are analyzed. As an example of application to systems with very large number of atoms, however, the participation function of some silica structures are calculated. Participation functions allows one to evaluate the degree of localization of specific electronic states inside the system. They are described in 3.4.

3.1 Calculation of the band Offset through the Density of States.

A fundamental tool for the exploration of the electronic structure of a solid is the Density of States (DOS), that shows the number of electronic states contained in an energy interval. The calculation of this function can be done using several techniques, once the electron density has been determined. In this particular case, we have used the tool provided in the CRYSTAL 98 program, that is based on a Fourier representation of the bands, and an expansion of the DOS associated with each band into an orthonormal set of Legendre polynomials. The authors have selected such a technique because it provides analytical expressions for the integrals that need to be evaluated, and allows highly efficient calculations [27]. A detailed derivation of the method can be found in [61]. There is no account of thermal broadening in the DOS.

The DOS as such does not provide any spatial resolution, that is, it does not discriminate where the electron states with certain energy are placed. However, it is possible to project the DOS function into a given subset of electronic states. The

Projected Density of States (PDOS) is often used to see how specific electron states contribute to the band structure of the solid. Through a selection of these orbitals, it is possible to find the states associated with specific regions of space.

An example of the use of PDOS to study the electron structure as a function of the distance can be found in the particular case of the Si/SiO_2 interface, also studied here. Yamasaki and co-workers [6] use the result of plane-waves density functional calculations to plot the evolution of the bands across the interface. To do so, they subdivide the unit cell into "slices" parallel to the plane of the interface, and evaluate the PDOS associated to the density contained inside each one of this elements. In principle, such a method is applicable independently of the technique and basis set used, as long as the electron density is known. However, in the calculations presented here (done using gaussian basis sets) it is also possible to exploit the localized nature of the basis functions to make the selection in an intuitive way.

The methods using gaussian basis sets assume that the orbitals are constructed as linear combinations of gaussian functions that are centered in the positions of the nuclei. Therefore, it is straightforward to select the electron orbitals that belong to a specific ion, and construct a PDOS. In order to obtain as much spatial resolution as possible in the direction perpendicular to the interface, the PDOS associated with the orbitals of each single ion is determined. Then, the set of projections can be arranged in the same order that the ions have in the direction perpendicular to the plane of the interface, to obtain a representation of the evolution of the band structure along this direction.

To illustrate the procedure, an example is shown in Figure 3.1 (we avoid, for clarity, to give details about the electronic structure calculation). The plots are displayed in a range of energies so as to show the states corresponding to the top of the valence band. Note how the edges of this band are placed at different energies as one considers atoms at different positions along the interface, showing the evolution of the band as a function of the distance to the junction.

This picture of the smooth change of the electronic structure at the interface is very similar to the results obtained by Yamasaki *et al.* [39] and the experimental measurements by Müller [5]. One interesting feature of this method is that allows one to detect the region of the junction that contains interface states. As it can be seen in Figure 3.1, this region extends to the first layers of oxygen region.

From now on, all the discussion will be centered in the calculation of the valence band offset (VBO). The reason why the calculation of the conduction band offset will not be addressed using this method is that, as has been already mentioned, none of the first-principles methods applied to extended systems are able to provide a correct description of unoccupied states of which the conduction band is formed. The evaluation of the VBO is straightforward by comparison between the edges of the band in two plots corresponding to the PDOS of atoms at each side of the interface, as it is also shown in Figure 3.1.

The method just described allows an easy determination of the valence band

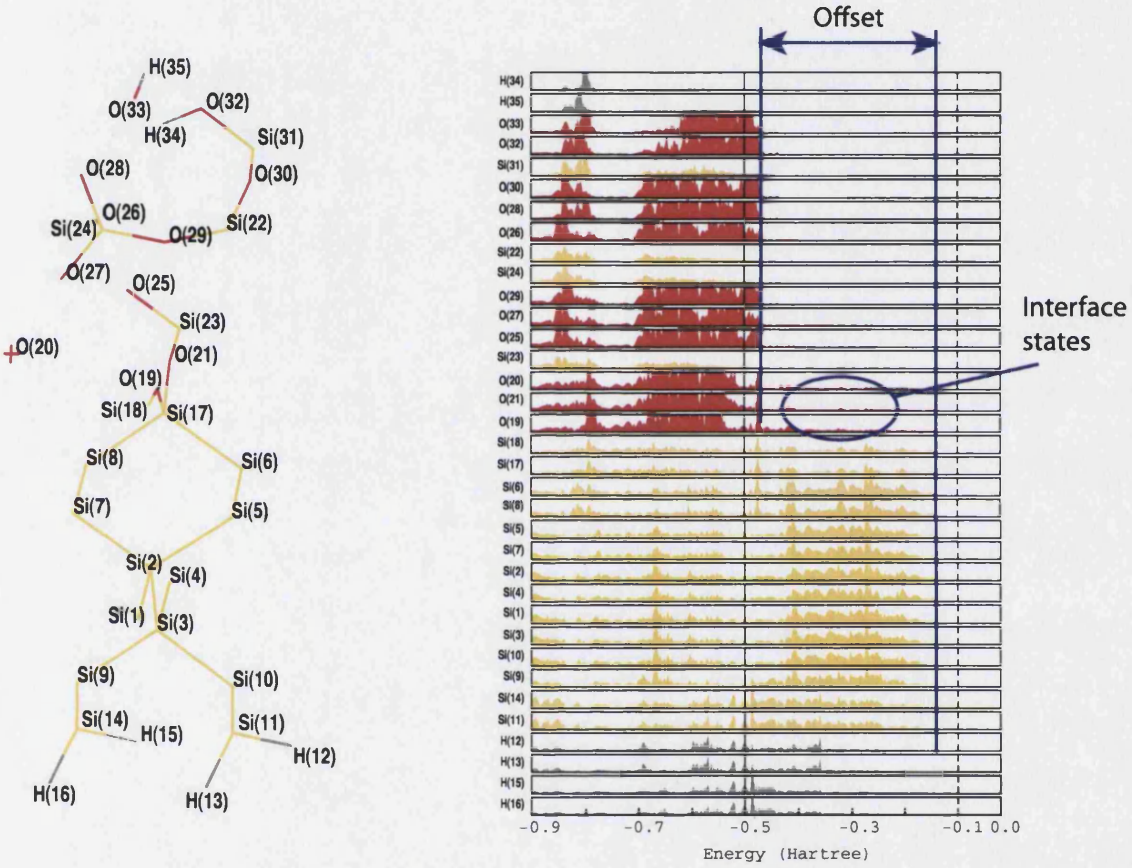


Figure 3.1: **First-Principles calculation of the valence band offset.** The evolution of the electronic structure across the interface can be explored by projecting the density of states over the basis function associated to each one of the atoms contained on the unit cell. The projections are then arranged in the same order as the atoms in a direction perpendicular to the interface. The band offset can be determined by locating the edges of the band at each side of the interface. The method allows also to locate the ions in to which the interface states are associated.

offset, with almost no additional cost to the calculation, and with the advantage of allowing an analysis of the band structure in terms of the chemical composition and bonding. However, it is obvious that the results will depend on the quality of the calculated electron density.

Since the calculation of the offset relies on obtaining the Density of States of the system, the accuracy of the results depend on the same factors that determine the quality of the electronic structure calculation. This study also intends to establish how these factors influence the results. More specifically, the comparative study will concern the choice of Hamiltonian, the quality of the basis set, and the model of interface.

3.2 Energetics of point defects.

We will discuss now several issues concerning calculation of energies involved in the processes of formation and ionization of point defects. The formation of a point defect often requires one to remove or add an atom to the system (which is assumed to be taken to, or from infinity). Evaluating the formation energy requires consideration of the total energy of this isolated atom, in order to keep the mass and energy balance correct. The evaluation of the atomic energies, when using periodic models, is actually tricky, as will be discussed in 3.2.1. In what concerns the ionization energies, the process involves the transition of an electron (at least) to or from the conduction band. Unfortunately, the position of the levels that form the conduction band is not predicted accurately with current methods. This problem, and a possible correction, will be presented in 3.2.2 and remaining subsections.

3.2.1 Calculation of atomic energies.

The evaluation of formation energies of some of the defects requires knowledge of the total energy of the neutral atoms. The calculation of these systems is non-trivial, since an incorrect representation of the exchange interaction and the single determinant wave-function implemented in the DFT-GGA Kohn-Sham scheme may eventually lead to incorrect predictions on the ordering of atomic electron levels. This happens especially in the case of electronic configurations close to degeneracy, like the triplet state of oxygen atom. In the plane wave implementation, there is an additional constraint on the density imposed by the shape of the unit cell.

In our calculations, we have to determine the energy of the atoms predicted by DFT with the smallest possible constraint. To do so, we used low-symmetry unit cells (rectangular and monoclinic). The size of the vectors on each cell was increased until convergence in the total energy was achieved.

3.2.2 Correction to the underestimation of the band gap.

We have calculated the ionization potentials and electron affinities for some of the defects, defining them as the energy required to move one electron to and from, re-

spectively, the bottom of the conduction band of the oxide. This is not necessarily the case in the real MOS structure, where the source of electrons may be the conduction band of silicon. The calculation of the affinities and ionization potentials for this alternative process is straightforward if the offset between the edges of the bands of the oxide and of silicon are known (see figure 3.3).

A real problem in the calculation of the affinities and ionization potentials appears because DFT techniques underestimate the width of the band gap considerably. We have assumed that the error introduced in our calculations has remained constant for all systems where the bulk material remains the same. We have estimated these errors by comparing the width of the gap of the perfect bulk with the experimental values. The theoretical gap in a band structure calculation can be found in two different ways:

- As the difference in energy between the highest occupied and the lowest non-occupied one-electron states.
- As the difference in total energies between different charged states of the supercell corresponding to the perfect lattice.

$$E_g(theor) = E(prf, q = -1) + E(prf, q = +1) - 2E(prf, q = 0) \qquad (3.1)$$

The change in the charge state modifies the population of the one-electron states, and therefore allows one to define the gap as a property of the occupied orbitals only. Note that the calculation of the charged states requires the introduction of the neutralizing background, possibly affecting the one-electron energies. A common problem also arises if the added charges are completely delocalised over the lattice; describing a metallic state often requires one to modify the computational setup (sampling of the reciprocal space).

None of these methods allowed us to obtain values of the band gaps that agree with experimental ones. By direct comparison of the two, we obtain the following correction.

$$\kappa = E_{gap}(Exp) - E_{gap}(Theor) \qquad (3.2)$$

where $E_{gap}(Exp)$ is the value of the experimental gap, and E_{gap}(Theor) is the value of the theoretical gap taken as the difference in total energies. E_{gap}(Theor) is calculated in this way (as opposed to the difference between one electron energies), because: a) then the gap is a property of the total energy and of the occupied orbitals only; b) there is no clear physical meaning of the one-electron energies in the DFT Kohn-Sham scheme, and c) we will be calculating the ionization potentials and affinities as a function of total energies as well.

These corrections will be applied to re-locate the position of the bottom of the conduction band of the bulk in the calculation of ionization potentials and electron affinities. The DFT technique describes accurately the valence band of the solid

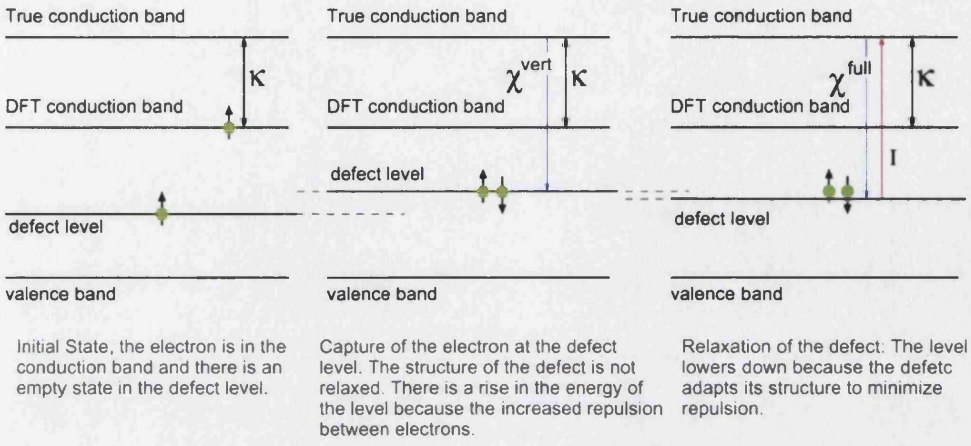


Figure 3.2: **Process of capture of an electron by a positively charged Oxygen vacancy.** The charged vacancy has only one electron in the level introduced in the gap. The electron comes from the top of the conduction band of the oxide. There is a difference between the true bottom of the conduction band and the calculated one, which causes the trapping energy to be underestimated by κ. This correction needs to be added with opposite sign to the ionization potential and electron affinity, since the flow of energy is opposite in these processes.

(formed by occupied states) and therefore further corrections are not necessary. Even if the method is clearly an approximation, it is possible to "update" the results if more accurate measurements of the band gap are made. This implies a shift of the predicted values by a constant. Note that the method is applicable only if there is no coupling between the defect and the states at the bottom of the conduction band. Otherwise, one should expect that both the defect levels and the states of the conduction band will shift together.

The sign of the κ correction is inverted in the affinities with respect to the ionization potentials. This is done to keep consistency with the criterion in the sign of the total energy described before (see figure 3.2 for a schematic representation of the process of trapping and releasing of electrons).

3.2.3 Calculation of the ionization potentials and electron affinities.

In chemistry, affinities and potentials are often defined as the energy required to extract or add an electron. The problem when simulating such process when using periodic infinite models is that there is no absolute scale of energies. On the other hand, in MOS devices electrons are usually coming from the metal or silicon layers, or from the conduction band of the oxide (see figure 3.3). As I already mentioned in 3.2.2, we have chosen the bottom of the conduction band of the oxide as reference energy.

Another important issue is that, unlike in Hartree-Fock methods, in the DFT-Kohn-Sham method the one-electron energies and orbitals cannot be given a direct physical meaning. This excludes, in principle, the possibility of obtaining the ionization potentials and affinities from a direct analysis of the spectrum of eigenvalues.

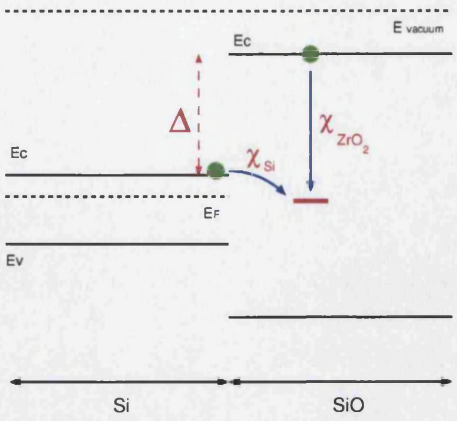


Figure 3.3: **Sources of hot charge carriers in a semiconductor/oxide interface.**

The electrons coming either from the conduction band of the silicon substrate or the conduction band of the oxide substrate can be trapped in a defect level (shown in red). The amount of energy required in each case (χ_{Si} and χ_{ZrO_2}) is different, but it is related through the band offset: $(\Delta) = \chi_{ZrO_2} - \chi_{Si}$

We have calculated the electron affinities and ionization potentials using the total energy of the systems directly. This is justified because DFT-Kohn Sham techniques have actually been constructed to calculate the ground state total energy accurately.

The naive approach would be comparing the total energies of two systems containing N and N±1 electrons. This, however, gives an incorrect description since there is no specification of the amount of energy of the extra electron. As mentioned before, the sources of electrons in the MOS system are well known (they are the conduction bands of silica and silicon) and the amount of energy that electrons have in these bands is also known.

For the calculation of the electron affinities χ (i.e. trapping of an electron), the electron is assumed to come from or to a state at the edges of the bands of the bulk oxide. This can be represented by calculating the electronic structure of a supercell of perfect lattice, with an extra electron. At this initial stage, a supercell containing the defect represents the vacancy in charge state q. In the final state, when the electron has been transferred from the bulk material to the defect, the final systems are represented by a neutral perfect lattice supercell and a supercell with the vacancy in charge state $q-1$. In microelectronics, it is assumed that both electrons and holes play the role of charge carriers [1] Therefore, a full characterization of the charge trapping ability should include the hole affinities. Holes are completely symmetrical to electrons, and thereby holes coming from the valence band are trapped into the defect state. In this case the perfect lattice supercell carries a positive charge in the

[1]The concept of a *hole* exists only because the valence band in a semiconductor or dielectric is entirely occupied, and therefore cannot conduct electricity. When one electron is taken away from this band, the empty state seems to have the properties of a positive charge carrier. A more detailed description of the properties of both electrons and holes can be found in many texts as, for instance, [7].

initial state, which is then transferred to the vacancy supercell in the final state.

The calculation of the ionization potential I_p, is symmetrical to the calculation of the electron affinity. Once more, two different systems are considered (a perfect cell and a cell containing the defect). In the initial state the vacancy holds an extra electron and the perfect lattice supercell remains neutral. In the final state, the electron has been transferred to the perfect lattice and the vacancy has charge q.

There is an issue concerning which defect structures should be considered, since the trapping of a carrier causes a rearrangement of the system to lower the repulsive energy. In the case of the positive charging process, we ensured that the vacancy containing the extra electron has had time to undergo the relaxation process.

In the case of negative charging, we have considered both the case where the final state of the vacancy is unrelaxed (vertical process), and the case where the vacancy is relaxed after trapping the electron. In the vertical process, both initial and final states have the same geometry: the relaxed geometry of the initial state. Note that the ionization potential and the relaxed electron affinity are symmetrical processes.

As was discussed in 3.2.2, one of the main limitations of the method comes from the fact that DFT-GGA calculations considerably underestimate the width of the gap. To avoid introducing this error we have applied the correction described in the mentioned subsection.

Using the total energies of the different supercells described above as data, the expressions that enable calculation of the defect levels are:

- Ionization potentials.

$$I(V_c^q) = E(prf, q = -1) + E(V_c^q) - E(prf, q = 0) - E(V_c^{q-1}) + \kappa \qquad (3.3)$$

- Electron affinities.

$$\chi_e^{proc}(V_c^q) = E(prf, q = 0) + E(V_c^{q-1}) - E(prf, q = -1) - E(V_c^q) - \kappa \qquad (3.4)$$

- Hole affinities.

$$\chi_h(V_c^q) = E(prf, q = 0) + E(V_c^{q+1}) - E(prf, q = +1) - E(V_c^q) \qquad (3.5)$$

Where $E(prf, q)$ are the total energies of the supercell containing a perfect lattice in charge state q, $E(V_c^q)$ are the energies of the vacancy with coordination c in charge state q. The $proc$ superscript in the electron affinities $\chi_e^{proc}(V_c^q)$ specifies whether the affinity describes a vertical ($proc = vert$) or a relaxed process ($proc = rlx$).

Note that the hole affinity does not require the introduction of the correction constant κ, since the process of trapping of a hole involves only the valence band, which is assumed to be correctly described by the DFT technique. Note also that the addition of the vertical electron affinity of the defect with charge q and hole affinity of the defect with charge $q + 1$ gives the width of the gap, as the difference in total

energies.

3.3 The Electron Paramagnetic Resonance technique.

The physical principle underlying the EPR experiments is the interaction between applied magnetic fields and the unpaired spins inside a given system. When an electromagnetic field is applied to a sample, the electrons experience a change of their energy that depends on their spin (see figure 3.4). Transitions between states of different spin are possible if the difference in energy ΔU is supplied, usually by means of another applied field perpendicular to the first. This technique, of course, makes sense only in systems where the spin populations are unbalanced. The presence of paramagnetic entities is detected as an absorption line in the spectra. This technique is also sensitive to the presence of nuclei with spin surrounding the paramagnetic centre, since they modify its energy because of the spin-spin interaction, as indicated also in figure 3.4. A very detailed description of the EPR technique can be found in the book by Weil, Boltzon and Wertz [62].

In principle, knowledge of the electron density does allow one to evaluate the hyperfine interaction over the ions. The interaction between nuclei with spin and the electronic spin can be defined by means of the spin-Hamiltonian $H = \mathbf{S} \cdot \mathbf{A} \cdot \mathbf{I}$. The equation is composed of tensors, since it is necessary to account for anisotropy caused by the presence of an external field. The tensors $\mathbf{S}$ and $\mathbf{I}$ stand for the spin of the system and that of the nucleus, respectively. The matrix $\mathbf{A}$ describes the coupling between the electron and nuclear spin. This matrix can actually be split into two terms:

$$\mathbf{A} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{bmatrix} = a_{iso}\mathbf{I} + \begin{bmatrix} B_1 & 0 & 0 \\ 0 & B_2 & 0 \\ 0 & 0 & B_3 \end{bmatrix} \tag{3.6}$$

here $\mathbf{I}$ stands for the unit matrix. The a_{iso} factor corresponds to coupling due to the presence of non-zero electron density at the nucleus. This is only true for s-type orbitals, which are spherically symmetrical and therefore make this contribution isotropic. The isotropic component can be calculated:

$$a_{iso} = \frac{2\mu_0}{3} g_N \beta_N g_e \beta_e < \rho^{spin} > \tag{3.7}$$

where μ_0 represents the permeability in vacuum, g_N and g_e correspond to the nuclear and electronic g factors and β_N and β_e correspond to the nuclear and electronic Bohr magnetons. The value of the expectation value of the spin density, $< \rho^{spin} >$ is the one that can be calculated by means of an electronic structure calculation. The $\mathbf{B}$ tensor represents the anisotropic contribution of the coupling. Although it is also possible to calculate this tensor from first principles results [30] we have not evaluated it in this study.

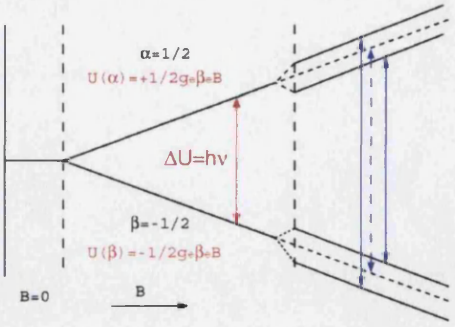


Figure 3.4: **Basis of the EPR experiments.**
When an external electromagnetic field is applied, there is a change in the energy of the electrons in depending of their spin. In paramagnetic systems, transitions between states are possible if the required energy ΔU is supplied. This is usually done by means of a second field, and presence of paramagnetic defects is detected as absorption in the second field spectra. Because of the interaction with spins of surrounding nuclei, there is a second and more subtle change in energy (hyperfine interaction). As a consequence, the energies that appear in the absorption spectra (solid blue), differ from the theoretical one (dashed blue) if no hyperfine splitting was present.

3.4 Calculations with systems of many atoms. Participation functions.

As an example of an application of the semi-empirical INDO method, I will present in chapter 7 an study of the localization of electronic states in amorphous silica. Such a type of analysis profits from the possibility of using periodic models containing hundreds of atoms.

One problem with the strong impact in semiconductor integrated circuits technology is the tunneling of electrons between the channel and the gate, which causes the degradation of the devices. It has been suggested that the presence of localized states in the oxide contributes to the tunneling process [63]. A very interesting study in the specific case of silica[2] has been presented by Ching along several publications[21, 22, 23]. Following the idea presented by this author, the use of one electron wave functions based in linear combinations of atomic orbitals (LCAO) offers a very powerful tool for detecting the presence of localized states, by means of a simple analysis of the calculated linear coefficients.

In the LCAO scheme each electronic orbital is defined:

$$\psi_i = \sum_\alpha c_{i\alpha}\phi_\alpha \tag{3.8}$$

and the c_{ij} coefficients are fitted during the SCF calculation. Following the probabilistic interpretation of Quantum Mechanics, if the electron state corresponds to a delocalized state, the electron will have equal probability to be found at any point in

[2]The study by Ching has been mentioned here because of the use that this author makes of the participation function analysis. Of course, there are many other authors which have focussed their efforts on studying amorphous silica (see, for instance, section 4.2.2, and the reviews by Wright and Hobbs in [64].

space and therefore all the coefficients squared (c_{ij}^2) should have very similar value. However, if the state is localized then the squared coefficients corresponding to the basis function ϕ_j localized in a specific region will have higher values than the rest.

In order to make use of this information, it is possible to define a participation function as follows:

$$P^{-1} = N \sum c_{i\alpha}^4 \tag{3.9}$$

where N equals the number of atoms. Note that if the electron state is delocalized, then $P = 1$, while if it is completely localized, then $P = \frac{1}{N}$.

Obviously, such a type of analysis requires one to have a system big enough so that you have true delocalized states, but this is not the only reason why the study of localization may require considering systems with a big number of atoms. As Anderson and Ching suggest [63, 65], the mechanism causing the localization of an electronic state is related to the symmetry of the potential, as it proves the fact that amorphous systems have states of this type, while in crystalline systems the states remain completely delocalized.

Part II

Presentation of the research.

Chapter 4

Electronic structure of the Si/SiO_2 interface

Aim. This chapter studies the possibility of determining directly the valence band offset at an interface by means of first-principles electronic structure calculation. The method is based on the analysis of the density of states projected over selected sets of orbitals. We aim to understand how different parameters of the calculations influence the results. In particular, we study the effect of using different Hamiltonians and different basis sets, as well as the dependence on the symmetry of the oxide. As a case study, we have selected the Si/SiO_2 interface. We address the technologically relevant problem of determining the minimum thickness of the gate oxide layer in the MOS structures.
Results. All the Hamiltonians tested provide the same qualitative description of the electronic structure, although the quanitative results differ from each other, and from the experimental value. The B3LYP functional as implemented in CRYSTAL gives the best agreement with experiment. Reliable basis sets require the inclusion of polarization functions. In our model of interface, the Si substrate is not thick enough as to provide the correct band structure of bulk silicon. The symmetry of the oxide layer influences the value of the offset; an oxide model with density closer to that of the amorphous phase is in better agreement with the experimental data. The interface states are located inside the oxide, along region several atomic layers wide. The size of this region is independent of the thickness of the oxide, and constitutes a limit on the thickness of the gate oxide layer.
Techniques employed. First Principles Hartree-Fock (HF). First-Principles Density-Functional Theory on its Local-Density-Approximation (DFT-LDA), and Generalized-Gradient Approximation (DFT-GGA). First-Principles Hybrid method (B3LYP).
Basis sets. Gaussian basis sets: STO-3G, 6-21G, 6-21G*, 6-311G*.
Models. Periodic slab models of the Si/SiO_2 interface, containing between 35 and 50 atoms.
Computer Facilities. Self-consistent calculations of the electronic structure done using the Cray-T3E supercomputer of the CSAR service at Manchester (UK). Post-processing (calculation of densities of states) done using a single processor workstation.

4.1 Introduction.

The development of planar micro-electronic devices relies entirely on the possibility of controlling the flow of charge carriers by alternating layers of materials with different electrical properties. At the interfaces, the carriers find energy barriers that keep them confined within certain regions of the device. The size of the barriers critically determines the performance of the system, and influences the possible improvements of the technology. The general tendency in planar micro-electronics is to rely on the

miniaturization of the device in order to improve its performance. There is, however, an emerging problem, linked to the change in the properties of the materials when they are disposed in too thin layers.

Up to the date, the most relevant heterostructure is the Si/SiO_2 interface, that plays a fundamental role in the Metal-Oxide-Semiconductor devices [1]. The oxide layer in the gate region of the MOS transistors is by far the thinnest layer in current circuits. Predictions are that, if miniaturization rate is maintained, in around ten years time the devices will become inoperative due to tunneling of carriers through the gate oxide and dielectric breakdown processes.

One of the main interests in the research of possible solutions is to develop a tool that will help to understand the mechanisms that cause the change in the properties of the materials, and to determine the ultimate limits to which current devices can be taken. The same tool will also be useful for designing new heterojunctions with very specific properties. Following this idea, there is a collection of band alignment theories that try to predict the values of the energy barriers at the junction. Many of these theories are only concerned with the chemical composition of the materials that form the junction, although some others also consider details of the micro-structure of the junction, like its roughness or the presence of defects.

First-principles electronic structure calculations can, in principle, be very useful for the type of studies mentioned above. Unfortunately, their elevated computational cost prevented their application to the study of heterojunctions until very recently. However, this situation is changing as the capability of computers increases and now considerably complex models of interfaces can be developed and studied.

In this chapter we have addressed the possibility of studying the energy barriers at an interface by means of first-principles calculation. The evolution of the electronic structure in a direction perpendicular to the junction is determined by studying projections of the density of states over selected sets of local orbitals. This method is a variant of that employed by Yamasaki and co-workers, where a plane wave basis set has been used. The use of localized sets, however, allows one to keep a chemical interpretation in terms of the position of the ions and their bonding. I think that such an interpretation can be useful for correlating certain features of the electronic structure to those of the microstructure of the system.

Since the method relies entirely on the accuracy of the *ab initio* calculations, it is necessary to explore how different factors that can be tuned in the calculation affect the results. This study focuses its attention on the effect of selecting different Hamiltonians, the effect of the quality of the basis set, and on the effect of different features of the models. In particular, we explore if the offset is affected by changes in the microstructure of the oxide, and by changes in the thickness of the oxide layers. I found these two factors especially interesting because they will help to understand if the complicated amorphous structure of the oxide can be simulated by simpler

[1] For a more detailed explanation on the working principle of the MOSFET devices, and the relevance of the Si/SiO_2 system, see section 1.2

crystalline models, in the first place, and to see if the energy barriers lower as the thickness of the oxide is reduces, in the second place.

As I have mentioned above, the Si/SiO_2 structure has special technological relevance, and it has been target of many studies, which have left a lot of information available. This is the reason why I have selected this system as a case study.

The chapter is organized as follows. Firstly, I present a bibliographic review of relevant issues. This includes the knowledge concerning the micro-structure and electronic structure of silicon dioxide and the silicon/silicon dioxide interface. Secondly, I discuss in detail the Hamiltonians, basis sets and models of interface in which I will base the comparative study. Thirdly, I describe in detail how the band offset can be calculated from the density of states of the system, and finally I present and discuss the results.

4.2 Bibliography review.

4.2.1 The role of silicon dioxide in planar microelectronics.

Silicon dioxide is almost the only material used, up to the date, as a gate oxide in silicon-based transistors [66]. Among other properties, the Si/SiO_2 system presents energy barriers high enough to prevent electrons and holes from moving across the oxide layer (see figure 4.1). This allows one to keep the carriers confined in the channel region, and establish the main current between source and drain[2].

The importance of the silicon-silicon oxide interface in the MOS technology is proven by a historical fact: the wide gap of almost forty years between the first patent of a MOS field effect transistor, by Lilienfield, and the first transistor of this type, which actually worked. The original devices did not work because charge was accumulating in the channel region near the interface. The use of the combined system of silicon and silicon dioxide overcame the problem. The charge accumulation problem was due to the existence of defects at the interface, caused by mismatch between the materials. This problem is greatly reduced in the Si/SiO_2 system because the fiexibility of the amorphous oxide allows one to minimize mismatching [67].

4.2.2 The structure of the silicon-silicon dioxide interface.

Despite being one of the most studied systems, the silicon/silicon dioxide interface is far from being well known. In general terms, it is not clear which are the structural features of the interface that determine the features of the electronic structure. A big effort is being paid in understanding which is the role of, for instance, the roughness of the interface, its stoichiometry composition or the presence of structural point defects like three-coordinated silicon ions, oxygen vacancies or interstitial hydrogen. The complexity of the problem is such that it is difficult to have a full view of the

[2]See section 1.2 for a more detailed description of the MOS transistor.

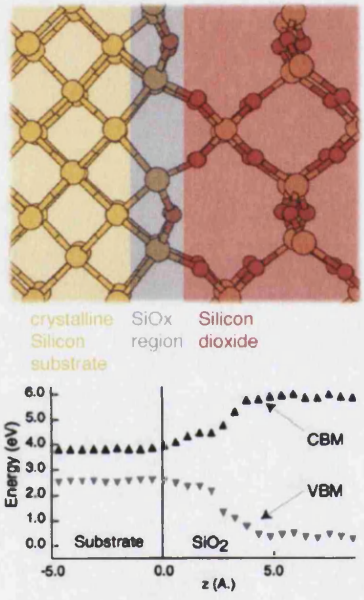


Figure 4.1: **The Si/SiO$_2$interface.**
Three types of regions can be identified: the crystalline silicon substrate, the SiO$_x$ region and the silicon dioxide region. At the SiO$_x$ region, the ions are not connected as in the bulk materials (for instance, Si ions bond two Si and two O), Although the junction can be very abrupt, the electronic structure (displayed below) evolves smoothly. The bulk properties, specially in the SiO$_2$ region, are displayed a few Å away from the interface. This figure is based in the work by Yamsaki and co-workers [6].

current situation. However, excellent reviews on the subject appear often, as, for instance those of references [10, 68, 3].

At present, one of the main problems of the simulation of the Si/SiO_2 interface is that there is not a representative model of its micro-structure. This is due to the fact that, in the MOS systems, silicon dioxide displays an amorphous structure, whose composition and structural features are difficult to characterize, both experimentally and theoretically. Despite this difficulty, many different approaches have been use to develop such a model.

In what concerns generation of the amorphous structure, the review by Wright (chapter 1 in [64]), lists three main strategies: crystal-based models, random network models and computer simulations. The crystal-based models use the fact that the local structure, to an extent of approximately 10 Å is very similar in almost all phases of silica, and justifies the main properties. The random network models are based on generating a structure based on the SiO_4 tetrahedral blocks, following certain rules of connectivity. The computer simulations involve a wide range of techniques. Models of the interface can be obtained by performing structural relaxation of reasonable initial configurations. In the molecular dynamics simulations, a crystalline structure is "melted" and then "quenched" to form the amorphous phase. In the reversed Monte Carlo methods networks structures are randomly generated and selected as to reproduce experimental data.

We will see now some examples of the methods listed before, applied to the specific case of generating the interface. The amount of such works is actually big, and therefore this list is not exhaustive, but rather intends to provide an idea of the evolution of the models and methodology.

Herman and co-workers have developed a model of the interface (described in the review by Helms, chapter 3 of [10]) after realization of the almost perfect matching between the Si(001) surface and the cristobalite (001) surface. The model constitutes an idealization of the true system, containing a perfectly flat interface, two crystalline systems and no defects, but does not reproduce correctly, however, the density of atoms in the SiO_x region (see figure 4.1 for a definition of this region). Pasquarello *et al.* also make use of crystalline-based interfaces for an analysis of photo-emission measurements [69]. A model of continuous random network was presented by Pantelides and Long [70]. The structure was constructed following several connectivity rules, and the free energy of the system optimised. The model shows that abrupt interfaces are possible, but the connectivity rules impose restrictions to the bond lengths and angles from the bulk SiO_2, while some experience with other interfaces shows that the situation may be considerably different (see [68]). Both crystal-based and random network models are in fact considerably old, and were designed without the present computing resources. Despite their success in justifying the main features of the interfaces, they describe too idealized surfaces. The use of computer simulation techniques has lead towards much more realistic models.

Monte carlo simulations have been employed by Ng and Vanderbilt [71], and

recently by Tu and Tersoff for constructing some models of Si/SiO_2 interface [72, 73]. The models by Ng and Vanderbilt display wide transition regions (around 20 Å thick) characterised by deficiency of oxygen atoms, which lead to th presence of positively charged Si species. Tu and Tersoff use a continuous network algorithm to generate the (defect-free) structures, and the Monte Carlo sampling to explore the thermodynamic ensemble of possible network topologies. The most relevant conclusion in this study concerns the key role played by the bringing oxygens beween each pair of Si atoms terminating the substrate region. This bridging oxygens eliminates half of the bonds of the Si surface, and therefore correct the mismatch between the bond densities in the two regions.

Work on reversed Monte Carlo simulations has been presented by one of its developers, McGreevy and co-workers [74]. The main objection to the use of this technique is that results depend directly on the experimental data used to generate the structure. Because of this, reversed Monte Carlo has often been considered as a fitting of 3N variables (the coordinates of N particles) to experimental data (see chapter one, in [64]).

Molecular Dynamics simulations can actually be considerably differently. In the classical scheme, some pair potentials are used to simulate ion interactions. The potentials are usually fitted to reproduce some data, either from experiments or from first principles calculations. Very large scale models of the interface, both flat and with steps and terraces, have been presented by Watanabe and Ohdomari [75, 76]. Their interatomic potentials were fitted as to reproduce *ab initio* calculations of energies and structures of small clusters and molecules. The results seem to be in good agreement with experimental measurements of the angular distributions. The success in reproducing the microstructure of the oxide is, however, just the very first step: the classical molecular dynamics technique does not provide any information concerning the electronic structure of the solid. Given the size of the models generated with this technique (Watanabe and Ohdomari models have around 4000 atoms), it is impossible to port the structures as a basis for a first-principles technique, even if no structural minimization is performed. Possible alternatives are to use a low-cost semi-empirical method, or to use the classical model of the interface as basis for embedding a quantum cluster.

The use of first principles molecular dynamics has also been employed widely in the generation of Si/SiO_2 interfaces. Demkov and Stankey have simulated a process of oxidation of a silicon substrate [46], within the size-limitations linked to the use of this computationally expensive technique (most of the models employed had side-lengths of 5.43 Å). Pasquarello and co-workers have dedicated considerable effort [77, 78, 69] to simulate realistic models of the interface using quantum molecular dynamics. In an earlier approach [79], the same authors performed structural relaxations using crystalline-based models as an initial configuration. The structural relaxation drives the system towards a more favourable configuration, but does not allow to explore the potential surface for different minima. In their publication [78],

a model of the interface is generated by simulating the oxidation of 3 layers of crystalline silica substrate. The oxidation process is actually simulated not by adding oxygen, but by putting together a piece of crystalline silica near the silicon substrate. The main problem with these simulations is that they were performed at a very high temperature ($\sim$ 6000°) to compensate for the short timescale (24 ps). This limitation is imposed by the cost of the computations at this level of theory. As a result, the structure obtained presented an oxygen deficient interface, which is corrected in a later model obtained using the same technique and presented in [77]. Additional oxygen atoms were added, preserving the bonding pattern but modifying the stoichiometry conveniently. The new model agrees with some experimental data.

A similar procedure is used by Yamasaki *et al.* on their work presented in [6, 39], although no detail concerning the molecular dynamics simulation is given in their publications. One of the targets of Yamasaki's study was to determine the stability of the different models as a function of the thickness of the oxide. The authors have done a systematic study of the stability of three different models of the interface, based on quartz, trydimite and cristobalite, as a function of the oxide layer thickness.

There are remarkable differences between the models presented by Pasquarello's group and the ones presented by Yamasaki's group. The latest model presented by Pasquarello agrees with certain experimental conditions, but still does not provide a description consistent with *all* experimental data. Yamasaki's models, present much higher symmetry, with clear resemblance to the crystalline phases. This is not extremely artificial since it has been shown that the region surrounding the interface presents to a certain extent crystalline structure. These interfaces are also very abrupt and do not justify properly some X-ray diffraction data. However, Yamasaki presents six different models constructed based on three different polymorphs of silicon. Each one of the phases is used for generating a model with a thin layer of oxide (2 monolayers) and a model with a thick layer of oxide ($\sim$5 monolayers). This collection of models, with so specific changes, is ideal for performing a systematic study of how the properties at the interface evolve as these changes take place. In fact, Yamasaki has shown how the stability of the interfaces changes with the thickness, and has compared how the bonding and angular distributions of each model compare with experimental data. However, these type of results are more oriented towards validating the models than to determine how different techniques and parameters in the model affect the picture of the interface that it is obtained.

4.2.3 The electronic structure of the Si/SiO$_2$ interface.

One of the main challenges on the construction of microscopic models of the interface is to relate the arrangement of ions with the electrical properties. A very interesting contribution from the experimental side has been presented by Muller *et al* in [5]. In this work the electrical properties are deduced from measurements of the unoccupied electronic states done using atomic-scale electron-energy-loss spectroscopy (EELS). A transmission electron microscope (TES) was used to access the states of internal

interfaces. The roughness of the interface was measured using X-ray reflectivity data. One important observation done during this research is that the tunnelling current through the oxide layer increases considerably if the interfacial regions from both sides of the oxide layer overlap. As the authors suggest and prove, the tunnelling reduces considerably if the thickness of the oxide is increased until both interface regions do not overlap.

The determination of this minimum thickness has been target of several works, both experimental and theoretical. In Muller's work, a structure with an oxide ~15 Å wide, represents an acceptable level of tunnelling current. This implies that the transition regions have widths smaller than half the value of this width (7.5 Å). Previous works reach similar conclusions; Alay *et al* [80] have found (by X photoemission spectroscopy) a constant value of the conduction band offsets of 3.2 eV, for oxide thickness larger than 16 Å and a value of the valence band of 4.36 eV for oxide thicker than 18 Å. This value of the offset is now quite widely assumed to be correct, as they follow similar values from a much earlier review by Balk [10]. Finally, Tsu *et al.* show that careful control of the thickness can allow smaller values (1.7 eV for the conduction band offset [81]).

Theoretical models for evaluating the tunnelling current through the oxide require an accurate determination of the band structure. This type of analysis has been performed for Städele *et al.*, [4] arriving to the conclusion that oxide layers as thin as 7 Å still present the correct band structure.

The tunnelling of electrons does not depend only on the thickness of the oxide layer, but also on the barrier height between SiO_2 and Si, i.e. the band offset. Yamasaki *et al.* show that in fact both the width of the band gap and the offsets evolve along the direction perpendicular to the interface [39]. In their model (and using the DFT-GGA technique for calculating the electronic structure), the oxide presents a slow change of the gap width during the first 2 Å and a more sudden change in the next 3 Å, until the properties of bulk SiO_2 are achieved. Tang *et al.* [8] used the same type of calculations (DFT-GGA), with less sophisticated models of the interface to evaluate directly the minimum thickness required to control tunnelling by keeping an appropriate value of the band offsets. They found that the value of the offset becomes constant in oxide layers with a thickness bigger than 4 monolayers (~7 Å). The valence band offset of the Si/SiO_2 interface has also been studied in the works by Demkov and Sankey [46] and, recently, by Tuttle [82]. Their results will be compared with ours in Subsection 4.3.5.

A final picture of the band structure of an ideal $Si\text{-}SiO_2$ system is shown in figure 4.2. The values that characterize the energy barriers, shown in this figure, have been taken from [10], and are widely accepted.

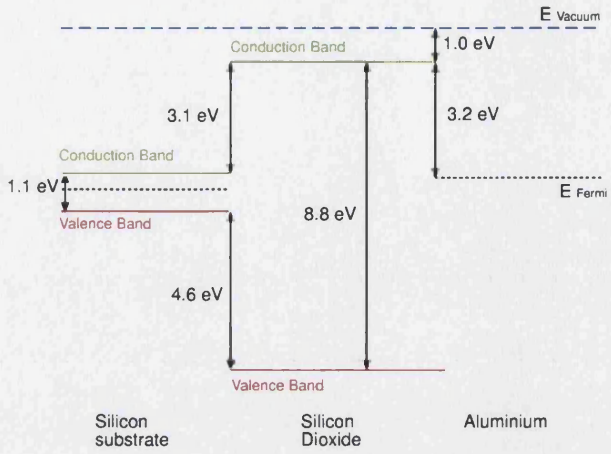


Figure 4.2: **Electronic structure of a MOS structure.**
This type of step-like diagram has been valid during many years as a representation of the electronic structure of a silicon-silicon dioxide-aluminium structure. However, the reduced dimensions of the devices in a short future will require to consider that the evolution of the bands is smooth along a direction perpendicular to the interface.

4.3 Presentation of the results.

4.3.1 Choice of Hamiltonians. The CRYSTAL 98 code.

As has already been introduced in 2.1, there are several first principles methods that allow calculating the electronic structure of extended systems, among which the Hartree-Fock (HF) and some of the Density Functional Theory (DFT) schemes are now of standard use. Recently, some hybrid methods combine features of both HF and DFT schemes with the aim of obtaining a better estimation of the contributions to the total energy due to electron exchange and correlation interactions.

Exploring the accuracy that some of these techniques can provide when describing a complex system such as the Si/SiO_2 interface, has been one of the aims of this comparative study. In particular, we have selected the HF method, two implementations of the DFT, the Local Density Approximation (DFT-LDA) and the Generalized Gradient Approximation (GGA), and the B3-LYP hybrid method. References where the detailed derivation of each one of these techniques can be found have been already given in 2.1.

The code employed in all the calculations of this chapter is CRYSTAL 98. The reason for using this package has been precisely that the code allows one to select the type of technique that is used for the evaluation of the electronic structure, without essentially modifying any other parameter of the calculation. The peculiarities of the method implemented in CRYSTAL 98, and of the use of local basis sets that are employed, have been also discussed in 2.3 and 2.4.1.

4.3.2 Choice of Basis Sets.

The other important factor to ensure accurate calculations is the quality of the basis set, as it should be large enough to provide flexibility for the fitting of the approximated solution to the wave function. However the use of periodic models makes a difference with respect to the molecular codes. CRYSTAL 98 employs a Gaussian basis from which the Bloch functions are then generated. The use of local basis sets in periodic systems has the inconvenience that very diffuse basis functions cause numerical instability during the calculation (see 2.3). Because of this, the quality of the basis set cannot be systematically improved by adding more diffuse basis functions, as is the case in molecular calculations.

Several works have already paid attention to the search for a reasonable basis set for the description of silicon dioxide. A very interesting example can be found in the work by Civalleri *et al.* [83]. Some of the most relevant conclusions are the following:

- The use of poor quality basis sets, with reduced variational freedom and small number of basis functions, keeps the electronic charge under constraint and, as a consequence, the system appears "too ionic". In extended systems, this effect leads to an increase in the intensity of the crystalline field, which raises or lowers the energy of the one-electron levels (depending on their occupancy).

- Since the main features of the SiO_2 system seem to be due to the local arrangement of atoms, it is essential to ensure that the basis sets are able to provide an accurate description of the local geometry. In this sense, the presence of polarization basis functions in the basis sets associated to the oxygen atoms seems to play a fundamental role in determining the value of the Si-O-Si angle.

I have considered the points described above in the choice of the basis sets that I have used in the study. In the first place, I included a STO-3G basis set, which is now assumed to be incapable of providing enough flexibility, although its has been very popular in many molecular chemistry applications. Such a set typically produces the effect mentioned in the first point above. To evaluate the role of the polarization wavefunctions, I have selected a basis set of reasonable quality, the 6-21G. The functions in this basis are an adaptation of the standard set in order to obtain a better description of the SiO_2 system. I will compare the effect of using such a basis set with and without basis functions (which correspond to *d* shells). Finally, I have considered a 6-311G* set as the set of best quality in this study. This set has been specially adapted for the study of silica polymorphs and proved to offer very good results both for the description of micro-structure and electronic structure (see work by Chivaleri *et al*). Note that the use of the 6-311G* in systems with a number of atoms similar to that of the models of interface (not related by symmetry) implies considering a large amount of basis functions.

The most important characteristics of the mentioned basis sets are given in table 4.1. In general terms, the sets used are standard and the complete information

	Silicon		
	s	*sp*	*d*
STO-3G	**3**	**3, 3** (0.16)	-
6-21G	**6**	**6, 2** (0.13)	-
6-21G*	**6**	**6, 2** (0.13)	**1, 1** (1.5)
6-311G*	**6**	**6, 2, 1** (0.13)	**1, 1** (1.5)
	Oxygen		
	s	sp	d
STO-3G	**3**	**3** (0.36)	-
6-21G	**6**	**2, 1** (0.37)	-
6-21G*	**6**	**2, 1, 1** (0.28)	**1, 1** (0.40)
6-311G*	**6**	**3, 1, 1** (0.28)	**1, 1** (1.20)
	Hydrogen		
	s	sp	d
STO-3G	**3** (0.16)	-	-
6-21G	**2, 1** (0.16)	-	-
6-21G*	**3, 1** (0.12)	**3, 1** (0.72)	-
6-311G*	**3, 1** (0.12)	**3, 1** (0.73)	-

Table 4.1: **Description of the Gaussian basis sets used.**
The four basis set shown have been tested in this study. The columns named s, sp and d shown how many of this type of shells are contained on each basis set. Each number in bold corresponds to a shell, and indicates how many gaussian functions are contained on its contraction. The numbers in brackets correspond to the value of the exponents of the most external gaussian.

concerning the exponents and coefficients can be found in [30, 84]. The exponents of the outer shells have been optimized for the description of silica, and they are displayed in the table. The number of shells and orbitals for each basis and the total number of shells in the calculation are also listed.

4.3.3 Choice of the models of interface.

As has been already shown in section 4.2.2, an impressive amount of work has been devoted to find models of the SiO_2 micro-structure and, by extension, to the SiO_2 interface. Fortunately, some of these models have allowed us to perform this study without the need of generating new ones.

In the work by Yamasaki *et al.*, already presented in section 4.2.2, it is possible to find the set[3] of interfaces presented in Figure 4.3. As can be seen in the Figure, the interfaces α-S and α-L have an oxide that displays a very similar structure, but a very different thickness. More specifically, the oxide layer in α-S contains 2 atomic layers, while the layer in α-L contains 5 atomic layers. On the other hand, the interfaces α-L and C-L have oxide layers of very similar thickness, while the structures of the oxides are very different, since they are based in the α-quartz, and α-cristobalite phases of silica, respectively.

Although the structure of the oxides in these models looks very much like that of

[3]The information concerning these structures was kindly provided by Dr. C. Kaneta, from Fujitsu Laboratories (Japan).

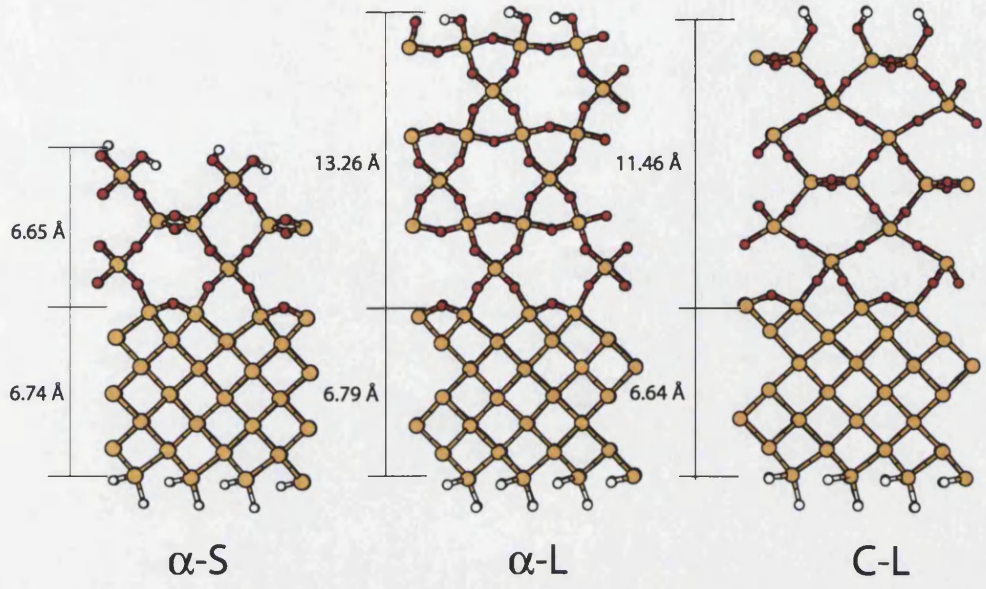


Figure 4.3: **Models of Si/SiO$_2$ interface.**

the corresponding silica crystalline phases, there is in fact certain differences. The most relevant structural features of each model can be seen in Table 4.2. The interfaces have been generated by joining an slab of crystalline silicon (with a (001) surface) and an slab of silicon dioxide, by means of first-principles molecular dynamics, using a plane-waves basis set, at DFT-GGA level, without imposing the constraint of the crystalline symmetry. As a consequence, there is an angular and bond length distribution. It also can be seen that some extra oxygen atoms have been added right at the junction, in order to reproduce correctly the density of oxygen found at the interface by experiments. There are finally, some extra hydrogen atoms used to saturate the dangling bonds in the surfaces of both the oxide and the substrate. For further information concerning the generation of the models and the analysis of their characteristics the reader is referred to the original papers [6, 39].

4.3.4 Setup of the calculations.

All the calculations presented here have a similar setup, that will be described now. All the interfaces have been built by generating an slab based on one of the models mentioned in previous section. The use of periodic boundary conditions allows performing part of the calculation in reciprocal space (in the search for the eigenvalues and eigenfunctions). From our tests, a set of five special κ-points selected by the Monkhorst-Pack method allows a convergence on the total energy up to 10^{-6} Hartree.

Some other important parameters in a calculation done with CRYSTAL98 are tolerances that controls the calculation of infinite Coulomb and exchange series [4].

[4]In CRYSTAL98 these series are truncated following the criteria established by the tolerances. The important terms are evaluated explicitly, while the neglected terms are approximated by means of a bipolar expansion. A detailed description of the method can be found in CRYSTAL98 user's

Model	α-S	α-L	C-L
Total number of ions	35	23	22
Number Si ions	18	23	22
Number O ions	11	21	19
Number H ions	6	6	6
Total Thickness (Å)	17.52	24.26	23.29
Thickness Oxide (Å)	6.65	13.26	11.46
Thickness Substrate (Å)	6.74	6.79	6.64
O-Si-O angle distribution (°)	[105-115]	[100-115]	[95-120]
	115	**105**	**110**
Si-O bond distribution (Å)	[1.62,1.66]	[1.63-1.66]	[1.63-1.68]
	1.66	**1.63**	**1.65**

Table 4.2: **Structural features of the models of Si/SiO$_2$ interfaces.**
The O-Si-O angular distribution and the Si-O bond distribution are described by providing the range (in square brackets) and the value for which the distribution has a maximum (in bold).

To keep the convergence beyond 10^{-6} Hartree it is enough to keep a set of tolerances equal to (6 6 6 6 12) (see CRYSTAL98's manual [30] for a detailed explanation of this notation).

When employing a DFT Hamiltonian, the CRYSTAL98 code makes use of an additional basis set, which is for the evaluation of the exchange-correlation contributions [30]. The authors of the code provide a database of basis sets that can be used for such purpose, and can be found in [85].

4.3.5 Presentation of the results.

The fundamental parameters that describe the electronic structure of the interface, for each different setup of the calculation, are displayed in Table 4.3.5. The value of the valence band offset has been determined following the procedure described in section 3.1. The values in the column "Error" display the difference between the calculated offset and the experimental value of 4.6 eV [10]. The gap has been calculated by finding the difference in energy between the top of the valence band and the bottom of the conduction band in the total density of states. Each one of the calculations has been named with a letter to ease the discussion. Note that calculations A, B, C and D have been performed using the same basis set and the same model of interface, and therefore allow one to study the effect of using different Hamiltonians. Calculations D, E, F and G have been performed using the same model and Hamiltonian, and allow one to study the effect of using different basis sets. Calculations D and H have been performed with the same Hamiltonian and basis set, and the models contain oxides of very similar micro-structure, but different thickness. Finally, calculations H and I have been performed using the same basis sets and Hamiltonian, and the models contain oxide layers of similar thickness, but

guide [30] and the description of the program by Roetti in chapter 8 of [27]. Comments on how this approximation can affect the results can be found in the chapter by Dovesi, in the same book.

	Technique	Basis Set	Interface	Offset	Error
A	HF	6-311G*	α-S	5.60	+1.00
B	DFT-LDA	6-311G*	α-S	2.13	-2.47
C	DFT-GGA	6-311G*	α-S	2.31	-2.23
D	B3-LYP	6-311G*	α-S	3.03	-1.57
E	B3-LYP	6-21G*	α-S	3.12	-1.48
F	B3-LYP	6-21G	α-S	2.40	-2.20
G	B3-LYP	STO-3G	α-S	1.85	-2.75
H	B3-LYP	6-311G*	α-L	3.42	-1.18
I	B3-LYP	6-311G*	C-L	3.96	-0.64

Table 4.3: **Results concerning the calculation of valence band offset.**
The error corresponds to the difference in eV of the predicted offset with respect to the commonly accepted [10] value of 4.6 eV. All values in eV.

different micro-structure.

I will start discussing the results concerning calculations done using different Hamiltonians (Calculations A, B, C and D). As could be expected, the HF technique overestimates both the band offset and the gap width. It is also well known that this is due to the complete neglect of the correlation effects, which causes a wrong positioning not only of the conduction band states but also of the valence band states. On the contrary, both DFT schemes underestimate the offset and the gap width, with the DFT-GGA results being closer to experiments than DFT-LDA ones, but by 0.1 eV. Finally, the hybrid B3LYP scheme gives an intermediate answer, not in exceptional agreement with experiments. These results, however, require further analysis.

Table 4.4 contains information concerning the electronic structure of both silicon and silicon dioxide layers, as given by each calculation. Focusing attention onto the gap widthin the silicon substrate, the calculation in better agreement with experiments would be the DFT-GGA. Such a straight match of the values leads, however, to an erroneous conclusion. Here, the width of the gap has been determined as the difference in energy between the last occupied and the first unoccupied one-electron state (edges of the bands in the projected densities of states). It is well known that any of the techniques employed describes far more accurately the occupied states than the unoccupied states. Therefore most of the error in the calculation of the gap comes from the incorrect positioning of the conduction band in the energy scale. Based in this conclusion, we will focuss in comparing the width of the valence band, which is composed of occupied states only.

Attending to the width of the valence band of silicon, and looking only at the results obtained with the same basis set and model of interface, the calculations done with the B3LYP hamiltonian seem to provide the results in best agreement with the experimental meassurement. Once more, numerical agreement should not be trusted without further analysis, since we cannot be sure that the few layers of

Calculation	Si		SiO_2
	Gap	Valence band	Gap
A	6.92	16.42	17.95
B	0.97	11.57	5.40
C	1.09	12.08	5.38
D	2.26	12.32	8.01
E	2.24	12.80	7.42
F	2.37	13.50	6.55
G	3.87	14.45	-
H	2.15	13.80	8.35
I	2.02	13.85	6.97
Exp.	1.1†	12.5‡	8.9†

Table 4.4: **Parameters of the band structure.**
The values shown correspond to the widths of the band gap for both materials, and the width of the valence band of silicon. They have been meassured directly from the projected densities of states of the interface model, and therefore are equivalent to differences of one-electron energies. All values in eV. Experimental data taken from † [10] and ‡ [86].

HF	DFT-LDA	DFT-GGA	B3LYP
16.42	11.57	12.08	12.32
18.9	11.95	-	13.08

Table 4.5: **Comparison of the electronic structure of the slab model with bulk calculations.**
The width of the valence band the Si substrate is compared with data from calculations of bulk Si, to understand if the ammount of layers of Si employed is enough for reproducing the bulk structure of the material. All calculations done using 6-311G* basis set and model α-S (note that the dimensions of the substrate layer are identical in all models). Data from the bulk calculations taken from the compilation in [28]. All energies in eV.

Si atoms included in the substrate layer are enough as to provide a band structure matching that of bulk silicon. In order to check this point, we have contrasted the data presented above with the compilation included in the work by Muscat *et al. [28]*, concerning the width of the valence band of bulk Si. This is done in Table 4.5. Note that in all cases, with independence of the technique employed, the value of the valence band width is lower than the corresponding bulk value. Assuming that all the calculations are of similar quality in what concerns the quality of the basis set and sampling of the reciprocal space, the conclusion is that addition of more layers of Si is needed to reproduce correctly the bulk structure of the silicon substrate. It is a well known effect that systems with small number of atoms display larger band gaps than the bulk material (quantum confinement). It is worth noting that the calculation corresponding to the models α-L and C-L display larger widths of the valence band than that of the model α-S. Although all three models, have identical dimensions of the substrate layer, the former models are those having similar thickness of the oxide layer, while the later is the one with smallest thickness of the oxide.

In what concerns basis sets (calculations D, E, F and G), it is evident that the basis sets of lowest quality (STO-3G, calculation G) are not able to provide a good description of the electronic structure. Furthermore, the need of polarization functions seems critical, by comparing the results obtained with sets 6-21G and 6-21G* (calculations F and E, respectively). Finally, by comparing results between the two basis sets of better quality (6-21G* and 6-311G*, calculations E and D) one can expect convergence in the results up to 0.1 eV, once certain quality has been reached. Unfortunately, as mentioned, making the outer basis functions more diffuse than those employed will cause problems of linear dependence when constructing the Bloch functions.

Centering the attention now on the features of the models, as has already been pointed out, the comparison between calculations D and H outlines the need of using models thick enough as to obtain a correct representation of the bulk states. Unfortunately, we only had models of two different thicknesses, and we could not test if the bigger ones where thick enough. However, by examining the graphs containing the projections of the densities of states (as the one in Figure 3.1, we could determine that the thickness of the region of oxide which contains interface states is the same (~5 Å), independently of the thickness of the oxide layer. The dimensions of the region containing the interface states are in agreement with previous experimental [5] and theoretical [39] work.

The intrinsic interface states appear because of the way ions at the junction are connected. Silicon ions, for instance, bond two silicons from the substrate and two oxygens from the oxide, while in the bulk materials silicon ion bonds either four other silicons (while in the substrate) or four oxygens (while in the silica). This difference in the bonding scheme leads to the interface states that appear in the gap of the oxide, which are generated as charge redistributes in the junction. The presence of these states is detectable only in a region of the oxide corresponding to several atomic layers. This evidences that silicon dioxide is more ionic than pure silicon. The electric field generated during the charge redistribution causes the electron states in the region of the oxide near the interface to shift, and justifies that interface states have energies inside the gap of silicon dioxide. The electrostatic interaction does not apparently affect the states associated to the silicon substrate, i.e. the crystalline field does not propagate significantly inside the semiconductor.

The fact that the dimensions of the region with interface states remain constant even if the oxide layer becomes thinner may have significant impact for miniaturization. As the oxide layer is reduced, the number of bulk states becomes smaller while the number of interface states remains constant. The interface states, however, are easier to access by the carriers, since they are inside the gap, and very close to the junction (so carriers can tunnel). This, effectively, should be seen as a reduction of the energy barriers and an increase of the leaking current.

Comparison between calculations H and I, shows that also the micro-structure of the oxide influences the band alignment. It is encouraging that the best result is

obtained for the model in which the oxide has cristobalite-like structure, since this is the polymorph whose density (2.33 g/cm^3) is closer to that of amorphous silica (2.21g/cm^3)[87]. This suggests that the models based in a cristobalite-like structure of the oxide may actually simulate successfully the properties of the amorphous oxide. Reading the result in a different way, it is not possible to simulate the electronic structure of the interfaces with an oxide based in any symmetry. In my opinion this result should be handled carefully and further tests are needed, to find out if the relationship between the density of the oxide and the band alignment is consistent.

We finally compare our results with those obtained by other authors. Tang *et al.* obtained a valence band offset of 2.56 eV, using a crystalline model of interface, and a DFT-LDA technique [8]. This is a similar result to ours (2.13 eV), and we notice that the model employed by Tang and co-workers has an oxide thickness of 16 Å, while the oxide layer in our calculation the model had a thickness of only 6.65 Å. We have found almost perfect agreement with the results obtained by Yamasaki and co-workers [39], which used the same models as us, a plane wave DFT-GGA approach, and a similar procedure for studying the evolution of the electronic structure along the interface. Also Tuttle has employed a DFT-LDA scheme, and obtained a value of 3.00 eV for the valence band offset of a model with an oxide structure based in α-quartz [82]. Furthermore, the same author presents a calculation of the band offset based in the ionization potentials (using calculations of surface models), which gives a value of 4.0-4.3 eV, very close to the experimental one. This results are certainly encouraging, although it is unclear to me how to extrapolate the methodology employed by Tuttle to the study of systems where the oxide layer displays an amorphous structure. On the other hand. the calculation based in the ionization potentials does not allow to evaluate the dependence of the offset on features of the junction, like the roughness or the presence of point defects. Finally, the results by Demkov and Sankey shows apparently a very good agreement with experimental data (their calculated offset is 4.65 eV). The authors, however, have employed a local basis set, but they have not studied the dependence of the offset on the quality of the basis set (which in their case is minimal). As we have shown here, this is a critical issue.

It is interesting to note that in most of the cases mentioned, the authors still make use of band-alignment techniques which require defining and locating a reference level, even if the studies already employ interface models.

4.4 Discussion of the results.

4.4.1 Calculation of the offset: capabilities and limitations. Possible ways of improvement.

As has been pointed out already, the most useful aspect of the technique employed here is that it does not rely on any particular energy level as reference, in contrast with most of the other band alignment methods.

On the other hand, it is obvious from the results that obtaining a correct de-

scription of the electronic structure is also not an easy task. In view of the results presented here, the B3LYP technique is probably the best choice of the Hamiltonian. Basis sets including polarization functions should be used.

Our models should be consider as a lower bound. The results depend on the thickness of the oxide layer. Future models should certainly improve in this direction. As we have already mentioned, thicker layers of silicon substrate are needed to reproduce correctly the band structure of this region. The case of the oxide is different, since the thickness that we have employed is very similar to the real thickness of true MOS structures which can now be produced.

Another issue that we have not been able to address here is the structural relaxation of the models to study their dependence on the Hamiltonians and basis sets. Unfortunately, CRYSTAL98 does not allow one to calculate the gradients over the ions, and therefore performing such relaxation is impossible. On the other hand, exploring the energy surface of such a complex system is a task that can only be faced by means of a molecular dynamics technique, in the ways as the authors of the models proceeded.

The impossibility to obtain a correct description of the unoccupied states prevents us from being able to apply the technique to the study of the conduction band offset, which is fundamental for understanding of the electron flow. A possible way to explore, in this direction, is the use of embedding techniques that allow modeling a region of the system at quantum mechanical level, while reproducing the long-range interaction at a simpler level. A good example of this techniques is the scheme developed by A. Shluger, P. Sushko and co-workers [35, 88]. Although the simulation of a complex interface is not yet straightforward in this embedding scheme, its study is certainly possible and affordable, and may become useful in the study of the band alignment problem in the Si/SiO_2 interface. The use of embedded cluster models opens the possibility of considering amorphous lattices generated by inexpensive classical molecular dynamics or reverse quantum Monte Carlo simulations, which are able to provide structures in good agreement with experiments. They also allow one to use more sophisticated quantum chemical methods that can provide a better description of unoccupied states.

Turning back to the method employed here, and to possible applications, I think that the key factor is the cost of the whole process. An evaluation of the offset following this procedure requires: a) to construct a model of the interface following a reasonable procedure; b) to perform a first-principles calculation of the electronic structure of the model, ensuring as much accuracy as possible and c) to evaluate the density of states of the system and generate its projections. From all these steps, the third one takes an infinitesimal part of the resources and the computer time. Some of the other band alignment techniques, on the other hand, require only a knowledge of the band structure of the band materials. In my opinion, such techniques have been useful up to now, since the MOS structures contained dielectric layers of big thickness, where the alignment was easy to deduce with simplified models. The

method employed here, instead, is useful to understand if particular features of the micro-structure, like its roughness, the density of the materials, their stoichiometry or the presence of defects, affect the band alignment.

Once the capabilities and limitations of the method have been discussed, it is worth considering that not all the sources of error may be due to the method itself, but to oversimplifications of the model. The most important of these simplifications will be discussed next.

4.4.2 Disagreement with the experimental results. Possible explanations.

As can be seem from all the results presented, there is a disagreement between the results obtained and the commonly accepted value of 4.6 eV for the valence band offset. In the best of the cases (B3-LYP calculation using a 6-311G* basis set and a model based on the cristobalite phase of the oxide), the offset is underestimated by 0.64 eV. There are several factors that can justify this difference.

Often the systems and mechanisms considered by theoretical studies are much simplified from those systems handled in experiments. Theoretical systems allow an accurate control of the conditions of the system, which is not always possible to achieve in laboratory. On the other hand, the approximations implemented in the techniques and the limitations of computational capacity often oversimplify the systems described in calculations.

As to what concerns the experimental measurement of the valence band offset, several methods have been employed, from high-resolution X-ray photo-electron spectroscopy [89, 80, 90] to electrical measurements [91, 92]. The results actually span a considerably wide range of energies. In the paper by Alay *et al.* [80] the authors remark in the introduction that electron current measurements predicted an offset of less than 1.0 eV for oxides thinner than 3.0 nm, while internal photo-emission experiments give a constant offset of 3.0 eV for systems with oxides thicker than 4.0 nm. Some other measurements of tunneling processes establish the offset at a value lower than 1.8 eV for oxides thinner than 3.1 nm, and close to 3.0 eV for oxides thicker than 3.6 nm. A scanning tunneling microscope-based technique predicts a constant barrier of 2.75 eV if the oxide has a thickness between 1.8 and 4.5 nm. Alay's group themselves have determined the offset to be 4.36 eV (with an accuracy of ± 0.1 eV), by means of X-ray photo-electron spectroscopy. A similar technique is used by Keister *et al.* [89], who have predicted an average offset of 4.44 (± 0.06) eV. The electrical measurements, however, predict much lower barriers, between 3.1 eV [91] and 3.5 eV [92].

Such a dispersion of data, especially amongst the results obtained using different techniques, suggests that either that the experimental conditions are very different and greatly affect the results, or that what is being measured is actually not the same thing. The following list highlights the features of the main techniques listed above:

- **X-ray photo-emission spectroscopy.** This setup allows direct scanning of the valence band of the system. The technique has a high spatial resolution, but it is constrained to explore the surface or the very first layers of the sample. This requires the interfaces to be grown under very controlled conditions. The uncertainty of the technique is around 0.05-0.07 eV. It is very important to note that after the photo-ionization process the system is left in an excited state. The photoelectrons generated in the substrate are usually replaced by carriers coming from the substrate holder, but in the oxide the photoelectrons are also replaced by electrons tunneling from the substrate at a much smaller rate. As a consequence, the oxide charges positively, and the energy and position of the band are bent. This effect is by no means accounted for in standard electronic structure calculations, and therefore the ground-state valence band does not correspond exactly to the detected one.

- **Scanning tunneling microscope-based techniques.** The main difficulty when using STM devices in the SiO_2/Si system is the insulator character of the oxide. This situation can be modified if the sample is irradiated with an electron beam. The beam excites electrons from the insulator valence band into the conduction band, making the system a conductor and therefore allowing STM probing. The I-V (intensity vs. voltage) and $dI/dV - V$ plots allow one to detect the gap width of the materials, and given the spatial resolution of the setup, the change in the width of the gap along the interface can be scanned. In the measurements presented in [93], the authors observe a too narrow oxide gap, which they attribute to an increase in the density of states due to the irradiation process.

- **Transport measurements.** Based on the evaluation of the capacitance and intensity as a function of the voltage. These methods are space averaging over the direction perpendicular to the interface. Because of this, the methods are especially affected by contact performance, unintentional doping or leakage currents.

Given that the main trend in the microelectronics industry has been to miniaturize the devices, the measurements obtained by photo-emission spectroscopy on thin films have been considered as the landmarks for the development of new devices, and the valence band offsets considered to be correct are between 4.4 and 4.6 eV. As already mentioned, this technique allows spatial resolution, essential in order to study the offset at the interface and to reduce the number of undesired effects.

4.4.3 The role of defects at the interface.

As can easily be seen, all the models of interface considered here are free of defects. However, the role of even simple defects like vacancies or atoms with different coordination (for instance, two or three coordinated silicon) is important and known to

severely affect performance of the MOSFET based devices by contributing to carrier tunneling or breakdown processes. In practice, the presence of such defects is unavoidable, or even allowed on purpose. A clear example of this is the use of Hydrogen for saturating the dangling bonds of the Si substrate surface previous to oxidation, in order to allow better control over the oxide growth process [94, 95].

The action of some of these defects over the barriers at the interface has been addressed experimentally. In the work by Perfetti *et al* [96], it is shown how the introduction of interlayers of hydrogen or cesium causes changes in both directions (depending on the dopant), and as big as 0.5 eV, in the values of the band lineup. Such changes are due to the fact that the dopant species bonds atoms at the interface, modifying the quantity and magnitude of the existing dipoles. Some other defects which cause the electrons to localize, like vacancies, are also known to affect the dipole distribution at the interface. The magnitude of their action is, however, much more modest, of the order of 60 meV [92].

In the work by Keister *et al*, the authors outline that the difference in the valence band offset between the interfaces grown on the (100) and (111) surfaces of crystalline silicon may be due to the action of defects at the junction.

It is possible, in principle, to use the method in this chapter for studying the effect of the presence of specific point defects in the electronic structure across the interface. However, I have not addressed the subject here, since it would have introduced more parameters, complicating considerably the comparative study. Some successful attempt of determining the effect of defects in the offset by means of supercell calculations has been done by Kageshima and Shiraishi, who found that the presence of H and O-H increases the value of the valence band offset in 0.1 and 1.3 eV respectively [97]. Certainly such numbers partially justify the difference between experimental measurements and the results presented here.

4.5 Conclusions.

In this chapter I have explored the possibility of studying the band alignment at an interface through an analysis of the density of states of a first-principles calculation. The method is applied to calculations done on periodic models of the interface, where local basis sets have been used.

The approach contrasts with other band alignment techniques, where only information concerning bulk materials is employed. Current computational resources allow already calculations of models of interfaces, and therefore open the possibility to apply methods like the one shown here. This avoids the need to define an absolute reference level for studying the alignment.

The case study has been the Si/SiO_2 interface, given its technological relevance. I have studied in particular the effect that different Hamiltonians, basis sets and characteristics of the models have on the prediction of the band offset. In what concerns the Hamiltonian choice, from all those compared (HF, DFT-LDA, DFT-

GGA and B3LYP), B3LYP gives best agreement with experimental measurements. It is necessary to consider basis sets with polarization functions that allow enough flexibility to the wavefunction.

The models of interface tested (taken from the work by Yamasaki *et al.*) contains layers of Si substrate which are too thin as to describe the band structure of the bulk material. It was possible to see that the interface states remain confined in the oxide region which is closest to the junction. The thickness of this region is around 5Å wide, and is independent of the total thickness of the oxide. Since total thickness of the oxide region in some of the models is very similar to that of some of the real MOS structures, this result could be relevant for technologicl applications. Electrons trapped in the region of the interface states would be able to tunnel easily. Therefore, the thickness of the oxide layer should be bigger than that of the region containing interface states.

Part of the disagreement of the calculated offset is certainly due to oversimplifications of the model, like the absence of defects in our models of interface. Certainly, further work is required to understand the role of defects in the band alignment, but this is not possible without using models of the interface. The most interesting characteristic of the technique used here is actually the possibility to understand how the band alignment depends on particular features of the junction.

Chapter 5

Defects in hafnia and zirconia

Aim. In this chapter some basic point defects inside zirconium and hafnium dioxides are characterized. The formation energies, electron affinities and ionization potentials of the anion and cation vacancies, substitutional Zr in HfO_2 and complex pairs formed by this substitutional defect and a vacancy. The capability of these defects to act as charge traps is discussed.
Results. The ionization potentials and electron affinities have been evaluated as ground-state properties by defining them in terms of differences of total energies of different charge states. A correction for the underestimation of the band gap has been considered. The relaxation of the structures as defects are formed and/or charged is small, due to the dense nature of the materials, and suggest that they display a considerable ionic character. The oxygen vacancies appear to be able to trap electrons, and neutral oxygen vacancies are also able to trap holes. The presence of hafnium impurities inside zirconia does not affect significantly the properties of the vacancies. Finally, the energy involved in exchanging ions between the oxides turns to be ten times smaller than the energy required to exchange ions in the metals. This partially justifies the difficulty to isolate zirconium from hafnium.
Techniques employed. First Principles Density-Functional Theory on its Generalized-Gradient Approximation (DFT-GGA).
Basis sets. Plane waves.
Models. Periodic models of monoclinic Zirconium dioxide and Hafnium dioxide, containing around 96 atoms.
Computer Facilities. Self-consistent calculations of the electronic structure done using the Cray-T3E supercomputer of the CSAR service at Manchester (UK).

5.1 Introduction.

Materials with a high dielectric constant (κ) are being considered as possible substitutes for Silicon oxide as a gate dielectric in MOS devices. This would allow one to increase the physical thickness of the gate to avoid the risk of carriers tunneling through the gate, without degrading the performance of the device. Prior to the employment of any of the high-κ materials in MOS technology, several tests have to be performed, like, for instance, the mechanical and thermal stability of the material over the Silicon substrate, and the density and ability of defects to act as charge traps. When it occurs, charge trapping by defects may contribute to some of the dielectric breakdown processes, which eventually make the device useless.

Zirconia and Hafnia have turned out to be strong candidates as gate dielectrics, fulfilling many of the mentioned prerequisites. Although most of the properties of

these materials have been already studied because of their wide scope of applications ranging from nuclear rods to jewelry, the properties of point defects have not been studied in detail.

Electronic structure calculations can be used to make an exhaustive characterization of point defects. The characteristic of these materials, with a strong ionic character, allow successful application of Density Functional Theory methods. In this chapter I will show how the energies involved in the formation of and charging of the defects can be evaluated accurately as a ground state property of the system, despite the intrinsic limitation of DFT to describe the unoccupied electronic states.

The chapter is organized as follows. In first place a bibliographic review on the previous research on the field is presented. Then, since all the setup of the calculations and models of defects is similar, it will all be introduced at once. Finally the results concerning each type of defect will be presented and discussed.

I will like to outline here that this research was performed in collaboration with A. S. Foster, from the Helsinki University of Technology and V. B. Sulimov, at that time working at University College of London. The results of the study have also been published in the papers [98, 99, 100].

5.2 Bibliography review.

5.2.1 Two strong candidates: Zirconia and Hafnia.

Two very attractive candidates, which are the object of very active research at present, are ZrO_2 and HfO_2. The permittivities for these materials are 22 and 21, respectively, allowing the use of a considerably thick layer of this oxide with a capacitance equivalent to very thin layers of silicon dioxide. Devices with a dielectric layer around 36 Å thick of any of these materials will have the same capacitance as devices with a layer of SiO_2 just 5 Å thick.

The band gaps of zirconia and hafnia are 5.8 and 6.0 eV, respectively, and the conduction band offsets are around 1.4 eV for ZrO_2, which is above the limit value of 1.0 eV required to avoid leakage current [12].

In terms of mechanical stability, usual processing temperatures actually rise up to 900°, and then it has been proven that ZrO_2 has a tendency to react with the Si at the interface, forming a significant SiO_2 and SiZr layers [11]. Because of the smaller permittivity of silicon dioxide, the effectiveness of the dielectric layer is reduced. There are some attempts to control the wideness of the SiO_2 layer by controlling the processing conditions [101]. The problem of incorporation of zirconium or hafnium to the silicon substrate and formation of the silicon oxide interfacial layer has also received some attention from theoretical studies [102] recently. Theoretical studies concerning different aspects of the growth of Si/ZrO_2 films has been presented by Brodskii *et al.* [103].

Although the formation of a silicon oxide layer is certainly not desirable, the formation of a silicate (Zr-Si-O or Hf-Si-O compounds) is not so inconvenient. The

silicate is formed from ZrO_4 and SiO_2 tetrahedral units. The band gaps and offsets seem to be close to the ones of the pure ZrO_2 high-k material oxide [12] and it seems that these materials present excellent behaviour in preventing leakage currents [104, 105]. Both the stability and electrical properties of these materials largely depends on their composition [105, 102]. The biggest drawback in the use of silicates is the low dielectric constant (around 11 for Zr compounds and 13 for Hf compounds). This may limit considerably the applicability of these materials as gate dielectrics in a few generations, although some research suggest that enhanced dielectric constants can be obtained by using SiO_2 and silicates alloys [106]. In any case, the dependence on the composition considerably complicates the systematic characterization of silicates and alloys. In this regard ideal oxide systems act as good paradigms for looking at more general properties.

5.2.2 Previous studies of the electronic structure of ZrO_2 and HfO_2.

One of the main problems of using ZrO_2 or HfO_2 is that they are fast ion conductors. In particular, oxygen has a very high diffusivity through these materials. In fact, this property has made these materials very attractive for application in catalysis and gas sensors. Since a detailed knowledge of the electronic structure is desirable to understand the mechanisms of diffusion and chemical bonding, several studies have been devoted to find it, even before the microelectronic industry found these materials interesting. One of the first studies was presented by Orlando et al. [107]. They used a Hartree-Fock level of theory, effective core pseudo-potentials and a valence band LCAO basis set. The cubic and tetragonal phases of zirconium were studied extensively. Experimental geometries were used, and some partial geometry optimisation was done for the tetragonal phase. As should be expected because of the use of a Hartree-Fock scheme, the band gaps appear clearly overestimated by a factor of more than 2. Stefanovich *et al.* performed, also using Hartree-Fock-based techniques (first principles and semi-empirical INDO), an extension towards all three phases [108]. This work shows that good accuracy for structural data and related magnitudes can be at *ab initio* Hartree-Fock level within 1% deviation from experimental data.

DFT-LDA calculations of these systems have been presented by French et al [109]. This last work, however, makes use of experimental equilibrium structures at room temperature. The band gaps are in this case underestimated by a factor of around 3/4, obviously because of the use of the LDA approach. The overall work presented by the authors is, however, very interesting because it also provides experimental data concerning the electronic structure. Finnis and co-workers [110] have also studied the relative stability of all three phases of zirconia, showing how a self-consistent tight-binding model developed by this group allowed a correct description of the energetics, close in accuracy to DFT-LDA results.

The most accurate calculation of the band structure, so far, has been provided

by means of the GW approximation for electron's self energy[1]. The authors show how DFT-LDA level calculations also provide correct structural information, and describe properly the total energy ordering of the different phases[111].

In what concerns to the structural properties, the amount of information available is actually very big (for reviews on both materials, see the articles by Kisi [112] and Wang [113]). Zirconia and hafnia are very well known materials in the field of ceramics. Nuclear industry takes advantage of hafnia because of its high neutron adsorption coefficient [113]. Certain instability problems presented by zirconium, which undergoes tetragonal to monoclinic transformation below 1000°, made it unsuitable for high-temperature applications, directing the attention of the research towards hafnium as an alternative material. The behaviour of both zirconia and hafnia at high-pressure phases has also turned out to be interesting because of their high bulk moduli, which makes them interesting as very hard and refractory materials[114]. The difficult experimental set up for high-pressure characterization has made theoretical models play an important role [115, 116].

One of the main fields of research, of interest both for ceramics and microelectronics industry, is the presence of defects in the ZrO_2 and HfO_2 structures. Addition of impurities like Ca, Mg or Y improves considerably the thermodynamically properties like strength and toughness. The stabilization effect of such ions in the cubic phase of ZrO_2 has been studied by Stefanovich et al [108] by means of the HF-INDO technique, and experimental measurements using ESR are shown by Ben-Michael *et al.* [117].

There are some other defects, which are more relevant for the performance of these oxides as gate dielectrics. This is the case of interstitial oxygen, and the vacancies of both anion and cation species. The presence of interstitial oxygen is obviously related to the process of diffusion of this species, affecting silicon oxide film growth. The vacancies, on the other hand, are important defects because they may show the ability to act as charge traps. In both cases, understanding the energetics and electronic structure of both types of systems is fundamental in order to ensure reliability of the dielectric. We have performed detailed DFT-GGA electronic structure calculations using a plane wave basis set to study this. Results are presented in [98, 99], and will also be presented and discussed in detail next. The other only work, to our knowledge, which has studied the oxygen vacancies is that by Králik and co-workers [111].

5.3 Description of the Structure of ZrO_2 and HfO_2.

In order to construct the models for our study we have considered the following data. Both Zirconia and Hafnia have several crystalline phases. At the pressure and

[1]The GW approach allows the calculation of excited state properties, and therefore the band gap, or difference between occupied an unoccupied levels. It was first systematically formulated Hedin, and first implemented in a computational scheme by Hybertsen and Louie. We will not give a description of the method here, since it has not been used in our research. Further references can be found in the work by Kralik *et al* .

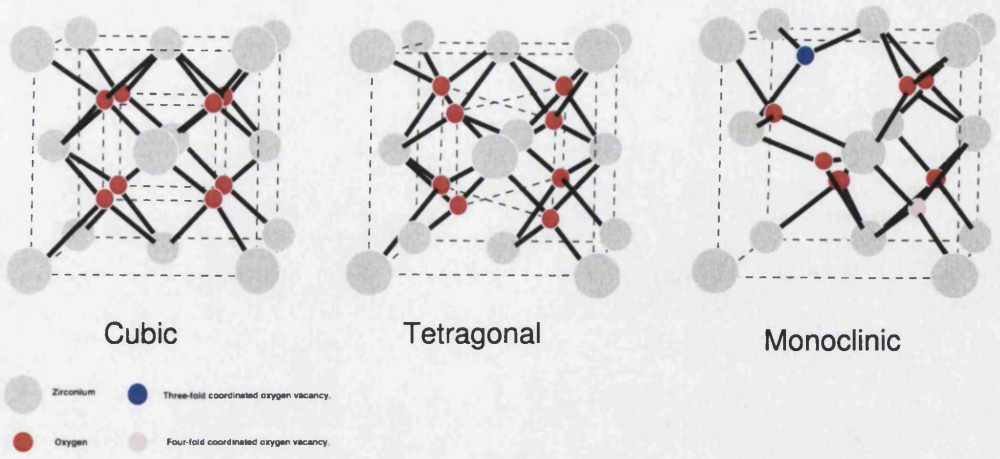


Figure 5.1: **Atomic arrangements of cubic, tetragonal and monoclinic ZrO_2 or HfO_2.**

temperature conditions in the manufacture and operation of MOS devices, ZrO_2 and HfO_2 show three different phases: monoclinic, tetragonal and cubic. Both monoclinic and tetragonal structures can be obtained by distortions of the cubic structure. Note that the monoclinic unit cell has the largest number of degrees of freedom (it is defined by three cell vectors and the angle β). The unit cells of all three are reproduced in figure 5.1. In table 5.1 some of the experimental values [118, 119, 113] for the relevant crystallographic parameters are shown. A more exhaustive set of crystallographic data for different temperatures can be found in the review by Wang [113]. There is a strong similarity between both materials, but is interesting to note that the volumes of the Hafnia unit cells are smaller in all three cases. This result can be predicted by theory, since the ionic radius of Zr (0.083nm) is larger than the one of Hf (0.084nm)

.

The transformations, at atmospheric pressure, from one stable phase to the next as temperature is increased are as follows [2]:

For Zirconia:

$$\text{Monoclinic} \overset{1720^\circ}{\Longrightarrow} \text{Tetragonal} \overset{2600^\circ}{\Longrightarrow} \text{Cubic}$$

For Hafnia:

$$\text{Monoclinic} \overset{2000^\circ}{\Longrightarrow} \text{Tetragonal} \overset{2870^\circ}{\Longrightarrow} \text{Cubic}$$

Although both materials follow the same transformation sequence, each phase transition always takes place at higher temperature in Hafnia than in Zirconia. This can be an interesting advantage to avoid phase transitions during manufacturing. In general, the qualitative description of both materials is identical, since Hf and Zr have the same electronic outer shell configuration. Because of the higher density of Hafnia,

[2] Data for these transformation temperatures have also been taken from the review by Wang [113]. As the author remarks, there is some disagreement in these numbers, which are given for guidance purposes only.

	Monoclinic Phase (space group $P2_1/c$)		Tetragonal Phase (space group $P4_2/nmc$)		Cubic Phase (space group $P4_{3m3}$)	
	ZrO_2	HfO_2	ZrO_2	HfO_2	ZrO_2	HfO_2
a	5.145	5.292	5.094	5.140	5.124	5.110
b	5.208	5.175	-	-	-	-
c	5.311	5.292	5.117	5.250	-	-
β	90.23	99.22	-	-	-	-

Table 5.1: **Structural parameters for ZrO_2 and HfO_2.**
Data have been extracted from the review [113].

however, differences in the energies of formation of defects, ionization potentials and electron affinities are expected. Also, as pointed out by Lowther *et al* [115], electron correlation seems to play an important role in explaining the differences between both materials.

5.3.1 The monoclinic phase.

Since the monoclinic phase is the most stable in the range of temperatures at which processing of MOS devices takes place, we have centred our attention on this system, and all subsequent models of defects are based on a supercell of this polymorph. This phase has been experimentally characterized in several works, which are summarized in the review by Wang [113]. Since Zirconia and Hafnia monoclinic phases are almost identical, the following description is applicable to both:

- Zr(Hf) has a formal charge close to +4, and it is seven-fold coordinated. The range of bond lengths goes from 2.057 Å to 2.285 Å in ZrO_2 and 2.031 Å to 2.254 Å in HfO_2.
- There are two types of Oxygen sites with three-fold (O_I) and four-fold (O_{II}) coordination, both with formal charges close to -2.
- The Zr(Hf) ions are arranged in layers parallel to the (100) planes, separated by Oxygen ions on either side.

Interatomic distances are shorter in Hafnia than in Zirconia. As an example, the average Zr-O_I distance is 2.090 Å and the average Zr-O_{II} one is 2.211 Å, while the corresponding distances in Hafnia are 2.086 Å and 2.197 Å respectively. Because of this disparity in bonding distances, the group of O_{II} ions resemble a square array more closely in HfO_2 than in ZrO_2.

5.4 Technical details of the calculation.

Since the study of the structure and electrical levels has been done at DFT-GGA level, in this section we provide all the details of the calculations. All calculations have been performed using the VASP 4.4.4. code [32, 120, 121].

5.4.1 Set up of plane wave DFT calculations.

We have used the spin-polarized Density Functional Theory Hamiltonian and the Generalized Gradient Approximation of Perdew and Wang, usually known as GGA-II. We have made use of Vanderbilt [53, 52] ultrasoft pseudo-potentials to simulate the cores of the ions. The pseudopotentials were generated considering the electron configurations $[Kr]4d^3 5s^1$ for Zirconium atoms, $[Xe4f^{14}]5d^3 6s^1$ for Hafnium and $[1s^2]2s^2 2p^4$ for the Oxygen atoms (the core electron configurations are shown inside the brackets).

The two fundamental parameters to adjust in a DFT plane-wave calculation are the number of plane waves of the basis set (through the cut-off energy) and the number of k-points considered in the sampling of the reciprocal space. The results of our calibration for Zirconia are shown in table 5.4.1.

We explored cut-off ranges between 200 and 600 eV in the case of Zirconia, and 200 and 700 eV in the case of Hafnia, and have found that a cut-off energy of 500 eV is sufficient to achieve convergence in the total energy of the order of 10 meV in any of the systems studied here.

The k-point sampling of the cell was done by means of the Monkhorst-Pack method. We tested the dependence of the total energy as the sampling in reciprocal space was varied from 1 to 60 k-points in the case of Zirconia, and 1 to 80 k-points in the case of Hafnia. Grids of 20 k-points in the former case and 10 k-points in the later provided accuracy within 10 meV in the total energy of the unit cell. The modelling of defects has required the use of supercells (which will be described in section 5.4.7), which are constructed by extending the unit cell of the monoclinic phases by two along all three dimensions. The symmetry of this superstructure causes some points in the reciprocal space to become equivalent, and therefore the amount of points to be sampled can be reduced significantly. A $2 \times 2 \times 2$ grid corresponds to a set of two special [3] k-points in the case of the 96 atom perfect supercell, and provides accuracy up to 1 meV in the total energy of the system, as it can be seen in table 5.4.1. The creation of a point defect inside the supercell lowers the symmetry of the system, therefore making necessary to employ a set of four special k-points. This set of k-points was used in all the calculations of defects.

5.4.2 Criteria for relaxation of structures.

Structural relaxations were done for the different defects in different charge states. For each type of defect, the neutral defect was generated by modifying the relaxed structure of the perfect lattice. The starting geometry for the relaxation of charged defects is that corresponding to the relaxed structure of the defect in another charge state. The discussion of the relaxation process will be addressed in terms of the

[3] When a supercell containing several unit cells is used, the Brillouin zone in reciprocal space is smaller than that corresponding to the unit cell. Because of the symmetry, each of the points contained in the reciprocal space cell of the supercell is equivalent to several k-points in the normal Brillouin zone. These points are known as special k-points.

Cut-off energy	Total energy	Number of k-points	Total energy
450	-916.90321	1	-916.698194
500	-916.87354	2	-916.968404
550	-916.87940	10	-916.968174

Table 5.2: **Convergence tests for the total energy of the 96 atom Zirconia supercell, as a function of the energy cut-off and number of k-points.**
The Table shows the total energy of the supercell for different values of these parameters. The Energy cut-off determines directly the number of plane waves that are included in the basis set. The choice of cut-off energy of 500 eV gives convergence in the total energy of the cell in the order of 10 meV.The k-point mesh was generated using the Monkhorst-Pack method (see text). The choice of a mesh of size 2x2x2 (containing 2 k-points) gives convergence up to 1 meV in the total energy.

change in energy and in the position of the ions between the unrelaxed and the relaxed structures. In all cases, the criterion to stop the relaxation was converging the forces over atoms to a value smaller than 0.05 eV/Å. A combination of Conjugate Gradient and quasi-Newton minimization algorithms was used in the relaxations.

5.4.3 Accuracy of the calculations.

To test the accuracy achieved with the set up just described, we have compared with experimental data the results obtained for the perfect structures of the three phases (cubic, tetragonal and monoclinic) of the oxides (ZrO_2 and HfO_2) and the hcp phases of the metals (zirconium and hafnium).

Typically, DFT-GGA calculations do not agree completely with the experimental structural data. DFT-GGA calculations predict longer bond-lengths, so the volume of the cell at equilibrium turns out to be larger than the experimental one. Because of this disagreement the experimental geometries cannot be used as the most stable in the calculations, as they appear strained. We therefore needed to first find the most stable geometries of the unit cells, as predicted by DFT.

We have performed full cell relaxations (i.e. position of the ions inside the cell, and cell volume and shape) of all three phase of both Zirconia and Hafnia. The results are shown in Table 5.3.

In the cases of the metallic systems, very big sets of k-points are required to obtain correct description of the delocalised states. In the cases shown here, sets of 250 k-points and cut-off energies of 250 eV were required to achieve convergence of 10 meV in the total energy of the cell. Due to the elevated cost of the calculations, we opted for a less time-consuming methodology than the straight optimisation of the structures. First. the positions of the ions of the basic unit cells of were relaxed for a series of fixed volumes. Then, the total energies obtained were fitted with a Murnaghan equation of state [122], and the equilibrium volume and total energy were found. A comparison of the structural parameters between the DFT-GGA

	ZrO_2		HfO_2	
	Calculated	Experimental	Calculated	Experimental
		Cubic		
Volume(Å^3)	32.97	32.97	32.49	32.77
a(Å)	5.090	5.090	5.07	5.08
		Tetragonal		
Volume(Å^3)	34.55	34.07	33.12	35.08
a(Å)	3.628	3.571	5.06	5.15
c/a	1.447	1.451	1.024	1.027
δz	0.049	0.057	0.051	-
		Monoclinic		
Volume(Å^3)	36.05	35.22	34.81	34.62
a(Å)	5.192	5.150	5.132	5.119
b/a	1.014	1.012	1.011	1.011
c/a	1.032	1.032	1.034	1.035
$\beta(°)$	99.81	99.23	99.78	99.18
X_X	0.277	0.275	0.277	0.276
X_y	0.044	0.040	0.044	0.040
X_z	0.209	0.208	0.209	0.207
$O1_X$	0.072	0.070	0.070	0.071
$O1_y$	0.338	0.332	0.333	0.332
$O1_z$	0.341	0.345	0.345	0.344
$O2_X$	0.447	0.450	0.448	0.446
$O2_y$	0.758	0.757	0.758	0.755
$O2_z$	0.479	0.479	0.478	0.480

Table 5.3: **Comparison of structural parameters between calculated (DFT-GGA level) and experimental data for the monoclinic phases of ZrO_2 and HfO_2.**
δz is the shift in the fractional coordinates of oxygen ions in the tetragonal cell with respect to their ideal position in the cubic phase. β is the angle (in degrees) between a and c. The coordinates of the non-equivalent ions are given in fractional units, and X stands for the metal species (Hf or Zr). Experimental data taken from [113, 123, 118, 119]

predictions and some of the experimental data are shown in Table 5.4.

On view of the results just presented, the general conclusion is that, despite a slight overestimation caused by the longer bond lengths, the results are in good agreement with experimental data. This essentially proves that the set up of the calculations was correctly done, and that further results can be trusted within the limitations of the theory.

5.4.4 The electronic structure of perfect lattices.

Since most of the processes in which we are interested involve the transition of carriers between different one-electron states, we analyse now how the electronic structure of the bulk materials are described in our calculations.

All three phases of both Hafnia and Zirconia qualitatively display the same band structure. Since the calculations have been performed using pseudo-potentials to simulate the electron cores, the discussion concerns the outermost bands only. A

Property	Calculated	Experimental
a(Å)	3.1797	3.1946
c(Å	5.0239	5.0510
$\frac{c}{a}$	1.58	1.58
V_0(Å^3)	21.99	22.32
$B_0(GPa)$	110	110

Table 5.4: **Comparison of calculated and experimental data for the metalic systems.** The data concerns bulk hcp hafnium. Experimental data has been taken from [124] and [125].

graphic representation of the density of states described here is shown in figure 5.2. The plot corresponds to the total density of states of the monoclinic phase of both Hafnia (solid black line) and Zirconia (dashed red line). The main characteristics of the density of states of these materials are:

- One valence band is composed of Oxygen 2s orbitals. This band is located at around -15 eV.
- Another valence band appears at 0 eV, and arises mainly from Oxygen 2p orbitals, with a small contribution from d orbitals from Zr or Hf . The proportion of this contribution is consistent with the fact that the covalent character of the Zr(Hf)-O bonding of these materials is very small and they can be considered ionic insulators.
- A conduction band formed by d states of Hf or Zr.

It is a well known fact that the DFT-GGA method is unable to provide a proper description of the unoccupied states (which for the conduction band of the oxides). As a result of the self-interaction interaction, the unoccupied states are positioned far below the experimental value, and therefore the gap between the valence and the conduction bands appears to be smaller than experimental values. The way this error is treated in our study is discussed next.

In any case, the values that we have obtained are reasonable within the limitations of the method. This can be seen by comparing the gap width of our calculation of monoclinic zirconia (3.41 eV) with the compilation of results obtained with other methods presented by Králik *et al.* [111]. The DFT-LDA technique gives a gap width of monoclinic zirconia equal to 3.12 eV, that is an even smaller value than the DFT-GGA result. This is consistent with the fact that the GGA technique provides a better approximation of the exchange-correlation energies than the LDA scheme. The GW approach, on the other hand, provides a value of the gap of 5.42 eV, much closer to the experimental value. This is, once more, also consistent, since the GW approach constitutes one of the most accurate techniques available.

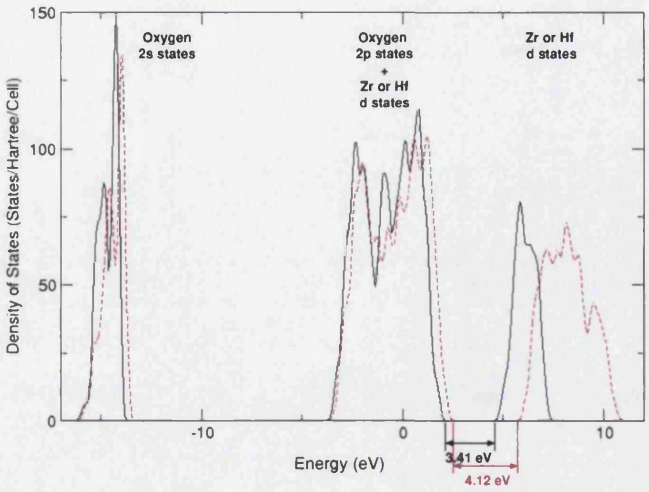


Figure 5.2: **Density of states of monoclinic Zirconia and Hafnia.**
The black solid line corresponds to the density of states of Zirconia, while the red dashed line corresponds to Hafnia. The widths of the respective band gaps, calculated as difference between one-electron energies, are also shown.

Material	Exp. Gap	One-elec. GAP	Tot. E. GAP	κ
ZrO_2	5.4†	3.19	3.41	2.21
HfO_2	5.68‡	3.92	4.12	1.76

Table 5.5: **Calculated and experimental values of the gaps of ZrO_2 and HfO_2.**
The first column contains the experimental values (see text for references), the second one the value of the gap calculated as the difference in one-electron energies, the third the value calculated as the difference in total energies, and the last one is the correction κ calculated with these values. All energies are given in eV. †Data taken from reference [127]. ‡Data taken from reference [128].

5.4.5 Correction to the band gap.

The value of the gap for both zirconia and hafnia, as given by experiments and as calculated by us, are shown in table 5.5. The value of the gap has been calculated as a ground -state property, as described in 3.2.2, and the corresponding κ correction calculated for these values is also shown in Table 5.5.

Earlier we noted that there is no overall agreement with the experimental values of the gaps of these materials. In the case of zirconia, the value for the experimental gap is given as 5.83, 4.2 and 5.4 eV depending on whether it is measured by UPS [109], EELS [126] or electron photo-injection [127] techniques. In the table we display the value obtained using the latter technique, because it is also the closest to that obtained by DFT calculation using perturbation theory at GW level [111] (5.4 eV). The experimental value for hafnia band gap (5.68 eV) has been taken from [128].

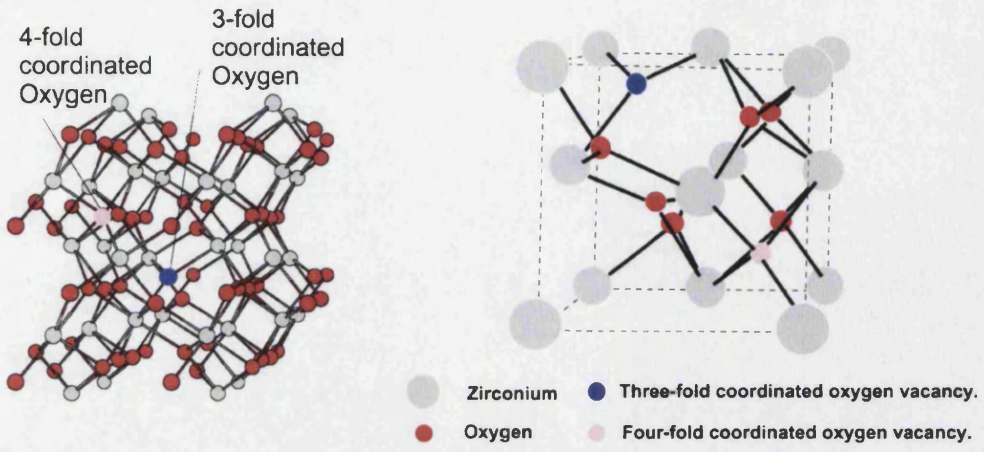


Figure 5.3: **Supercell of ZrO_2 (or HfO_2)**
The cell, containing 96 atoms is shown on the left. It has been constructed by replicating the crystallographic cell of monoclinic Zirconia (Hafnia) shown on the right. The red circles indicate Oxygen ions, the grey are either Zirconium or Hafnium. A 3-fold-coordinated Oxygen is highlighted in blue and a 4-fold coordinated oxygen in pink.

	Oxygen	Zirconium	Hafnia
Total energy (eV)	-1.97	-2.23	-3.36
DFT Config.	$[1s^2]2s^22p^4$	$[Kr]4d^25s^2$	$[Xe4f^{14}]5d^26s^2$

Table 5.6: **Atomic energies.**
Calculated following the procedure described in 3.2.1. They have been employed in the evaluation of the formation energies of defects.

5.4.6 Atomic energies.

Table 5.6 shows the atomic energies employed to calculate the formation energies of the defects. They have been calculated following the procedure described in 3.2.1. The second row lists the electronic configuration employed as reference for the fitting of the pseudopotentials.

5.4.7 The supercell models of defects in Zirconia and Hafnia.

All calculations of defects in Zirconia and Hafnia were done using a 96 atom unit cell as basis. This supercell was generated by extending the 12-atom monoclinic unit cell by two in three orthogonal directions. In this structure, the neighbouring defects (generated by periodic translation of the supercell) are on average 10 Å apart. At this distance the contributions to the total energy coming from interactions between these defects are, in all cases, smaller than 0.1 eV (the interaction between defects has been estimated following the method described in 2.2.1).

In figure 5.3 we show a schematic representation of the supercell, highlighting the sites corresponding to three-fold and four-fold coordinated oxygens.

5.5 Oxygen vacancies.

Oxygen vacancies (i.e. the removal of one Oxygen atom from its site in the crystalline network) are one of the simplest types of point defects that exist in the oxides. In many cases, vacancies have been reported to act as charge traps. Such a behaviour is undesirable if the oxide is being used as a gate dielectric, because the trapped charge may contribute to increase the leakage current and to trigger a breakdown process.

The aim of this study is giving a characterization, as complete as possible, of the oxygen vacancies in both Zirconia and Hafnia. I have divided the analysis in blocks:

- Structural data and energies of formation.
- Analysis of the electronic structure.
- Ionization potentials and electron affinities.

5.5.1 Description of the structures and formation energies.

The monoclinic structure of both Hafnia and Zirconia contains two different sites for Oxygen, characterized by a different coordination (see figure 5.3). We have performed a systematic study of both, and results are presented mostly through tables and pictorial schemes to allow an easier comparison. To differentiate each type of vacancy, I will make use of the notation V_c^q, where c stands for the coordination and q for the charge state.

The vacancies were generated by removing each of the different types of oxygens atoms (three-fold coordinated or four-fold coordinated, see figure 5.3). In all cases we remove a neutral atom, and not an ion in the corresponding charge state in the crystal.

A summary of the results of the relaxation of the structures is shown in Table 5.5.1. The first thing to point out is the similar behaviour of both materials. The very densely packed structure of these oxides only allows a very small relaxation of the neutral vacancies. The displacements in this relaxation process are 0.5%-1.0% of the Zr-O or Hf-O bonding distances. In the case of positively charged systems, displacements due to relaxation are in the order of 5%-10% of the bonding distances. This is caused by the reduction in the screening effect by the electrons trapped in the vacancy level. As these electrons are removed, the first cations surrounding the vacancy experience a greater repulsion between them. In the case of the negatively charged vacancies, the displacements are similar to those of neutral vacancies. The reason is that the extra electron is delocalized throughout the entire supercell.

The formation energies of the neutral vacancies have been calculated assuming the following process takes place:

$$\boxed{\text{Perfect}} \Longrightarrow \boxed{V_4^0} + \boxed{O_{atom}}$$

	ZrO_2		HfO_2	
Vacancy type	Relaxation energy	Range of the displacements	Relaxation energy	Range of the displacements
V_4^0	0.11	0.01-0.02	0.06	0.01-0.02
V_4^+	0.47	~ 0.1	0.61	~ 0.1
V_4^{2+}	0.74	~ 0.1	0.84	0.1-0.15
V_4^-	0.10	< 0.02	0.09	< 0.07
V_3^0	-	-	0.09	0.01-0.02
V_3^+	-	-	0.65	0.1 - 0.17
V_3^{2+}	-	-	0.83	0.1-0.2
V_3^-	-	-	0.04	0.01-0.02

Table 5.7: **Energies (in eV), displacements (in Å) and formation energies involved in the relaxation of the Oxygen vacancies in Zirconia and Hafnia.**

	ZrO_2	HfO_2
V_3^0	8.90	9.48
V_4^0	8.88	9.46

Table 5.8: **Formation energies (in eV), of three- and four-fold coordinated Oxygen vacancies in Zirconia and Hafnia.**

The oxygen atoms are assumed to be far enough as to have no interaction with the vacancy site. The initial system is calculated as a 96 atom supercell of perfect monoclinic oxide, the vacancy as a 95 atom supercell, and the oxygen atom as isolated (described in subsection 3.2.1). Note that the number of atoms and the number of electrons between the initial and the final states are equal. The formation energy of the vacancy can then be defined:

$$E_{for}(V_c^q) = E(V_c^q) + E(O) - E(prf) \tag{5.1}$$

where $E(V_c^q)$,$E(O)$ and $E(prf)$ are the total energies of the vacancy, the oxygen atom and the perfect lattice respectively. The values of the formation energies were obtained using equation 5.1, and are shown in table 5.8.

For each material, the similarity between both values is remarkable despite the difference in coordination. In the case of charged vacancies, we do not consider their direct formation, but rather the processes of charging of the defects. The energies involved in such processes are the ionization potentials and affinities described in subsection 5.5.3. Prior to this calculation, it is necessary to have a clear picture of the band structure of the system and of the perturbation introduced by the vacancies.

5.5.2 The electronic structure.

Neutral oxygen vacancies introduce a double-occupied one-electron level into the band gap of the oxides. With respect to the top of the valence band, this level is

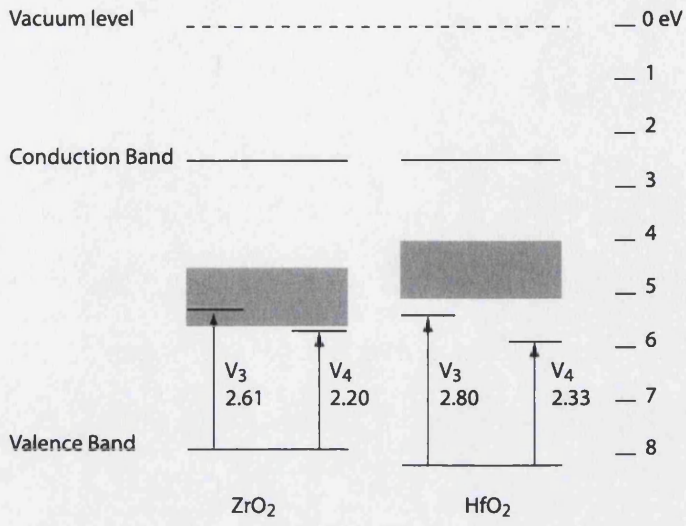


Figure 5.4: **Defect levels of different vacancies in zirconia and hafnia.** The width of the band gap and the position of the bottom of the conduction band with respect to the vacuum level (electron affinity of the material) are those detected experimentally, as given in [127, 128, 12]. The calculated position of the levels is given with respect to the top of the valence band. The estimated position [12] of the silicon gap (1.1 eV, as given in [10])is also shown as grey bands. All energies are in eV.

placed at 2.61 eV and 2.20 eV in the three-fold and four-fold coordinated vacancies in ZrO_2 and 2.80 and 2.33 eV in the three-fold and four-fold coordinated vacancies in HfO_2. These results are displayed schematically in Figure 5.4. Since it is well known that the unoccupied states are not correctly described by our approach, we have displayed in this Figure the position of the conduction band with respect to the valence band as provided by experiments (i.e. we used the esperimental band gap width). The experimental electron affinity of the material allows also to estimate the position of the vacuum level with respect to the bottom of the conduction band. It should be kept in mind, however, that this is an approximated estimation, since the electron affinity depends in the dipoles formated at the surface of the material. We have also included the position of the silicon band gap, as estimated by Robertson [12]. Note how the defect level corresponding to the three-fold coordinated oxygen vacancy in zirconia falls into the same range of energies as the Si band gap. Also some other levels (V_4 in ZrO_2 and V_3 in HfO_2) are very close to the edge of the valence band of Si, suggesting that tunneling of electrons from the valence band of Si to these defects levels is very likely to happen.

The reason for the difference in the position of the defect levels between these materials is that hafnia has shorter cell vectors than zirconia has. The potential well in which the electrons in the vacancy level are trapped is narrower in hafnia, and therefore the energy of these electrons lies slightly above the energy of the corresponding levels in zirconia.

We can compare some of our results with the only other work which has studied the vacancies. Králik and co-workers have performed accurate GW calculations,

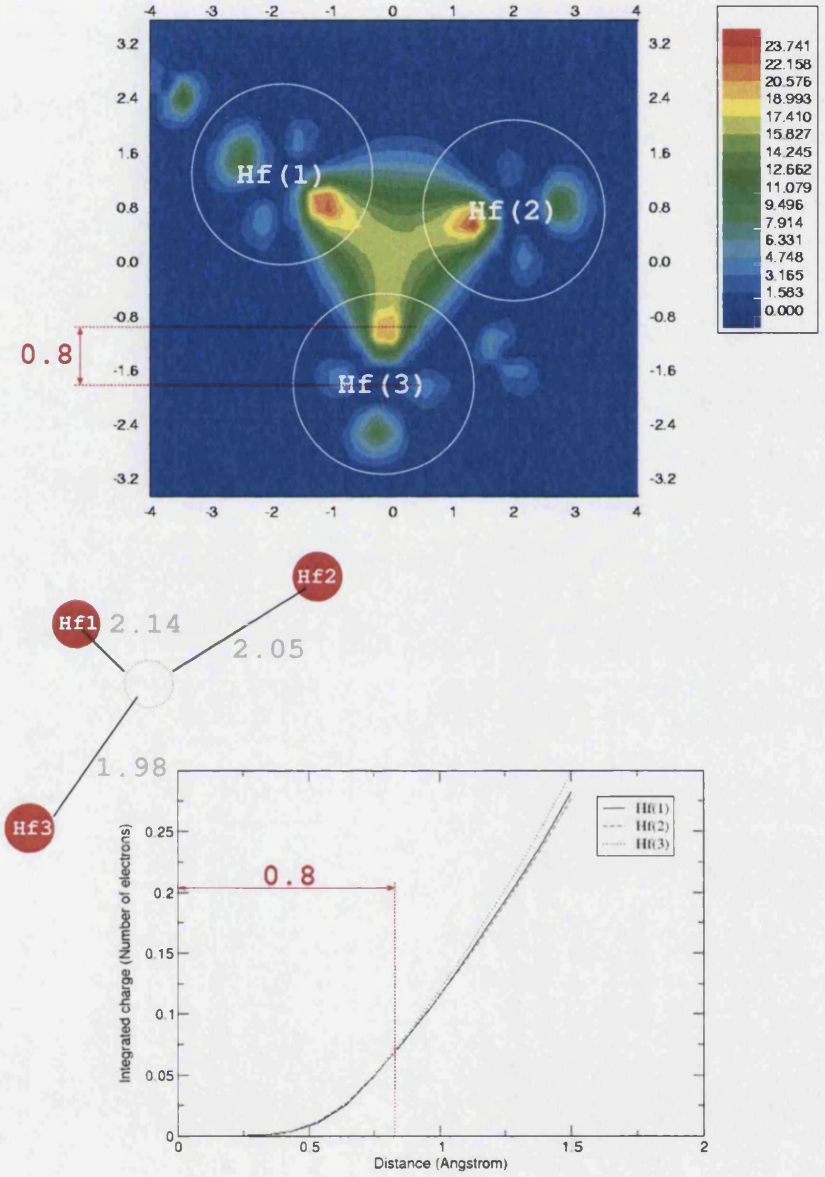


Figure 5.5: **Partial charge density maps of neutral three-fold Oxygen vacancy in HfO_2.** The density has been projected over the last occupied state (defect state) and mapped onto a plane containing the first neighbours (Hf atoms marked in the pictures). Because the vacancy site is not containded in this plane, the maxima of the density (red regions) appear to be near the hafnium ions. However, an integration of the electron density inside the spheres shown in the plot shows that actually only a portion of the density is contained in these regions. The distances along the axes are given in Å.

although their model concerns an oxygen vacancy in cubic zirconia only [111]. They describe the vacancy level being 2.1 eV below the bottom of the conduction band, which in their calculations turns out to be 5.55 eV. Therefore the distance of the vacancy level with respect to the top of the valence band is 3.45 eV, which is at least 1 eV above the value position of the levels found in our calculations. It is still difficult to compare these results, not only due to the difference in the structure of the materials (cubic against monoclinic), but also because the unit cell in the model by Králik an co-workers is considerably smaller (their cell contains 11 atoms, while ours contains 96). As a consequence, the interaction between defects in their case should be strong. Indeed, this is pointed out by the authors themselves in basis of the study of the dispersion of the defect level throughout the Brillouin zone. This interaction should unavoidably rise the energy of the level associated to the vacancy.

Turning back to the analysis of our results, both three-fold and four-fold coordinated vacancies in the two oxides have a similar behaviour. The following description can therefore be applied to all different types of oxygen vacancies.

It is possible to evaluate the localization of the electrons in these states by making a projection of the electronic charge associated to the defect state, as shown in figure 5.5. The doubly occupied electron state is localized in the region surrounding the site of the vacancy and limited by the first neighbours shell of hafnium atoms. The plane on which the density has been projected contains all three hafnium atoms, but the site of the vacancy is not contained on it. Because of this the graph may be misleading, showing that the electron density tends to accumulate near each one of the hafniums (red regions). To prove that the electrons accumulate on the vacancy site, it is possible to integrate the partial electron density corresponding to this state over spheres of radius 1.5Å (indicated in the figure as white circles). The maximums in the density map (red regions) are at a distance approximately equal to 0.8 Å. The integrated charge density, however, shows that the main part of the electron density is further away from all hafnium atoms. Note also how the differences in distance between the vacancy site and the hafnium atoms also affect the amount of charge contained on each sphere (seen as a difference in slope).

The first excited state can be determined by calculating the vacancy system in its triplet state. This level is very close to the bottom of the conduction band and it is much less localized. The charge density is maximum at the neighbouring Hf sites. We have considered the possibility of a singlet to triplet excitation of the system as a "vertical process" where the triplet (final state) has the same geometry as the relaxed singlet (initial) state. The difference in total energies between the singlet and triplet states is 1.20, 1.30 and 1.59 eV in the four-fold coordinated vacancy in zirconia and three-fold and four-fold coordinated vacancies in Hafnia, respectively. It is interesting to note that these energies agree to 0.1 eV with the single particle energy differences between the last occupied and the first excited levels. The exact values for these differences are 1.11, 1.22, 1.66 eV, once more for the three-fold and four-fold coordinated vacancies in ZrO_2 and HfO_2.

	ZrO_2					
	V_3^0	V_3^{1+}	V_3^{2+}	V_4^0	V_4^{1+}	V_4^{2+}
I_p(D)	3.45	3.53	-	3.80	3.98	-
χ_e(D)	-2.25	-2.68	-2.89	-	-3.33	-3.54
χ_h(D)	-2.72	-2.51	-	-2.10	-1.86	-
	HfO_2					
	V_3^0	V_3^{1+}	V_3^{2+}	V_4^0	V_4^{1+}	V_4^{2+}
I_p(D)	3.41	3.73	-	3.88	4.10	-
χ_e(D)	-1.80	-2.76	-2.93	-1.90	-2.26	-2.26
χ_h(D)	-2.92	-2.75	-	-2.42	-2.42	-

Table 5.9: **Ionization potentials, electron affinities and hole affinities for Oxygen vacancies in ZrO_2 and HfO_2, Ionization potentials (I(D)), electron affinities (χ_e) and hole affinities (χ_h) for Oxygen vacancies in ZrO_2 and HfO_2.**
The affinities correspond to fully relaxed configurations of both initial and final states of the process. All energies are given in eV.

When one electron is removed from the neutral singlet state, the vacancy becomes positively charged, and the electron remaining in the vacancy level remains localized. If, on the contrary, an electron is added to the neutral vacancy, it becomes strongly delocalised along the entire supercell. Because of this, the relaxation of the negatively charged vacancy is small, as aforementioned. The reliability of this result, however, may be affected by at least three factors. First, the DFT-GGA calculation predicts a too narrow band gap, so the added electron may be incorrectly placed in an state belonging to the conduction band. Second, the neutralizing background may introduce an unphysical interaction (see discussion in section 2.2.2). Finally, it may be necessary to consider a setup of the calculation that will allow a more accurate calculation of the delocalized state.

5.5.3 Ionization potentials and electron affinities of the defects.

As shown in the previous subsection, oxygen vacancies introduce a level in the gap of the bulk material. The ability of this defect to act as a charge trap depends on the amount of energy required for an electron to move between the defect level and any other state. The energies involved in these processes are, by definition, the ionization potential and electron affinities, and actually depend on the energy of the incoming electron. We have assumed here that the electron is moving to the closest empty one-electron state, located at the bottom of the conduction band of the oxide. It is worth noting, however, that in hypothetic MOS structures built with this oxides, it is more likely that the sources of carriers will be the valence and conduction bands of the semi-conductor substrate (as we have just discussed in previous subsection).

The ionization potentials, electron affinities and hole afinities of oxygen the different types of oxygen vacancies in both Hafnia and Zirconia are summarized in Table 5.9. These quantities have been calculated as described in subsection 3.2.3.

As expected, the energy required to trap an electron is different if the relaxation

	V_3^0	V_3^{1+}	V_3^{2+}	V_4^0	V_4^{1+}	V_4^{2+}
χ_e(D)	-1.80	-2.76	-2.93	-1.90	-2.26	-2.26
χ_h(D)	-2.92	-2.75	-	-2.42	-2.42	-

Table 5.10: **Comparison between vertical and fully relaxed electron affinities (in eV) of Oxygen vacancies in HfO_2.**

of the defect as it changes its charge state is not taken into account. The difference in energies, which corresponds to defect relaxation, is of the same order as the relaxation energies shown in subsection 5.5.1, involved in the processes of formation of the defects. The calculated electron affinities correspond to extreme cases of the real processes. This is because in the vertical processes no relaxation is considered, while in the relaxed processes a full relaxation is considered. In real systems, a relaxation will take place when the carrier is trapped, but it may not be a full relaxation if the carrier does not remain long enough on the site.

5.6 Cation vacancies, substitutional defects and complex pairs.

We have also looked at some other types of defects in pure Zirconia and Hafnia: the cation vacancy in HfO_2, substitutional Zr in HfO_2 and the complex pair Oxygen vacancy-substitutional Zr in HfO_2. The reason for choosing these defects is the difficulty found in eliminating Zirconium impurities from Hafnia: we wanted to investigate whether the presence of these impurities affects the electrical properties of the Oxygen vacancy and how. In case this does have an effect, the the situation will then be very similar to that of Ge impurities in Silica, which are of very much interest for high speed communications technology (see for instance [129] for further references).

The principal commercial sources of Zirconium are Baddeleyite (Zirconium oxide) and Zircon (Zirconium silicate). Both minerals are also the main source of Hafnium [113], which is naturally contained in approximately 1.5-3.0 %. The process of isolation of both metals is difficult, and most of the commercial Zirconium and Zirconium oxide contains Hafnium as an impurity, in the same proportion found in nature. This is also the reason for the high cost of Hafnium with respect of Zirconium, and of the mistake in the original determination of the atomic weight of Zirconium, due to the undetected presence of Hafnium.

The source of the difficulty found when separating Hafnium and Zirconium is the very similar behavior of both species in terms of chemical bonding. This is expected, since both elements have the same external electronic configuration. In order to evaluate the amount of energy involved in the process of formation of the defect, it is necessary to consider first the formation of the cation vacancy and then the addition

of the impurity.

5.6.1 Formation of the Hf vacancy. Calculation of the formation energy.

In the first instance, we generated the neutral Hafnium vacancy by removing the *atom* from the perfect supercell, following an identical procedure as for the Oxygen vacancy. The energies involved in the process of formation of the neutral Hf vacancy in HfO_2 can be determined from the equation

$$E_{for}(V_{Hf}) = E(V_{Hf}) + E(Hf) - E(prf) \tag{5.2}$$

If we use the energy for the isolated Hf atom, $E(Hf)^{atomic}$=-3.36 eV, calculated using the periodic scheme as described in section 3.2.1, the formation energy of the Hf vacancy is $E_{for}^{atomic} = 23.50$ eV. It is important to note that in this case the resultant energy does depend on which the atomic energy E(Hf) is taken for Hafnium. In order to avoid the problems involved in the calculation of atomic energies using periodic DFT techniques, it is possible to evaluate a system of bulk metallic Hafnium and find the energy per atom of this system. Using such reference makes sense because in the process of formation of the oxide, when used as gate dielectric [4], metallic Hafnium may be clustered inside the material. In our calculations, we obtained an atomic energy for metallic Hafnium equal to $E(Hf)^{metallic}$=-9.88 eV, which gives a formation energy of $E_{for}^{metallic} = 16.9$ eV. Finally, one can assume that the removed Hafnium remains in equilibrium with Oxygen vapour (this situation also agrees with the process of growth of the oxide over the Silicon substrate in MOS devices). In this case the energy of the Hafnium atom will be equal to $E_{Hf} = E_{HfO_2} - E_{O_2}$, where E_{HfO_2} is the energy of monoclinic bulk Hafnia and E_{O_2} is the energy of the Oxygen molecule. Molecular Oxygen can be calculated more accurately that atomic Oxygen if using periodic DFT [5]. Using this procedure the energy of Hf is $E(Hf)^{oxide}$=-20.83 eV and the energy of formation of the cation vacancy is $E_{for}^{oxide} = 5.76$ eV.

All these results are summarized in table 5.6.1. The process is endo-energetic and the amount of energy involved is significant. This is because of the large number of Hf-O bonds that need to be broken in order to extract the Hf ion.

The lattice relaxation involved in the creation of a Hf vacancy is larger than in the case of the Oxygens, and has energy equal to 1.57 eV. During this relaxation, the Oxygens move between 0.1 and 0.2 Å away from the vacancy site. This indicates that the Hafnium ion (with formal charge +4) provides a strong screening to the

[4] The growth of the oxide layer over crystalline substrate (when the oxide is not native silica) is done by deposition of the metal and by subsequent oxidation at high temperature in an atmosphere rich in molecular Oxygen.

[5] Actually, it is possible to estimate the error introduced by the DFT calculation of the atomic Oxygen by evaluating the dissociation energy of the Oxygen dimer by means of the formula $\Delta E^{theory} = E(O_2) - 2E(O)$. Using $E(O)^{atomic}$=-1.97 eV, the dissociation energy is ΔE^{theory}=5.88 eV, which is still above the experimental value ΔE^{exp}=5.17 eV. Also the electron affinity of the Oxygen atom [124] is overestimated by 0.2 eV.

Type of Hf	E(Hf)	Formation energy
Atomic	-3.36	23.50
Metallic	-9.88	16.90
Oxide	-20.83	5.70

Table 5.11: **Formation energies of the Hafnium vacancy in HfO_2.**
The value of this energy depends on the atomic energy of Hf. The table shows the values of the energy of atomic Hf, (bulk) metallic Hf and Hf in equilibrium with molecular Oxygen vapour. All three results depend on the accuracy of the DFT technique (see text for details). All energies are given in eV.

Oxygen atoms that surround it, and this screening disappears as soon as the ion is removed.

It is interesting to note that all the states introduced by the creation of a Hf vacancy are contained in the valence band, and therefore we do not expect this defect to play any important role as a charge trap. The abundance of cation vacancies is also not probable, since a high energy is required to create these type of defects. This energy may, however, be supplied by the process of incorporation of the Zr impurity.

5.6.2 Substitutional Zr in HfO_2.

For the study of the formation of this substitutional defect, we have considered directly the substitution of a Hf atom by a Zr atom at the same site. In a similar way as described for the other defects, the energy involved is

$$E_{for}(SZr) = E(S_{Zr}) + E(Hf) - E(Zr) - E(prf) \tag{5.3}$$

where E(Hf) and E(Zr) are the atomic energies for the isolated atoms. Using the atomic energies calculated in section 3.2.1 the formation energy turns out to be $E_{for}^{atomic}(SZr) = 0.86$ eV; if the bulk metal references are used, then $E_{for}^{metallic}(SZr) = 0.55$ eV ; if the oxide references are used $E_{for}^{oxide}(SZr) = -0.05$ eV. These results are reasonable, since the incorporated Zr atom is able to create bonds very similar those of Hf atom which it substituted, and therefore the energy gain due to bonding is similar to that required to break the previous bonds. These results also explain the difficulty in separating the metals from the minerals. The exchange of ions between the oxides is a process that almost does not require any energy, while the transfer of ions from the oxide to a metallic cluster requires as much as half an electron-volt per ion.

When the substitutional Zr is added to the vacancy site, the system undergoes a relaxation in the opposite direction, but with displacements of the same order of magnitude (0.1-0.2 Å), as those in the creation of the cation vacancy. The Oxygen atoms, however, are not able to return to their positions as in the perfect monoclinic Hafnia. This is consistent with the fact that the Zr-O bond in Zirconia is longer that the Hf-O bond in Hafnia.

In order to establish if the substitutional defect was able to act as a charge trap, we removed and added an extra electron to the relaxed system. In both cases (positive and negative charging) the hole and the electron (respectively) remained delocalized all over the supercell. This behavior is similar to that of the supercell of perfect Hafnia, and outlines once more the similarity of Hafnium and Zirconium in their electronic structure. This system does not mimic the behaviour of Silicon and Germanium systems, but structure considered here (monoclinic Hafnia) is much denser than that of silica polymorphs at normal pressure. One should also be aware that the DFT technique used may not be accurate enough to describe the existence of shallow states, due to its poor description of unoccupied states.

5.6.3 The substitutional Zirconium-Oxygen vacancy complex pair in Hafnia.

It is not probable that the industrial production of MOS devices based on using Hafnia as a gate dielectric (should this occur) will employ a material free of Zirconium impurities. Both experimental evidence and studies like the one presented in the previous section show that the process of isolation of one of the metal (or an oxide based only on one of them) is difficult and therefore, in economical terms, expensive. The use of Zirconium-contaminated Hafnia should not be, however, a problem, if its presence does not modify the essential properties of the material. The very similar chemical behaviour of both Zirconium and Hafnium ensures similar mechanical properties. In what concerns the electrical activity of the defects, in the previous section it is also shown that substitutional Zirconium is not itself a charge trap. In order to determine how the presence of this impurity affects the properties of some other defect that can trap charge, we have studied the complex pair formed by an Oxygen vacancy and a substitutional Zirconium.

The complex pair has been constructed using a supercell containing a three-fold coordinated Oxygen vacancy, replacing one of the Hafnium atoms neighbouring the vacancy. The relaxation process is more complex than in the single defects, but it can easily be understood in terms of the interaction between ions. The Oxygen atoms in the vicinity of the vacancy move towards the empty site. This happens because part of the repulsion is compensated by the attraction felt by these anions to the Zirconium atoms. Also feeling the presence of the Zirconium, some Hafnium ions displace away from the defect site. In all cases the displacements are very small, between 0.1 and 0.04 Å in the case of the Oxygens and 0.01 Å in the case of the Hafnium ions. The rest of the ions do not experience any displacement due to the presence of the defect.

The complex pair introduces a level inside the gap of Hafnia, 2.7 eV above the top of the valence band. This result is very similar to that of the single three-fold coordinated Oxygen vacancy (2.8 eV). We have also calculated the electron affinities and ionization potentials of the complex pair, by successively charging the defect. The behavior of the charged states mimics precisely those of the single vacancy. The

electrons in the neutral and single positively charged vacancies remain localized at the site of the vacancy, and the charge density around the Zirconium ion remains very similar to that of the Hafnium ions also neighbouring the vacancy. The presence of the Zirconium nearby the vacancy does not modify the electron affinities and ionization potentials in more that 0.1 eV.

5.7 Conclusions.

In this chapter I have presented the results of a study of oxygen vacancies (V_O) in zirconia and hafnia, the hafnium vacancy in hafnia (V_{Zr}), the substitutional zirconium in hafnia (S_{Zr}) and the complex pair S_{Zr}+V_O in hafnia, using a DFT-GGA scheme and plane waves basis set.

All defect structures in all charge states have been relaxed, to allow estimation of the energy involved in the relaxations after the formation or charging of the defects. In general terms, the relaxations do not involve big displacements of the ions, because of the dense nature of the material. All the relaxation processes can be justified in terms of the attractive and repulsive interaction between ions of different charge, and the rise or decrease of screening provided by electrons as they are added or removed. This is indicative that the system has strong ionic character, although an analysis of the Density of States of the materials shows that the top of the valence band is formed by a mixture of contributions from oxygen p orbitals and hafnium or zirconium d orbitals. This shows that the system retains some covalent character, which can make its modeling by means of simpler techniques inaccurate.

The energy involved in the process of formation of the anion vacancies is much lower than that involved in the formation of the cation vacancies, in both hafnia and zirconia. In both cases, the values of the calculated energies depend in the accuracy with which the value of atomic energies can be determined. The evaluation of the formation energy involved in the incorporation of a substitutional zirconium atom into hafnia, shows that the process requires very little energy for exchanging ions between oxides, but ten times more for the exchange of ions with the bulk of one of the metals. This justifies the difficulty in isolating zirconium from hafnium in nature.

The oxygen vacancies of both types (three- and four-fold coordinated) appear to be able to trap electrons when positively charged. The calculation of the defect levels shows that they should be able to trap electrons that tunnel from the bottom of the conduction band of silicon. The neutral oxygen vacancies are also able to act as hole traps. Negatively charged oxygen vacancies apparently do not act as traps. This result, however, may be affected by the incapacity of the technique to display the spectrum of unoccupied states correctly. The presence of a neighboring zirconium atom substituting a hafnium in the vicinity of the oxygen vacancy does not seem to essentially modify the properties of the defect. All the ionization potentials, electron affinities and the position of the level introduced in the gap differ from those of the isolated vacancy by less than 0.1 eV.

In conclusion, the potential application of either zirconia or hafnia as gate dielectrics should account for the effect of the charge trapping by even simple defects. The localized nature of the charge trapped may participate in breakdown mechanisms similar to those appearing in silica, and in any case the charging of the oxide will certainly have an effect on the band structure at the interface.

Chapter 6

Hydrogen in α-quartz

Aim. The structure and diffusion of atomic hydrogen inside α-quartz is studied. The energy barriers for hydrogen atom diffusion along selected directions are evaluated by calculating the adiabatic potential. The zero-point energy is accounted for *a posteriori*. The results are compared with EPR data by calculating the isotropic hyperfine constants.
Results. The adiabatic potential for hydrogen has two minima in the large channels along the c-axis of quartz. Hydrogen becomes polarized when placed inside silica, but does not establish any chemical bond with the lattice. The adiabatic barrier for hydrogen difussion along the c-axis is 0.21 eV, but lowers down to 0.13 eV if zero-point energy is taken into account. The classical treatment of the H nucleus, and the approximated treatment of the exchange-correlation energy constitute two important sources of error in this system. The hyperfine isotropic parameters cannot be described accurately, due to a large dependence on the quality of the basis set.
Techniques employed. First Principles Density-Functional Theory on its Generalized-Gradient Approximation (DFT-GGA).
Basis sets. Plane waves.
Models. Periodic models of \alpha -quartz, containing 73 atoms.
Computer Facilities. Self-consistent calculations of the electronic structure done using the Cray-T3E supercomputer of the CSAR service at Manchester (UK).

6.1 Introduction.

The two physical processes that prevent reduction of the size of a MOS transistor beyond certain limits are the tunneling of electrons through the gate dielectric layer and the breakdown of the dielectric [130]. The former process is undesirable since increases the leakage current and drives the consumption towards unaffordable levels, but it is not destructive, while the latter is. The breakdown of the dielectric implies a destruction of its micro-structure, and eventually causes device failure.

The mechanisms of a dielectric breakdown are complex, and probably not unique. In the specific case of the SiO_2 (by far the most used gate dielectric), there is quite clear evidence that hydrogen plays a fundamental role. Paradoxically, the presence of hydrogen is also desirable, since it helps to passivate the dangling bonds at the interface, reducing drastically the number of surface states. Because of both its beneficial and degrading effects, understanding the chemistry and physics of hydrogen inside the MOS systems will allow one to optimize device reliability.

Several works have tried to study particular steps of the mechanisms and structures in which hydrogen participates when placed inside silica. In particular, there is several works that make use of the Density Functional Theory to evaluate the electronic structure of the SiO_2:H system. On the other hand, modeling the complex MOS structures, where Silicon Dioxide displays an amorphous structure, is a difficult problem, as well as is interpreting the experimental data concerning this system. A possible alternative is to study the behavior of hydrogen inside one of the many crystalline phases of silicon dioxide. Such a system should allow one to understand the fundamental mechanisms and interactions of hydrogen inside silica, while rendering the analysis of the results considerably simpler.

In the particular case of the study presented here, we will center attention on the study of the diffusion of atomic hydrogen through α-quartz. The choice of such a host is due to the fact that there is actually a very interesting source of experimental information: Electron Paramagnetic Resonance (EPR) experiments. These type of measurements are able to provide information concerning the micro-structure of a paramagnetic center (as is the case, since atomic hydrogen has an unpaired electron).

We have explored the possible paths of diffusion for hydrogen, and the energies involved in the process, by sampling the adiabatic potential surface for hydrogen inside the silica host. To do this we have performed several structural relaxations of a periodic model of the system, at Density Functional Theory level. We have also considered the validity of the adiabatic approximation in the extreme cases of very light particles, as hydrogen atom is. This has been done by evaluating the zero point energy of this particle at its equilibrium position. The results have been compared with those coming from EPR experiments in two different ways: a) by comparing the possible most favored sites for hydrogen, as predicted by EPR and as predicted by our calculations, and b) by calculating the isotropic hyperfine constants that can be measured directly in the experiments.

The chapter is organized as follows. In first place, I will review the knowledge in the subject - presented in a bibliographic review. This review includes the suggested models of dielectric breakdown in which hydrogen takes part, the data concerning the diffusion of Hydrogen through silicon dioxide, and a brief overview on the basis of the EPR technique. In second place I will discuss the motivation for choosing the DFT periodic models and the setup of the calculations. Finally, I will present the results, and discuss them in depth.

6.2 Bibliographic review.

6.2.1 Hydrogen in MOS systems.

The study of the role of hydrogen inside the MOS systems is certainly among the classics of the micro-electronics technology. It is well known that many steps in the fabrication of MOS systems, in which wet reactors are often employed, allow the introduction of this impurity inside the device. The elevated temperatures of

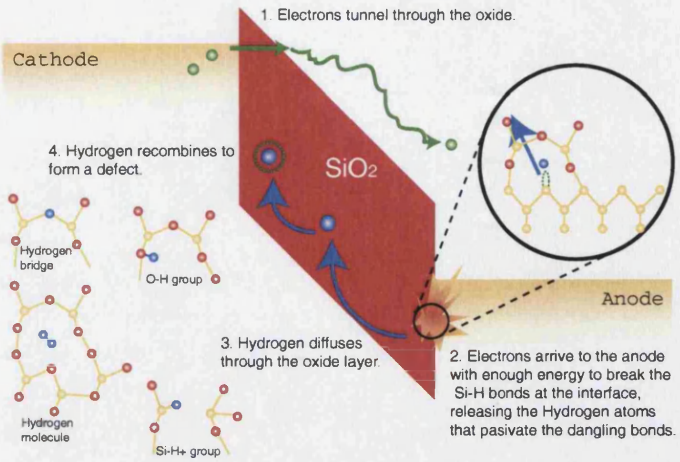


Figure 6.1: **Generation of hydrogen-related defects in a MOS system.**
Hot electrons can cause the release of hydrogen from the interfaces, that will diffuse through the oxide and recombine forming some of the defects shown. The schematic representation is inspired in that of [135], also shown in [3].

some of the steps of the manufacturing process and the light nature of hydrogen cause its fast-diffusion through the hosts, and make it very difficult to control its concentration.

The presence of hydrogen is far from being undesirable. One of the first problems found on the development of planar circuits was the high density of defects caused by lattice mismatching at the interfaces. hydrogen is capable of saturating the dangling bonds and therefore reduces the amount of surface states. Also, implantation of high doses of hydrogen allows one to generate a highly resistive layer that separates a thin layer from a substrate [131].

On the other hand, many hydrogen related defects, like atomic and molecular hydrogen, Si-H^+ and O-H groups, and bridging H (a hydrogen trapped at the site of an Oxygen vacancy) influence the electrical behavior of the Si/SiO2 system [132]. Some of these systems introduce defect levels in the band structure of the host, and therefore are able to act as charge traps. These traps reduce the insulating capability of the oxide and eventually trigger its breakdown. The generation of these defects in many cases takes place during the operation of the devices, as it is illustrated in Figure 6.1. If the gate dielectric layer (made of silicon dioxide) is thin enough, electrons can tunnel through, and arrive to the anode interface with enough energy to break the bonds of hydrogen atoms that are saturating dangling bonds (this energy can be as low as 2 eV [133]). These atoms can then diffuse rapidly through the oxide and recombine, forming one of the mentioned defects. Some other non-desired processes in which hydrogen takes part are several chemical reactions in the SiO_2 films, such as the nitridation via the use of NH_3 molecules (used during the manufacturing process). A review of the chemistry of hydrogen inside SiO_2, and its negative effect in the MOS systems has been presented by Poindexter [134].

Part of the understanding concerning the mechanism of release of hydrogen from

passivated bonds at the interface has been provided by the substitution of hydrogen by deuterium. A significant improvement of the immunity against electrical stress of the deuterium-treated-MOS systems with respect to hydrogen-treated ones has been reported by several groups (see [3] for further references). It has been suggested that the release of the hydrogen (deuterium) atoms at the interface happens through the excitation of the vibrational Si-H (Si-D) modes by the hot carriers.

There is a wide collection of works that address the issue of the potential role of the hydrogen-related defects as charge traps. Electronic structure calculations are convenient tools to help in the interpretation of experimental results, since they allow one to obtain details of the micro-structure, to locate the position of the defect levels with respect to the band structure of the solid, and to determine the energies involved in the formation and charging of the defects. Very interesting examples of this type of approach can be found in the works by Blöchl [17], and Van de Walle and Tuttle[136].

The H_2 molecule, for instance, has been reported to be useful for the reduction of the defects inside the silica network by helping to dissociate O_2 molecules that then bond low-coordinated Silicons and/or fill vacancy sites [137]. Theoretical predictions (periodic DFT calculations) done by Blöchl suggest that H_2 is inert in non-defective silica [17], while Edwards (using a cluster model) outlines the need to consider the quantum nature of hydrogen nucleus in order to justify properly the behavior of such system[14].

In what concerns atomic hydrogen, the most recent theoretical calculations point towards the idea that the neutral species is thermodynamically not stable, and should rather recombine to give a H^+/H^- pairs of H_2 dimers. This conclusion was based on periodic DFT calculations both by Yokozawa and Miyamoto [18] and Blöch [17], and of molecular cluster calculations by Edwards and co-workers [13]. The argument apparently contrasts with experimental evidence. The Electron Paramagnetic Resonance (EPR) setups are able to detect the unpaired spin associated to the neutral hydrogen. The EPR spectra shows the presence of hydrogen inside α-quartz at low temperatures, but the signal disappears above 100-130 K [138]. The most straightforward interpretation of this result is that hydrogen remains as an isolated neutral entity at low temperatures, while above the mentioned temperature becomes highly mobile and/or recombines, so the unpaired spin disappears. Edwards discussed that the recombination of neutral hydrogen should actually be regulated by the height of the diffusion barriers (in the case of the formation of H_2) and by the activation energy required to take an electron from a neutral hydrogen to the conduction band of SiO_2 (in the case of the H^+/H^- pairs, since an electron needs to be transfered from one H to the other).

The study of the barriers for diffusion has been addressed at very different levels of theory, from macroscopic kinetic models [139] and classical molecular dynamics [140] to first principles Density Functional calculations [19, 18]. Not very surprisingly, the range of predicted values spreads considerably. The fact that hydrogen is a very light

element makes it difficult to construct models based on classical descriptions. As an example, Bongiorno *et al.* [140] have performed a classical molecular dynamics to study the diffusion of hydrogen inside α-quartz, and they estimated that the height of the energy barriers for diffusion is 1.7 eV, and concluded that hydrogen diffused through the narrower channels in the *alpha*-quartz structure, along the c-axis of the crystal [1]. The authors argue that very good agreement is achieved with some first-principles calculations performed using an STO-3G set. It is a well established fact, however, that such a basis set does not provide an accurate description of the silica polymorphs [83].

In contrast with these results, first principles electronic structure calculations using plane waves basis set and a DFT-GGA Hamiltonian have been used by Tuttle [19] to perform the same type of study, under the assumption that the silicon dioxide host does not offer any response to the presence of hydrogen (frozen phonon approximation). In this study the value of the barrier turns to be 0.2 eV. In calculations of the same quality, Blöchl shows that actually the preferred site for hydrogen inside quartz is the center of the wide channels in quartz. Unfortunately, the accuracy of any of these works is difficult to determine, since the spread of the experimental values for the diffusion barriers is similar to that of the theoretical works, in a range from 0.07 to 1.98 eV (see [140] for further references).

As can be seen, the amount of studies on the behavior of hydrogen inside the MOS systems (more particularly, in the oxide region) is considerably big and, while certain aspects have become more clear, some others are still puzzling. The use of electronic structure calculations can be very helpful, but it is also necessary to consider their limitations in the accuracy that they can provide, and in the complexity of the systems that can be studied. In the next section I will present the system studied and for the technique employed on it.

6.2.2 Choosing the case study: crystalline hosts.

Most of the experimental data related to the study of diffusion are based in electrical measurements (see, for instance the work by Cartier *et al.* [133]). These techniques find some difficulties to identify which is the species that is actually diffusing (i.e. which is the charge state of hydrogen, or if it is atomic or molecular hydrogen). This, as is suggested by Tuttle, may be one of the reasons why there is such spread of the measured energy barriers [19]. There is, however another type of experimental setup that allows one to detect the presence of atomic hydrogen: the EPR technique. As it has already been mentioned, the temperature over which the signal associated to the unpaired hydrogen spin fades away can be related to the energy required by hydrogen to recombine. Furthermore, an analysis of the hyperfine spectra allows to determine which is the closest environment of hydrogen.

In view of the information that the EPR techniques are able to provide, the

[1] A schematic representation of the structure of α-quartz can be found further in the text, in Figure 6.2.

calculations presented here will focuss on two issues: a) determining if there are stable sited for hydrogen inside silica, and b) determine the energy involved in the diffusion of this species through the host.

The first point will be addressed in two different ways. Firstly, it is possible to explore the energy surface of the system as a function of the coordinates of the constituent ions, and therefore detailed information about the equilibrium structure can be obtained an compared with that suggested from the analysis of EPR data. This experimental information is obtained (as will be described in section 3.3), through the fitting of a phenomenological Hamiltonian to match the values of the hyperfine constants. The second possibility, is then to calculate directly the value of the hyperfine constants from the information from electronic structure calculations. The difference between both approaches is that in the first one the experimental technique is providing indirect information, while in the second one it is the theoretical technique that does so.

The basic idea of this study is to test, in a system for which there is direct experimental information available, if the theoretical method is capable of providing a description of the basic mechanism that justify the behavior of hydrogen inside silica. I have not expected to provide any conclusion that could be applicable to the more complex case of MOS systems. Such a target is now within the limit of the affordable with present computational resources and it is yet to be proved if results concerning crystalline materials can be extrapolated to amorphous systems. If the technique is validated, however, it can be assumed that future progress of the computing technology will allow such studies.

Up to my knowledge, the EPR data available [20, 138, 62] has always considered α-quartz as a crystalline host, and therefore it is logical to consider this host also in the calculations. The use of crystalline silica as host for hydrogen is very convenient, since it eases the interpretation of the results. In crystals, the translational symmetry reduces the number of possible sites for the impurity, while in amorphous materials the number of possible different sites is virtually infinite and forces one to elaborate an statistical sampling. From the point of view of calculations the translational symmetry of the crystals allows a considerable saving of computational resources, by using periodic models.

6.3 Presentation of the results.

6.3.1 Set up of the calculations.

The work presented here is based on DFT-GGA calculations using a plane wave basis set. At the time of realization of this study, there was no code available that would allow one to perform structural relaxations using periodic models at the Hartree-Fock level. Such types of calculations are essential to explore the energy surface of the system, and determine the energy barriers.

The evaluation of the hyperfine constants, however, required the use of a all-

electron basis set, since the isotropic hyperfine signal depends essentially on the electron density associated to some of the inner core electrons. All-electron basis sets are avoided when using plane waves, due to the elevated number of basis functions required to describe properly the localized core functions. This has forced us to consider using two different codes, and split the study in two steps.

The first stage of the study has targeted the sampling of the energy surface for hydrogen, employing the advantages of the VASP code, that allows efficient structural relaxations using plane-waves DFT-GGA. In a second stage, selected geometries obtained from the structural relaxation have been ported and recalculated using CRYSTAL98, which is based on the use of localized basis sets and therefore allows the use of all-electron basis sets (although does not allow one to relax the structures). The consistency between both types of results has been checked and will be presented later. Prior to the detailed description of the calculations and the results, I think that it is worth providing a very basic overview of the EPR technique, that, as I said, is the fundamental source of the experimental results that will be used as reference.

The adiabatic barriers for hydrogen inside quartz were explored using VASP 4.4 code [120, 121, 32], which implements a periodic Density Functional scheme based in the Generalized Gradient Approximation known as Perdew and Wang 91 (GGA-II). Given that hydrogen has a single electron, spin-polarization was included in the calculations. The basis sets are formed by plane waves, and the cores of the atoms are simulated by means of Vanderbilt ultrasoft pseudo-potentials [53, 52]. The pseudopotentials were generated considering the electron configurations $[Ne]3s^2 3p^2$, $[1s^2]2s^2 2p^4$ for the Oxygen atoms and $1s^1 1p^0$ for the hydrogen atom (the core electron configurations are shown inside the brackets). Although using a pseudopotential for the hydrogen atom may seem surprising, since it only contains one electron, the authors of the code justify its use because allows a correct description of the hydrogen atom and H_2 dimer while keeping the cut-off energy low [32] (later on, I will present some tests performed to check whether this statement is valid or not).

The first stage of the calculation was to optimize the geometry of the host lattice. This work was done using a supercell containing 36 atoms, resulting from a $2 \times 2 \times 1$ expansion of the crystallographic unit cell. A schematic representation of this cell can be seen in figure 6.2. The sampling of the reciprocal space was done using a set of ten special k-points, and the cut-off energy was kept to a value of 514 eV, in order to obtain the diagonal elements of the stress tensor accurately. During the calculation of the adiabatic barriers for hydrogen the cut-off was reduced to 400 eV, which still allowed an accuracy up to 10 meV in cohesive energy.

During the optimization of the structure no symmetry constraint was kept over the positions of the ions. The geometry of pure α-quartz is reproduced as follows. The calculated unit cell parameters are $a = 5.026$ Å, $b = 5.021$Å, $c = 5.509$Å. Since these are "temperature zero" calculations, they are to be compared with measurements at low temperature as, for instance the following experimental values [141], taken at 13K: $a = b = 4.902$ Å and $c = 5.400$Å. This means that our calculations overestimate

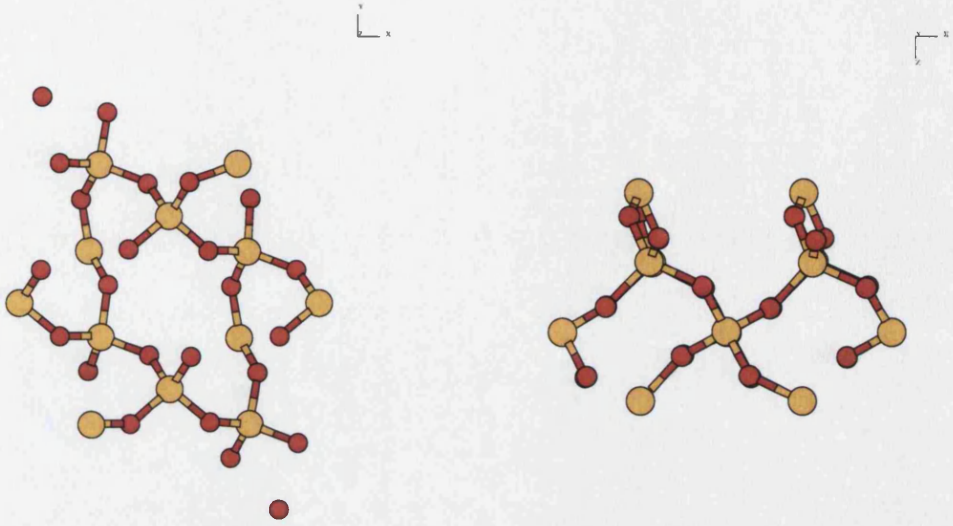


Figure 6.2: **Supercell of α-quartz.**
It has been constructed as a 2x2x1 extension of the crystallographic unit cell. The supercell contains 36 atoms (without hydrogen inside).

the length of the cell parameters by an average of 2.2% of the true values. Concerning the local structure, Si-O bond lengths are on average 1.618 Å for the long bond and 1.615 Å for the short bond. These values also overestimate the experimental values, but by less than 1%. In what concerns the electronic structure, the band gap of Silica, measured as the difference in energy between the one-electron energies associated to the last occupied and the first non-occupied orbitals, turns to be 6.07 eV, in contrast with the 8.8 eV detected experimentally [10]. This underestimation is typical of the DFT schemes.

6.3.2 Determination of the lowest energy site for hydrogen.

The determination of the lowest energy site was done by performing geometry relaxations using a set of different starting geometries, which differed basically in the position of hydrogen with respect to the lattice. In all of the cases tested, the minimum distance between hydrogens is 5.5 Å. This distance is enough to ensure that there is essentially no contribution to the total energy caused by interaction between defects in different cells (evaluated as described in 2.2.1). The change in total energy involved in the structural relaxation is 0.18 eV. This amount is significantly bigger than that obtained by Tuttle [19] for hydrogen inside cristobalite.

The lowest energy site found is displayed in figure 6.3; hydrogen is placed along the c-axis of the crystal, contained in a plane perpendicular to this axis that also contains several Oxygen atoms. These atoms are the closest neighbours to the hydrogen, being the minimum O-H distance 2.13 Å. This result is among the possibilities expected; by intuition, hydrogen should tend to be in the center of the channels, because the symmetry of this position minimizes the action of the crystalline field.

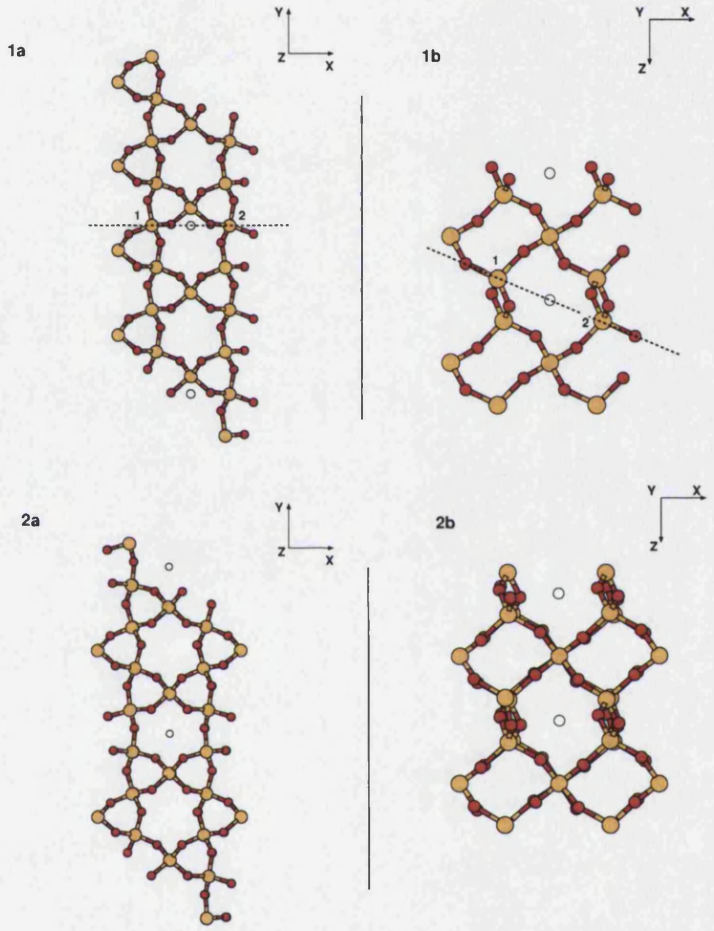


Figure 6.3: **Schematic representation of possible sites for H inside α-quartz.** The two top figures (1a and 1b) display the position predicted by EPR experiments (figures 1a and 1b), which is named $M_{EPR}^{unrelax}$ along the discussion in the text. The two bottom figures (2a and 2b) correspond to the position predicted by total energy minimization using a DFT-GGA Hamiltonian and a plane-waves basis set, which is named as M_{DFT} in the text. The dashed line in 1a and 1b corresponds to a two-fold axis in which hydrogen apparently sits, as it is deduced from the symmetry of the hyperfine tensor obtained in the experiment [62]. The axis also links the sites of the Si ions marked as 1 and 2. Although the view from a plane perpendicular to the Y axis (1b and 2b) is the same for both the experimental and the calculated sites, a view from a plane perpendicular to the Z axis (1a and 2a) shows considerable difference.

The position found as global minimum (M_{DFT}) seems to agree with the results obtained by Blöchl [17], which is the only other calculation done at DFT level using α-quartz as host lattice. However, this is not the site that is proposed by Weil and co-workers as deduced from their EPR measurements [20, 138], whose position is also indicated in figure 6.3. This site ($M_{EPR}^{unrelax}$) has two-fold symmetry, and the axis (also marked in the figure) goes through two Silicon atoms, each one at an opposite side of the cavity. The disagreement between the position predicted by Weil and the result of our calculations can be see clearly if the system is projected in a plane perpendicular to the c-axis (see figure 6.3). The minimum detected by EPR stands 1.30 Å away from the closest point contained in a c-axis, and it is 2.09 Å away from the closest Oxygen (its closest neighbour).

In order to understand if the position suggested by the EPR experiments corresponds to a minimum on the total energy surface, we have performed a structural relaxation, taken as starting geometry that of $M_{EPR}^{unrelax}$. We have in fact found such a minimum (M_{EPR}^{relax}), which stays 0.13 eV above the minimum M_{DFT}, found previously. The position of the second minimum M_{EPR}^{relax} does not correspond exactly with that predicted by the EPR measurements, since it is displaced 0.29 Å away from the two-fold symmetry axis. The difference between both sites is shown in figure 6.4.

It is interesting to note that in the displacement from $M_{EPR}^{unrelax}$ to M_{EPR}^{relax}, hydrogen moves towards the lattice, indicating the existence of a possible interaction with the Silica host. In this position the shortest H-O distance is 2.32 Å, therefore indicating that the hydrogen tends to move away from the Oxygens. A plot of the partial charge density corresponding to the minima is shown in figure 6.5. The plane of the figure contains the hydrogen atom and two Oxygens. As can be seen, there is very little transfer of charge from the hydrogen towards the Oxygens. The cumulative integral shows, as in the case of M_{DFT}, that the electron density associated to the hydrogen state is much more delocalized than in the case of isolated atomic hydrogen.

The most surprising fact in these results is that both minima found are actually placed in the same channel inside α-quartz (as its shown in Figure 6.3). There is only an empty region between both minima, and therefore it seems reasonable to think that hydrogen will be free to move towards the lowest minimum. A possible explanation for this behavior is that the second minima is actually due to the interaction between hydrogen and the host. The mechanism that rules this interaction should be the electrostatic interaction due to polarization of hydrogen.

To justify this hypothesis we have analysed in detail how the electron charge distributes around hydrogen. VASP allows to isolate the charge density associated to a specific band. In this system, the last occupied one-electron state is single-occupied, and the energy associated to this state is placed inside the range of energies corresponding to the gap of silica. We have therefore identified it as the level introduced by hydrogen.

In first place, we have studied the confinement of the electron density associated

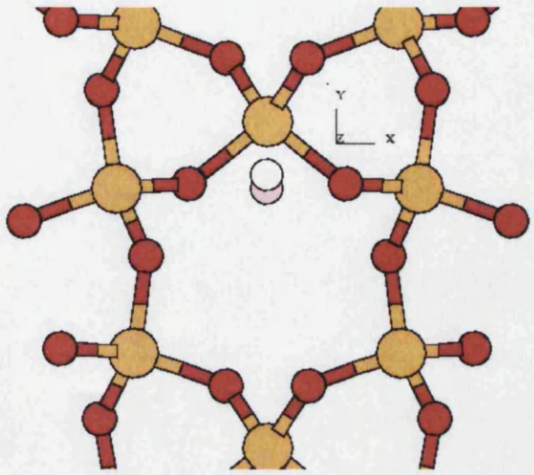


Figure 6.4: **Results of the structural relaxation.**
Using the experimental EPR geometry as starting point (M_{EPR}^{relax}, shown as a pink circle). hydrogen atom displaces towards a position nearer to the lattice (M_{EPR}^{relax}, shown as a white circle), and out of the axis that links both Silicon atoms to the right and the left of the initial position.

to this state. To do so, we have integrated the density contained in a sphere with center in the position of hydrogen's nuclei. By finding the value of this integral for different radii of the sphere, we obtained a representation of the amount of electrons as a function of the distance from hydrogen's site, as shown in Figure6.5. The graph displays the integrated density for both minima (plain and dashed lines). The changes in the slope of the curves indicate that the distribution of charge is not homogeneous. This is due to the interaction with the neighboring ions. To obtain a reference of the confinement, we compare the integrated density with that of isolated atomic hydrogen (dotted line). The calculation for isolated hydrogen has been performed using identical setup than that used for hydrogen inside quartz, but the unit cell contained a single hydrogen atom. It is possible to see how the electron in isolated hydrogen remains more confined than when H is inside quartz; In the case of isolated hydrogen, 0.99 of the electron is contained within a sphere of 2.3 Å, while in both cases corresponding to H inside quartz, it is necessary to consider spheres of radius 5 Å to find the same amount of electron density contained inside the sphere. It is therefore true that the hydrogen nuclei loses part of the screening provided by the electron, since the former becomes more delocalized.

The integration of density does not provide much information about the shape of the density, since the integration is performed in spheres. To obtain more insight into this issue, we have projected the density associated hydrogen's state into several planes. The results are shown in Figure 6.6. The planes have been selected as to contain the point at which hydrogen nucleus is placed. This point appears at the center of all plots, and each of these graphs shows a square region 5 Å wide. The planes have been selected to be parallel to the X- and Y-axis, X- and Z-axis and Y- and Z-axis, and there is a collection of these projections for each one of the two minima.

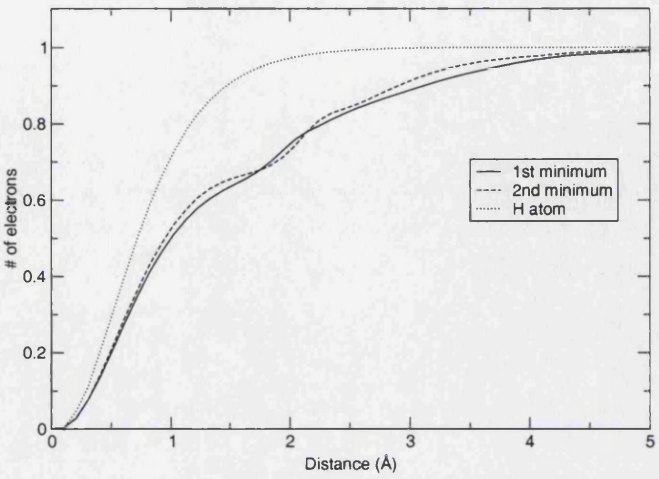


Figure 6.5: **Confinement of the electron of hydrogen atom.**
To see how confined is the electron trapped at the hydrogen atom, we have integrated the charge density associated to the last occupied orbital of the system in spheres with center at hydrogen's site. At any of the minima, the electron extends over a volume bigger than the one occupied by the electron of free isolated hydrogen (also shown for comparison).

When placed at any of the minima, part of the electron density associated to hydrogen is transferred to its first neighbors. Note that the planes shown here do not contain any of these neighbors. This redistribution of the density should cause the appearence of multipoles, although the amount of charged which is transferred is very small. Since the relative position of the neighbors with respect to hydrogen's site is different for each minima, the orientation of the multipoles is also different.

One possible cause of the stabilisation of the second minima, given the appearence of multipoles, is the interaction between defects of different cells. However, the estimation of the defect-defect interaction (as discussed in 2.2.1) showed us that the energies involved are not significant.

In principle, there is no reason to believe that several other minima may appear as a consequence of the same mechanism. Exploring such possibility, as well as the barrier between the two minima that we have described, was out of our possibility, given the cost of the calculations.

6.3.3 Determination of the adiabatic barrier along the c-axis.

Taking as starting point the lowest energy site found in our calculations, we explored the changes in total energy found by hydrogen as it moved along a direction parallel to the c-axis of the crystal. This displacement seems to us as the most probable since hydrogen will not find any ion on its way, and will be able to stay in symmetry positions. The adiabatic curve was determined by performing several structural relaxations in which hydrogen was constrained not to move along the c-axis, to

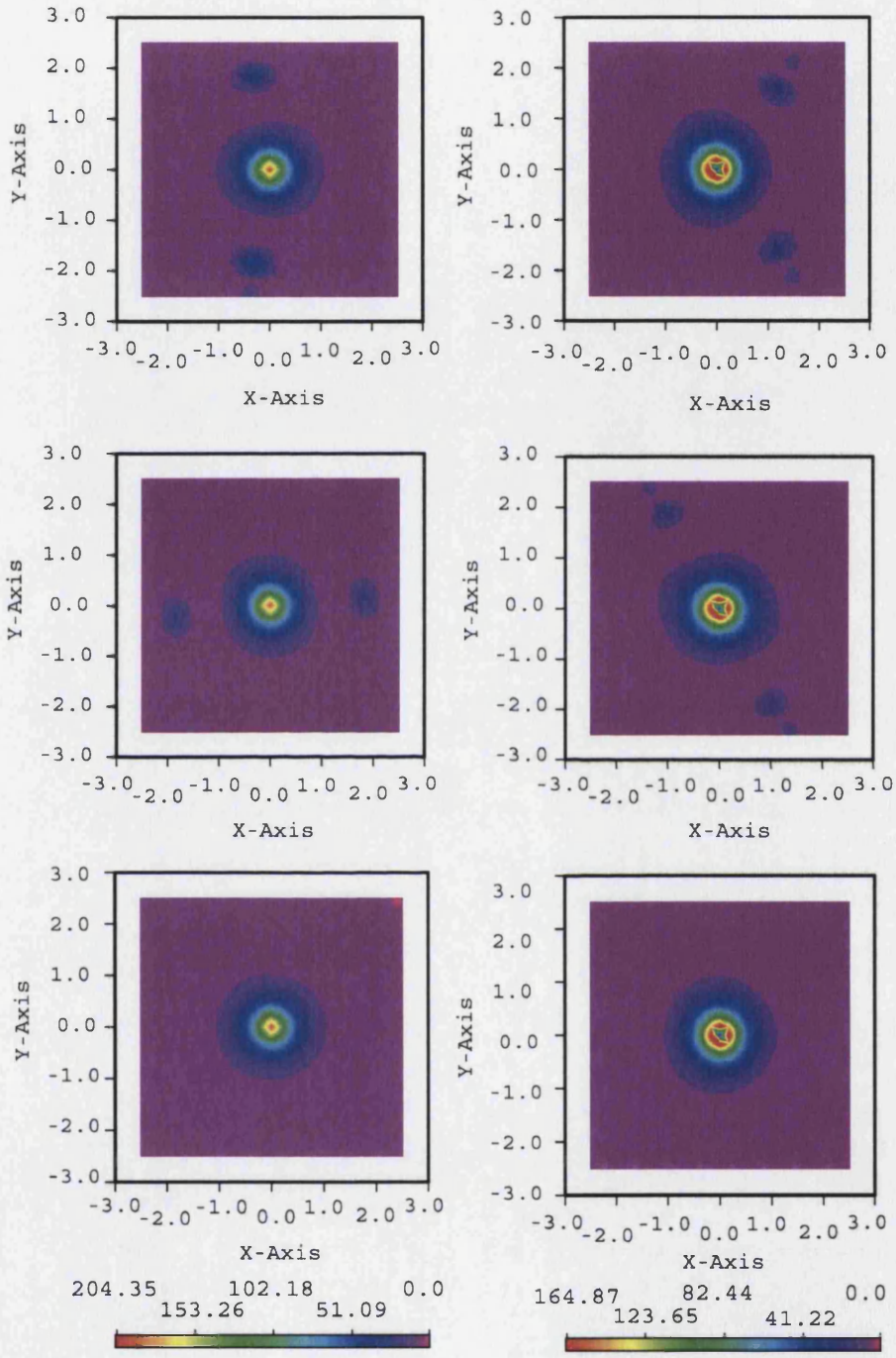


Figure 6.6: **Analysis of the density at the minima for hydrogen inside α-quartz.** These graphs corresponds to projections of the density associated to the last occupied state of the system. This state can be attributed to the presence of hydrogen (see text). The plots on the left column correspond to the lowest minimum, and the plots on the right column to the second minimum. Hydrogen atom is placed at the center in all graphs. From top to bottom, the graphs correspond to planes parallel to the XY, XZ and YZ axis. The scales on the bottom correspond to all the graphs in the same column. Units of the density are ($e^{-2}/\text{Å}^3$).

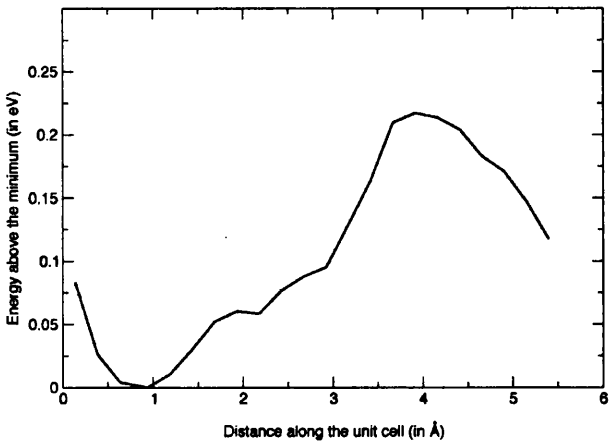


Figure 6.7: **Adiabatic potential for hydrogen inside SiO_2.**
It has been constructed by calculating first the most relaxed structure, and then displacing hydrogen atom along the Z axis. The graph shows the change in the total energy of the unit cell as this displacement takes place. Notice that the potential has the periodicity of the cell, and it is asymmetric around the minimum.

prevent this atom from returning towards its minimum energy site. Most of the atoms of the lattice were allowed to relax their positions. Also hydrogen was left free to move in a plane perpendicular to the c-axis.

The reason why not all the positions of atoms of the supercell were allowed to relax is to be found in the way VASP performs selective structural relaxation. During the relaxations, the code first performs a self-consistent calculation to determine the electron density, then used this information to evaluate forces over each ion. If some ions are selected not to move the computer makes the value of the gradient for the specific ion zero. Then the atoms are displaced following the gradients and a new self-consistent calculation is started. Making some of the gradients zero is equivalent to add a force of the same magnitude and opposite sign to all the other ions, that will be consequently displaced. In the case presented here, fixing the position of hydrogen in c-axis may have caused all the rest of the cell to feel a force in the direction opposite to the displacement. Then the whole supercell could have displaced itself as a block, as to reach once more the minimum energy configuration. In order to avoid this some of the ions in the edges of the supercell were kept fixed, serving as "anchors" for the lattice structure.

The relaxation of the position of the ions belonging to the host keeping hydrogen atom partially fixed is equivalent to assuming that the vibrational frequencies of the host depend weakly on the impurity [142]. This approach is less drastic than assuming that only hydrogen moves in response to the interaction with the lattice.

The results of the calculation of the adiabatic potential are shown in figure 6.7. The first thing to notice is that the potential displays translational symmetry, but it is not symmetric around the minimum. This is, however, consistent with the trajectory, and it is related to the distance that the hydrogen keeps with respect to the host atoms. In appearance, there is a secondary maxima at about 2 Å in the

horizontal scale shown in the graph. Such maxima is most likely due to not having required enough accuracy in the relaxation of the structure. Improving the quality of the calculations, however, would require to push the convergence criteria both in the convergence of the total energy and the gradients. As a result, the cost of the calculations will increase substantially, since it will be necessary to resolve changes of less than 1 meV per atom inside the cell.

Given that the differences in total energy between the minimum and maximum of the adiabatic curve are in the order of tenths of eV, the criteria for stopping the structural relaxations was to obtain convergence in the gradients up to 0.01 eV. The total height of the barrier is 0.21 eV. This value is very close to that obtained by Tuttle [19], and the experimental value for the activation energy for hydrogen diffusion presented by Cartier and co-workers [133]. We think, however, that the value of the real barrier should in fact be lower. In the following section, I will discuss several arguments which may be affecting the accuracy of the description.

6.3.4 Comparing with experiments. Hyperfine constants.

The isotropic hyperfine constant can be calculated once the electron spin density is known, using the expressions already described in 3.3. The calculation of the isotropic hyperfine constant requires performing all-electron calculations, since the spin density over the nucleus is given by the contributions of s orbitals only. This is the reason why we employed the CRYSTAL98 package, which uses local functions as a basis set.

In this case the basis for Silicon (6-21G*, outer exponents $\alpha_{sp} = 0.13$ Bohr^{-2} $\alpha_d = 0.5$ Bohr^{-2}) and Oxygen (6-311G*, outer exponents $\alpha_{sp} = 0.28$ Bohr^{-2} $\alpha_d = 0.6$ Bohr^{-2}) considered have been optimized for the description of α-quartz by Civalleri *et al* in [83] (this set has also been employed in the study presented in chapter 4, where it is described more extensively). This set describes correctly the geometry of the system when used together with the UHF Hamiltonian, with an error in the distances within 0.01 Å in the cell vectors and 0.002 Å in the Si-O distances. The average Si-O-Si angles are reproduced with an error of 1.3° and the O-Si-O average angles are reproduced with an error of 0.1°. The stability of different silica polymorphs is also predicted correctly. For the hydrogen, I have used a double-ζ basis set containing polarization functions. This set has been employed previously in the calculation of EPR parameters by Pacchioni *et al* [15]. Also some other lower quality basis sets have been used for testing the dependence of the results on the quality of the basis.

The geometry of the cell used in the calculation corresponds to an "idealization" of the structure found as lowest energy site using the plane-wave DFT technique. The relaxed host lattice was substituted by a perfect lattice generated by exploiting the full point symmetry displayed by α-quartz. The information concerning the size of the crystallographic cell and position of ions inside was taken from experimental data. Hydrogen is placed on its lowest energy site as described in section 6.3.2.

The code is able to evaluate the hyperfine isotropic constants on request, using

tabulated values of the g factors for the nuclei (in the case of hydrogen g_H=5.5856948). All the data concerning the values of the different constants can be found in the manual of the code [30]. The value of the isotropic constant for atomic hydrogen inside α-quartz is $a_{iso}(H) = 44.13mT$. This should be contrasted with the experimental value of 51.80 mT [138]. The reasons of the disagreement can be: a) due to difference in the position of the site for hydrogen and/or b) due to error introduced in the calculation of the constants.

Unfortunately, the second source of error has a significant weight in this case. The isotropic constant is evaluated through an expression that depends directly on the spin density of the system. The quality of the density, in comparison to the 'real' density, is affected by many factors, the most significant being probably the Hamiltonian and the basis sets. To test how the choice of these parameters affected the calculation of the hyperfine isotropic parameter, I have repeated the calculation with identical geometry and different Hamiltonians and basis sets. The basis set employed, apart of the one already mentioned, are the standard 6-21G and 6-21G* (value of the exponent of the polarization shell is $\alpha_d = 0.6$ Bohr^{-2}).

Table 6.1 shows a comparison of the values obtained for the hyperfine isotropic constant for hydrogen trapped inside quartz and also for free hydrogen in depending of which basis set and which Hamiltonian has been used in the calculation. Although the agreement with experimental values (also shown in the table) is not bad with certain setups, it is discouraging that the results are so sensitive. The calculations done at Unrestricted Hartree Fock (UHF) level allows one to check the dependence on the basis set. The UHF technique has been chosen because it allows one to evaluate the exchange interaction exactly. On the other hand, we are aware that the UHF calculation may be problematic due to the fact that the wavefunction is not an eigenfunction of S^2.

The poor results obtained with basis sets 1 and 2 are due to the fact that the hydrogen basis set does not contain any polarization function. The results obtained with basis 4 and 5 sets are of similar quality, only difference between sets being the polarization functions on Si basis. This indicates that the EPR signal over the H site is not strongly influenced by the polarization functions of the lattice. The basis set optimized for α-quartz by Civalleri (basis 3) gives the best result obtained at UHF level. In what concerns to the performance of the Hamiltonians, the UHF technique allows the best numerical agreement with the isotropic constant of hydrogen inside quartz. However, the difference between the isotropic constants of confined and free hydrogen is several times bigger than the one detected experimentally. Obtaining such difference is important because it indicates the change in the confinement of hydrogen's wavefunction. The DFT-GGA technique, also used in the plane wave calculations already presented, provides the closest estimation to that difference.

Certainly, there is space for comments on the setup of the calculations done for calculating the hyperfine constants. The geometries have not been relaxed and therefore an arbitrary amount of strain is artificially introduced in the electron den-

Basis set	Hamiltonian	a_{iso} H α-quartz (mT)	a_{iso} H Free (mT)
Basis 1			
Si 6-21G			
O 6-21G*	UHF	70.01	39.84
H 6-21G			
Basis 2			
Si 6-21G*			
O 6-311G*	UHF	100.97	39.84
H 6-21G			
Basis 3			
Si 6-21G*			
O 6-311G*	UHF	50.62	42.87
double-ζ			
Basis 4			
Si 6-21G			
O 6-21G*	UHF	49.85	42.87
double-ζ			
Basis 5			
Si 6-21G*	UHF	49.93	42.87
O 6-21G*	DFT-GGA9	44.36	42.50
double-ζ			
Experimental		**51.80**	**50.68**

Table 6.1: **Dependence of the calculated isotropic constant on the basis set and technique.**

sity. However, from my point of view, a significant conclusion is that the value of the isotropic hyperfine constants depend critically on the quality of the basis. This affects not only the complex system, but also to the simplest case of the isolated hydrogen atom. This conclusion is also in agreement with the problems found by Blöchl when calculating the hyperfine parameters of the same system, using an augmented plane waves method [17]. Therefore, we cannot trust in the data concerning the calculation of the hyperfine constants to contrast our results with those obtained on the EPR experiments.

Of course, there is still the possibility to compare the structures obtained through the structure relaxations with the predictions concerning the structure of the defect done through the analysis of the EPR experiments, like the models proposed by Weil and co-workers [20]. Also in this case, however, care should be paid to ensure that the level of theory employed is able to describe the system properly.

6.4 Discussion of the results. Sources of error.

The results presented so far, concerning the structural relaxation of the defect and the scan of the adiabatic potential along the c-axis of the crystal, are in good qualitative and quantitative agreement with previous studies. More precisely, the size of the energy barrier is very similar to the value obtained by Tuttle for hydrogen in cristobalite [19], and the preferred site for hydrogen inside α-quartz is essentially as that predicted by Blöchl [17]. This, however, does not guarantee that we are actually describing the real system correctly. My main concern is that hydrogen is extremely light, and therefore should display response to even very weak interactions. Different points that come to my mind as possible sources of error are:

- **The approximated treatment of exchange and correlation in the DFT scheme.** The exchange and correlation interaction are described in DFT by means of functionals of the electron density. Paradoxically, the case of hydrogen becomes then particularly difficult to treat, since it has an electron, and therefore the exchange functional will provide a non-zero contribution to the energy, while in fact the electron in hydrogen is unpaired and the exchange should be exactly zero.

- **Validity of the adiabatic approximation.** The electronic structure calculations are done in the assumption that the nuclear motion is decoupled from that of electrons (adiabatic approximation). The case of hydrogen is once more extreme, since the nuclei is composed of a single proton. Neglecting the quantum nature of hydrogen may be then a too drastic approach.

I will discuss the possible effect of this problems in detail now.

6.4.1 The treatment of exchange interaction in DFT.

The problem of the arbitrary (and non-systematic) contribution to the exchange energy due to the approximated treatment of this type of interaction is in fact a well known problem in computational chemistry [143, 144]. In order to have an estimation of the amount of error introduced by this problem, I have performed a calculation of the single hydrogen atom using the same DFT setup. The size of the unit cell, energy cut-off and k-point sampling ensure convergence on the total energy up to 0.1 meV. The final periodic system reproduces a "gas" of atomic hydrogen with atoms distant 12 Å apart.

As expected, the description of hydrogen atom by DFT-GGA is not so accurate. The formation energy for the atom turns out to be -12.53 eV, to be compared with the well known experimental value of -13.6 eV [124]. The single atom contains one electron and therefore the electron-electron interaction should not exist at all. However, the contribution from the GGA exchange-correlation functional is +1.92 eV. Furthermore, the dissociation energy of the H_2 dimer, is 2.21 eV, much smaller than its experimental value of 4.52 eV [124]. This proves the fundamental role of exchange in the H_2 bond.

The underestimation of exchange interaction is, with no doubt, one of the main sources of error in these methods, but estimating how it affects the system that we have studied is difficult. State of the art computational chemistry methods offer the possibility to improve the treatment of correlation, and this has proven to also improve the accuracy of weakly bonded systems containing hydrogen [144]. These methods are, however, at a very early stage of development in solid state disciplines. In addition to the low values found for the activation energies, it also should be kept in mind that this is actually an upper bound estimation. The reason is that kinetic energy of nuclei is completely neglected in this approach. This, as I will show in the next section, may be a drastic approximation in the case of hydrogen atom.

6.4.2 Hydrogen as a light particle.

In the structural relaxations presented here, the ion cores are considered as classical particles. The ions are displaced along the direction of the forces that are evaluated once the electron density has been determined. The Schrödinger equation is solved only for the electron system, and therefore only this particles are considered as quantum particles. This approximation may not be appropriate for determining the trajectory of hydrogen. The behavior of such a light atom may need to be justified by means of the quantum mechanics, as the amounts involved in, for instance, vibrational motion are very small.

As a first approach for accounting for these effects, we have explored the vibrational spectra of hydrogen atom inside the potential well constituted by the adiabatic potential. This well has been approximated by fitting a Morse potential to reproduce the potential curve in a region around the minimum. The reason why we have considered this type of potential instead of, for instance, a parabolic type one, is

because the Morse potential can account for the asymmetry around the minimum.

$$U(r) = U_0\{1 - exp\,(-\beta(r - r_e))\}^2 \tag{6.1}$$

The different energy levels associated to the Morse potential be calculated as follows [145].

$$G(v) = \omega_e(v + \frac{1}{2}) \tag{6.2}$$

$$\omega_e = \beta\sqrt{\frac{U_0}{\mu}} \tag{6.3}$$

where μ corresponds to the reduced mass. In this case, we assume that the solid presents an infinite mass with respect to that of hydrogen and therefore $\mu = m_H$. The three parameters which are actually fitted are U_0, β and r_e. For the fitted potential, these values are $U_0 = 0.290984$ eV, $\beta = 0.630729$ Å^{-1} and $r_e = 0.815781$Å. There is only one bounded state with energy lower than the barrier. This level has an energy 0.16 eV above the minimum of the potential curve. The results are shown in figure 6.8. The adiabatic potential is shown as a thick solid black line. The fitted Morse potential is shown as a blue dotted line. The bounded level is shown as an horizontal red dashed line.

This result indicates that atomic hydrogen is in practice almost free to diffuse along the oxide, and the activation energy is as low as 0.05 eV. With such a low barrier tunneling of the particle should be very probable. A fact to worry about is that the order of magnitude of this energy is the same as the order of magnitude of the stopping criteria for the relaxation. However, if the results are true, then they are in certain contradiction with the interpretation of the EPR experiments, in which the signal detected below 100 K was associated to hydrogen, and the fading of this signal above the mentioned temperature was linked to the recombination of the impurity. Assuming that hydrogen has such low barriers for diffusion, then the next question to understand is what are the magnitudes of the energies involved in the processes where hydrogen recombines.

6.4.3 Stability of hydrogen in quartz. Formation energies and dimerization.

There is actually two different proposals of mechanisms for the recombination of hydrogen inside Silica: the transfer of one electron from one H to another, to form a H^+/H^- pair, and the formation of the H_2 dimer. The former process does not require both hydrogen atoms to comes very close together, since the electron is lost first to the conduction band of the host. The second, instead, requires the formation of the molecular bond.

On the other hand, Blöchl [17] and Yokozawa and Miyamoto [18] have both studied the stability of isolated charged states of hydrogen inside silica. The study of a

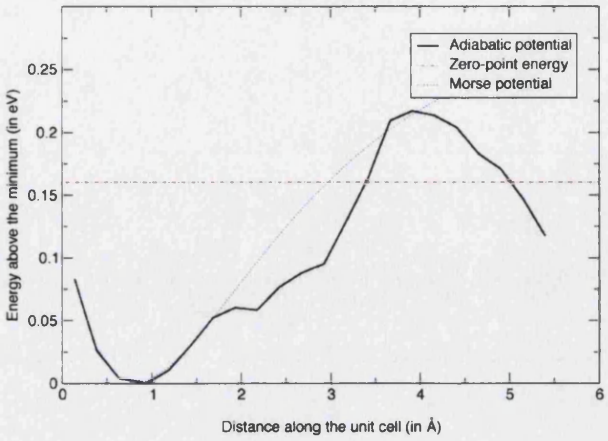


Figure 6.8: **Vibrational states of hydrogen inside quartz.**
Being such a light atom, the amount of energy required for hydrogen vibrational motion may be of the same order of the barrier. We have estimated the vibrational energy by means of a Morse potentials (see text), showed as a dotted line. The potential has been fitted to reproduce the region around the minimum. There is only one bound state, show as a dashed line.

charged system in periodic models, however, requires including a charge neutralizing background to avoid having a system of infinite charge. This background causes an unphysical interaction when the extra charge remains localized, like it seems to be in the case of H^- [13]. In the opposite case, from our experience in the studies on Zirconia and Hafnia (see chapter 5) and in agreement with Pacchioni and co-workers results [34], DFT may be inaccurate when describing the behavior of holes in insulators.

Because of these problems when studying the charged states of hydrogen with periodic Density Functional techniques, I have not calculated the charged states of hydrogen and I will center my attention only on comparing the stability of atomic hydrogen with respect to that of hydrogen dimer inside α-quartz. The incorporation energies for both atomic hydrogen and the H_2 dimer are defined:

$$E_{incorp}(X) = E(SiO_2 + X) - E(SiO_2) - E(X) \tag{6.4}$$

where X stands for either H or H_2, $E(SiO_2+X)$ is the energy of the relaxed supercell containing the impurity, $E(SiO_2)$ is the energy of the supercell without the impurity, and $E(X)$ is the energy of the isolated impurity. In the VASP code, the total energy of the system is defined as the cohesive energy required to separate the system into non-interacting *atoms*, and therefore $E(H) = 0.0$ eV. The energy of the dimer has been calculated ensuring convergency up to 1 meV, and its value is $E(H_2) = -2.21$ eV.

In the case of H_2 inside Silica, I have calculated the most stable geometry for the system using a similar setup as for the calculation of atomic hydrogen's lowest energy site. The molecule orientates as to place its axis parallel to that of the crystal (c-axis), in the same channel where atomic hydrogen finds its lowest energy site.

Following the definition from equation 6.4, the incorporation energy for hydrogen atom in quartz is $E_{incorp}(H) = -0.38$ eV, and that of H_2 is $E_{incorp}(H) = -3.57$ eV. From these numbers it is straightforward to deduce that hydrogen will tend to dimerize when placed inside silica. There is still room for the study of the reaction of formation of the dimer in detail, understanding the energy barriers and how they are influenced by the presence of the crystalline host. This study, however, is complex and has not been addressed here.

6.5 Conclusions.

We have applied plane wave Density Functional structural relaxations in the study of the stability and diffusion of hydrogen in α-quartz. In agreement with previous calculations using similar techniques, we have found that hydrogen remains inert and does not tend to bond to the lattice. However, the ion becomes polarized since the electron charge in the state associated to hydrogen becomes delocalized and therefore the nucleus partially looses the screening of its positive charge.

Density Functional techniques may present the drawback of using an approximate expression for evaluating the exchange interaction. This error becomes especially evident in the case of the hydrogen atom given that it has a single electron and the self-interaction error is important. Up to the degree of accuracy provided by DFT-GGA, the hydrogen potential surface seems to be very smooth with minima of the order of 0.2 eV, in agreement with the study done by Tuttle [19] using quantum molecular dynamics and with the experimental barriers found by Cartier *et al* [133]. This should be considered as an upper estimation, since hydrogen is a light particle and therefore its quantum character may be important. We have estimated that the first vibrational level of hydrogen stands 0.13 eV above the minimum of potential energy. In this case tunneling processes should be very likely. In agreement with the discussion by Edwards [13], the dimerization of hydrogen is favored and therefore hydrogen should recombine rapidly when diffusing. These results are however difficult to match with the experimental evidence that shows that hydrogen is detected as an atomic entity below 100 K.

Another surprising result is the existance of two stable minima for an hydrogen atom inside a single channel. This can be explained if at least one of such minima is due to interaction with the host. However, the shallow nature of the minima found, and the limitations of the DFT technique makes us regard these results with some concern. Unfortunatelly, the cost of the calculations required for a deeper study was too elevated for us to afford.Finally, we have tried to calculate the values of the hyperfine isotropic parameters that can be also determined experimentally. These parameters are very sensitive to the choice of both Hamiltonian and basis sets.

In conclusion, we have tried to determine how helpful standard plane-wave DFT can be in the understanding of the behaviour of atomic hydrogen inside silica. The subtle nature of the interactions in this system places it at the boundaries of the set of systems that can be studied using the DFT techniques.

Chapter 7

Development of an INDO code.

Aim. This chapters presents the work done for upgrading, optimizing and including extra features to the code MOSYM, that performs electronic structure calculations at INDO level.
We also have fitted a set of INDO parameters for the study of systems based on silicon dioxide, silicon nitride and silicon oxynitride. These materials are of great interest for the micro-electronics industry due to their role as gate insulators. The accuracy achieved when studying such systems is tested.
Results. The overall performance of the code has been increased dramatically through a substitution of the diagonalization routines in the self-consistent algorithm. We have added a new correction to the total energy and gradients, based on a library of pair potentials, that allows avoiding problematic situations, like the unphysical attraction between oxygen ions. We have enabled the possibility to define some of the parameters as functions of the charge, so they can be recalculated at each self-consistent cycle.
With the fitted set of parameters, we have successfully simulated the structures of amorphous and crystalline silica, α- and β-silicon nitride and silicon oxynitride. The bond-lengths are reproduced with less that 0.1 Å of difference with experimental data, and most of the angles are given with an error of less that 5°. The electronic structure shows the correct order of the electronic levels, as well as the hybridization, necessary to create the covalent bonds.
Finally, we present examples of applications of the code on the study of systems of many atoms, whose calculation is beyond the possibilities of current first-principles techniques. In particular, the participation functions, that allow one to detect localizes states in systems of many atoms, has been successfully calculated for two models of Silica containing more than 500 ions each.
Techniques employed. Semi-empirical Intermediate Neglect of Differential Overlap (INDO) technique (based in the self-consistent Hartree-Fock scheme).
Basis sets. Localized Slater-type orbitals.
Models. Periodic models of SiO_2 containing up to 648 ions. Periodic models of Si_3N_4 and Si_2N_2O containing up to 224 ions.
Computer Facilities. Single processor workstation.

7.1 Introduction.

The continuous rate of growth of the capability of computers to handle, process and store information has opened the possibility to apply first-principles electronic structure calculation techniques to the study of considerably complex systems. This has not always been the case, as in early days of the discipline the small capacity of computers constrained the range of applicability of *ab initio* codes to very few and very small systems. The semi-empirical schemes were born as simplifications of the first-principles schemes that reduce significantly the cost of the calculations.

The common feature to all semi-empirical techniques is the avoidance of the evaluation of certain expressions, that are substituted by sets of parameters. The values of these parameters are fixed before the calculation, usually in basis of experimental data, or information from accurate first-principles calculations. There is two main consequences of using this type of approach: a) the extra approximation that implies not using the exact calculated values introduces an added inaccuracy on top of that of first-principles results; and b) the results become dependent in the set of parameters used.

In practice, the use of semi-empirical methods requires a careful setup of the calculation, especially in what concerns the choice of the parameters. The parameterization is usually done by selecting a set of reference systems, from which there is experimental or first-principles information available. The reference systems are chosen as to be similar, but easier to calculate, than the system under study. The value of the parameters are fitted as to reproduce as good as possible selected properties of the reference systems. It is then assumed that approximately the same quality of the description will be achieved on the main system. The transferability of the parameters, that is, the range of systems to which they can be applied, is an important issue.

The general trend among many groups of research has been to stick to the use of first-principles methods as basis for their studies. The obvious reason is that the accuracy in first-principles methods can be controlled more systematically, and based on only intrinsic properties of the system (i.e. the number of particles and the spatial distribution of the nuclei). This has relegated most of the semi-empirical techniques to a second plane, where actually their main capabilities are rarely exploited.

Since semi-empirical methods have been designed to be computationally more efficient that first-principles methods by definition, the range of applicability of the former methods will always be wider than that of the later ones. In the particular case of solid state systems, the reproduction of the long range interaction is a main concern. *Ab initio* techniques are constrained to exploit intensively the symmetry of the systems or to use embedded cluster models. Semi-empirical techniques, using the same computing resources, are able handle much bigger models instead, reducing the constraint. As I already mentioned, it is necessary to be careful in what concerns which properties of the system can be reproduced, and with which accuracy, but the semi-empirical methods can be very useful tool for the exploration beyond the limits of the *ab initio*. Some examples will be given inside this chapter.

Unfortunately, the increasing popularity of the first-principles techniques has drifted most of the efforts on the development of computer codes towards their implementation. Most of the programs that are able to perform semi-empirical calculations have become obsolete.

With this perspective in mind, we have established a collaboration project [1] with

[1] Actually, Fujitsu has provided the funding for the elaboration of this Thesis, in the framework of this collaboration.

Fujitsu Labs (Japan), in which we agreed to develop a semi-empirical code that will allow its use as a tool for preliminary analysis of the structure of complex systems. More specifically, the intention was to target materials that could be employed as gate dielectrics, including Silica. A particular feature of such systems is that they display an amorphous structure, very difficult to reproduce with the limited sizes of the *ab-initio* models.

The basis for the project has been the updating of the MOSYM code [2] (formerly known as SYMSYM), which is an implementation of the self-consistent semi-empirical INDO method. We also have included some extra features that will offer more flexibility and the possibility to correct for the error introduced by the approximations. Together with the modification of the program, we have also tried to obtain a set of parameters that will be useful for the study of the dielectric materials just mentioned.

This chapter is particular in the sense that it does not target any specific physical problem, but rather the development of a computational tool and its set up (this includes both the program and the set of parameters). The structure of the chapter is also slightly different from the others presented in this part of the Thesis. Firstly I will describe the MOSYM code and its features, as well as the issues that we decided to address in order to improve its performance and capability. Then the work done on each one of this points and the achievements are presented. Secondly, I discuss in detail the problem of the parameterization, and the strategy followed to find a set of parameters that will allow the description of Silicon dioxide, Silicon Nitride and Silicon Oxynitride. The tests concerning the accuracy that can be achieved using this parameterization, and the transferability of the set are presented next. Finally, I will present an example of potential application of the code, and some preliminary results, for the study of a complex model of Silicon Dioxide containing far more atoms than is currently affordable for a first-principles code.

7.2 Updating and development of the MOSYM code.

7.2.1 Main features of the program.

The MOSYM code is an implementation of the Intermediate Neglect of Differential Orbital (INDO) scheme, which in turn is based on the self-consistent Hartree-Fock theory. Therefore, the INDO technique seeks for the best single-determinant approximation to the ground state wave-function of a multi-electron system. This wave-function is composed of one-electron functions (orbitals), that are defined as linear combinations of atomic orbitals. The self-consistent procedure optimizes the linear coefficients of each one of the orbitals .

One of the most time consuming parts of the first-principles HF calculation is the evaluation of the multi-center integrals that represent the interaction between

[2] Current public version of MOSYM remains under licensing of the University College London (UCL). The code authors are L. Kantorovich, A. Livshits and F. Lopez Gejo.

different electrons, and that are used to construct elements of the Fock matrix. In the INDO technique, the evaluation of the integrals corresponding to electron densities with overlapping close or equal to zero is avoided. Such an approximation, known as zero-differential overlap, is the basis of a whole family of schemes, of which the INDO forms part, developed by Pople and co-workers. The main difference between these methods is the extent to which the approximation is applied. In particular, the INDO scheme is the simplest of the schemes that allows discriminating between states that come from the same electron configuration, by considering the one-center integrals that represent the exchange interaction. A more detailed derivation of the INDO scheme, and a detailed discussion of the physical meaning of the parameters has already been presented in section 2.5. The original derivation of the method can be found in the series of papers by Pople *et al.* [59, 57, 58, 56], or in the book by Pople and Beveridge [54]. The MOSYM code actually implements a modification of the original scheme, suggested by Shluger [60] to account for the extended nature of the nuclei.

The MOSYM code is able to handle three different types of models: isolated clusters or molecules, embedded clusters and periodic systems. In the embedding scheme the quantum cluster is surrounded by a lattice of point charges, that allows one to reproduce the effect of the crystalline field over the cluster. The periodic model makes use of the Large Unit Cell scheme, to allowing sampling of the reciprocal space in the Γ point only [3]. Such a versatile choice of models turns out to be very useful in the stage of parameterization. For instance, it is possible to choose molecular systems as reference for the parameterization, and then move towards extended systems easily. In any of the cases, MOSYM is able to recognize and exploit the symmetry of the system, reducing both the storage and the cost of the calculation. The program can perform selective geometry optimizations of the positions of the ions, in any of the types of models.

Some other additional features of interest of the program are the following. MOSYM allows one to control the occupation of the orbitals, and to define the multiplicity of the system. Several tricks can be used for helping and/or allowing the convergency of the self-consistency calculation. More specifically, it is possible to define a shift of the energy of the unoccupied states, to avoid degeneracy, and/or to define a mixing between the Fock matrices of two contiguous steps in the self-consistency calculation. Finally, the code is able to select different work regimes, adapting itself to the existent resources of disk space and memory. The most efficient regime makes extensive use of memory and reduces input/output operations.

The basic structure of MOSYM, with all the features described above, was already defined for the time that this work of upgrading was started. We decided, however,

[3] The reason for this choice is, once more, because of computational efficiency. The calculations in the Γ point avoid the need to use complex numbers. Handling complex arithmetics was a heavy task on the early days of development of the code. The sampling in Γ point is usually enough for studying insulating materials, but is not valid for the study of the metallic systems, where an accurate description of delocalised states is required.

that in order to be able to use it for the study of systems with many atoms the performance should be considerably increased.

In the specific case of the study of oxides, we also found some trouble when optimizing the geometries, due to the appearance of an unphysical attraction between the Oxygen ions, that caused the system to collapse. We have designed an implemented a new correction to the total energy and gradients (that are the basic information for the structural optimization) based on a pair potential scheme. We have also developed a systematic way of defining the shape of the corrective potential based in simple first-principles molecular calculations. The tests (that will be presented here) show that such simple approach solves successfully the problem.

Finally, we considered that in some of the systems that could be interesting to study, such as some point defects inside the oxides, there was a strong charge re-distribution during the self-consistent calculation. We thought that the set of parameters will become more flexible if at least some of them will be defined as functions that could be recalculated at each self-consistent iteration. Unfortunately, we did not have the chance to perform the necessary test completely, but I have included the description of the charge dependent parameters, and how they have been implemented in the code, with the hope that they may become helpful in future research.

All these different issues will now be addressed in detail in the next sections.

7.2.2 Improvement of the performance of the code.

The Hartree-Fock equations are handled in matrix form, and solved through a diagonalization of the Fock matrix that produces the eigenvalues and eigenvectors of the system at each self-consistent cycle. The diagonalization involves multiplications of matrices and therefore scales like N^3, being N the number of basis functions of the system. We identified this step as one of the most time-consuming parts of the calculation.

Despite being a clearly defined problem, the calculation of eigenvalues and eigenfunctions is not trivial. The underlying reason is that the algorithms are based on an iterative scheme (grand strategy) for reaching the diagonal form of a matrix through similarity transformations [146]. Usually, applying such a procedure directly (transforming the matrix until the deviation from the diagonal form is small enough) does not lead to efficient algorithms. Instead, the matrix can be transformed in its tridiagonal form and then factorization methods are used. These techniques (named QL and QR methods) are based on splitting the matrix into sub-matrices of low dimension, whose eigenvalues can be found efficiently. If the tridiagonal form has been calculated correctly, the method scales linearly with the order of the matrix. A detailed derivation of the methods can be found in the references [146, 147].

I have introduced in MOSYM an implementation of the QR and QL techniques based in state of the art linear algebra libraries LAPACK [147] and BLAS [148]. In order to test the efficiency of the new routines in contrast with the old versions, I

Number Atoms	Number Basis functions	Total CPU time OLD	NEW
9	48	6.00	**1.68**
18	96	10.49	**2.82**
36	192	24.92	**7.31**
72	384	196.53	**40.00**
108	576	871.03	**139.76**
162	864	3592.41	**453.40**
243	1296	5955.29	**1380.09**

Table 7.1: **Test of performance of the code.**
The table displays the CPU time (in seconds) required to complete a self-consistent calculation, as a function of the number of basis functions employed. The time in plain letters corresponds to the calculations done with the old version of MOSYM, and the numbers in bold correspond to the time employed by the new version, where the diagonalization routines have been modified. The scf calculation has been converged up to 10^{-9} a.u. on the total energy.

have performed a timing of the different routines for different test systems. All the tests have been performed on a single AMD Athlon processor with a clock frequency of 1009 Hz, a cache of 256 Kb and 512 Mb of RAM memory. The MOSYM code has been compiled using the `g77/gcc 2.96` compilers [4] as provided in the Mandrake 8.2 (2.96-0.76 mdk) distribution of Linux operating system. The size of the arrays inside the code has been defined big enough as to allow the fastest calculation regime even for the biggest system tested (avoiding unnecessary use of hard disk). The timing has been done using `gprof 2.11.92.0.12` [5].

In order to test the code performance and scalability, I have calculated the electronic structure of a set of supercells of different sizes (the system was α-quartz). All the calculations were converged up to 10^{-9} a.u. on the total energy. The results of the timing are presented in table 7.1, where the size of the systems is listed by both number of atoms and number of basis functions (each function is doubly populated). Timing is given as total CPU time [6] of the calculation.

The results of this timing are also displayed in Figure 7.1 for a more intuitive comparison. The graph shows the relationship between the cost of the computation (measured as time) and the number of basis functions considered. The solid black line represents the performance of the code after the modifications mentioned here, and the red line the performance of the previous version. The green dashed line is just the function $y = x$. In any case, it is possible to see how we have managed to increase the calculation rate substantially. This has allowed us to face the calculation of considerably complicated systems as, for instance, the example in section 3.4,

[4] The MOSYM code is entirely written in standard FORTRAN 77. However, the GNU compilers are based in a set of libraries written mostly in C.

[5] `gprof` performs a profile of the program, displaying a timing of the routines and a call tree. The call tree allows to determine the percentage from the total time consumed by a given set of routines (for instance, those involved in diagonalization).

[6] Note that the real (wall) time involved in the calculation is usually considerably larger than the CPU time, since big intervals are required for the input/output operations (i.e. reading and writing on the hard drives and screen display).

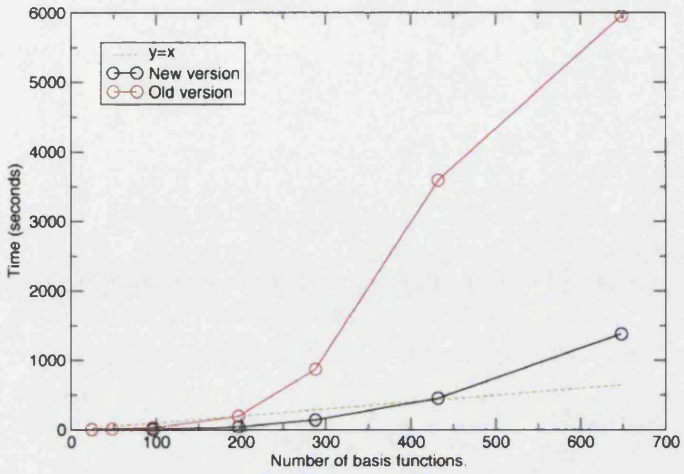


Figure 7.1: **Performance of the code.**
The plot shows the time (in seconds of CPU) employed in a self-consistent calculation (converged up to 10^{-9} a.u. in the total energy) as a function of the number of basis functions employed.

which concerns a model of Silicon Dioxide containing 648 ions. This calculation was performed in a few hours of single processor workstation. A similar *ab initio* calculation, if possible, will require exploiting the full resources of a massive parallel supercomputer.

7.2.3 Implementation of the pair-potential correction.

In the particular case of the systems containing oxygen, an unphysical attraction between the oxygen atoms causes great distortion of the micro-structure and, eventually, its collapse (this problem has been already mentioned by Sokol [149]. This effect was not noticeable in systems where the periodicity kept the system under a meta-stable minimum, but appeared as soon as the system was perturbed (generating a vacancy or displacing an ion far from its equilibrium position, for instance).

The algorithm used in MOSYM seeks for the set of atomic coordinates that minimizes the total energy of the system. The atoms are displaced in the directions indicated by the gradients of the total energy. This information is therefore crucial for driving the structural relaxation towards correct structures. In fact, much more simple non-quantum mechanical models can be used for generating the correct structures and determining many of the mechanical properties of a system accurately [150]. Following this idea, SYMSYM already implemented a simple pair potential correction to the total energy of the system based in the distance between oxygens only (this work was done by M. Szymanski). Presumably, the code should be able to explore the correct potential surface and provide the correct structure of the system. Such a simple procedure, however, did not give any positive result because both the pair potential used was not appropriate and the code followed the incorrect paths of relaxation, since it was using the information provided by uncorrected gradients.

In the new version of MOSYM, we have implemented a more elaborated version

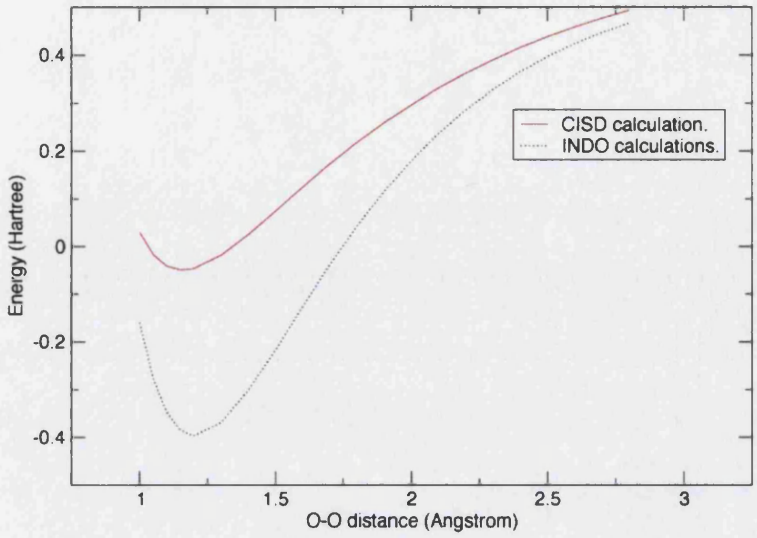


Figure 7.2: **Dissociation curve for O_2 molecule, obtained with INDO and CISD techniques.**

of the pair potential correction. The code is able to use a collection of different potentials, compiled as a library, for any given pair of atomic species. It is possible to choose on input both the functional shape of the pair potential and some parameters of the potential that are left free to be fitted following any desired criteria. Both the total energy and the gradients are corrected before the structural relaxation routine is given the information. Note that, as implemented, the pair potential is applied *as a correction* to the total energy, and it does not provide by itself the correct total energy surface.

The new feature of the code was used for correcting the problem of the attraction between oxygen ions as follows. In first place, we have used the program GAUSSIAN 98 [55] to obtain an accurate representation of the adiabatic curve for the O_2 molecule into two oxygen atoms in triplet state. This calculation was done using a 6-311G* basis set and the CISD technique, therefore taking partially into account the correlation effects that are not considered by construction in the single determinant Hartree-Fock approach and are partially responsible for the incorrect behavior. The same curve was also calculated at INDO level using MOSYM and a reasonable set of parameters.

Results are showed in figure 7.2. Since the comparison of absolute energies does not make sense between results obtained with different techniques, the adiabatic curves for the oxygen molecule have been calculated at each point by subtracting twice the energy of the atomic oxygen in its triplet ground state from the total energy of the molecule. As can be seen in 7.2, the INDO technique predicts a very similar equilibrium distance, but the depth of the potential well is much higher than in the CISD results (0.34 a.u. equivalent to 9.25 eV). It is also remarkable the difference between the shapes of the curves.

Although these results outline the need for correcting the INDO results in order

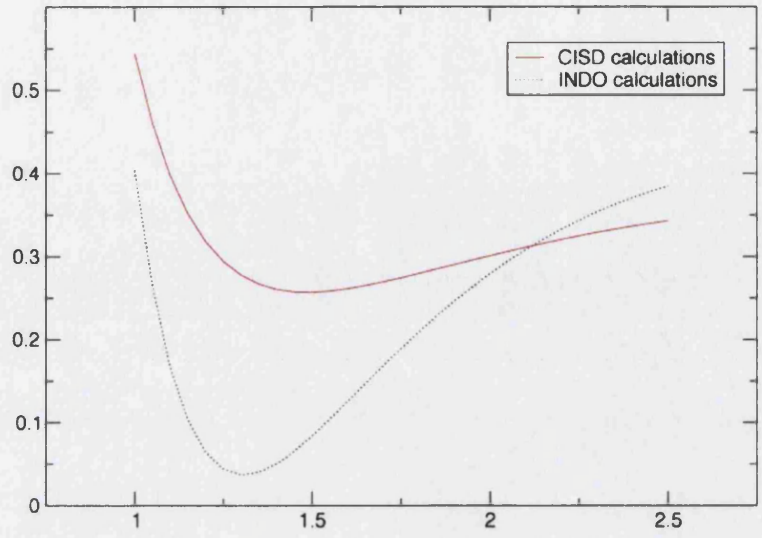


Figure 7.3: **Dissociation curve for O_2^{-2} molecule, obtained with INDO and CISD techniques.**

to obtain a correct representation of the total energy surface, they cannot be applied to correct the parameterizations of oxides. The reason is that the formal charge of Oxygen ions inside Silicon oxide is very different from that displayed in the neutral O_2 molecule. Because of this, the potential well describing the behavior of the latter does not represent that found by the oxygen ions in the oxide.

In silicon oxide, oxygen ions have charges between -1.0 and -2.0. In order to reproduce this situation, the same type of adiabatic curve has been represented, but for the O_2^{-2} molecule being dissociated into two O^- ions. The results are presented in figure 7.3. In this case the potential well predicted by the INDO technique is much more deep. In addition, and more consistently with the process of collapse of the structures during the relaxations, the equilibrium distance between oxygens is smaller by 0.15 Å in comparison with the CISD results. Also the difference in slope between both curves to the right of the minimum is large, leading to much bigger gradients directed towards the minima in the INDO calculations.

The correction has been constructed by assuming that the CISD calculation with the 6-311G* basis set provided accurate results. We have found the difference between the curves displayed in figure 7.3. Results are displayed in figure 7.4. The mentioned difference has been evaluated in all the points marked as green crosses. If this correction is added to the total energies obtained using the INDO technique, the total energies obtained with the CISD method should be obtained. The simplest function that can reproduce correctly this curve within a wide enough range of O-O distances is a polynomial of 7^{th} degree:

$$f(x) = \alpha \sum_{i=0}^{7} c_i (|r - r_o|)^i \tag{7.1}$$

where r corresponds to the distance between pairs of atoms. In this specific case

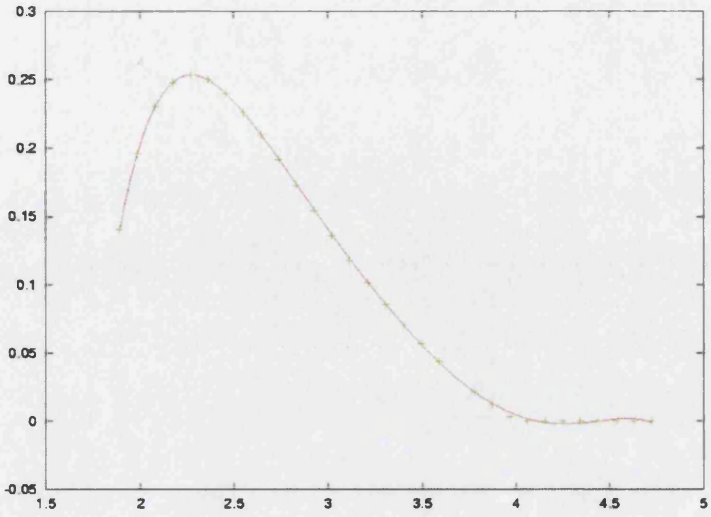


Figure 7.4: **Fitting of the pair potential correction.**

The green crosses show the difference between the dissociation curves of O_2^{-2} calculated at CISD and INDO level. The solid red line shows a polynomial of 7^{th} degree fitted to this data.

the coefficients are $c_0 = -0.001238754$, $c_1 = -0.05579981$, $c_2 = -0.3097622$, $c_3 = -0.5355361$, $c_4 = -0.3416556$, $c_5 = -0.10048775$, $c_6 = -0.008715564$, $c_7 = 0.001035221$, and the cut-off radius $r_0 = 4.72$ a.u. (2.5 Å). Note that the function converges smoothly to zero at this distance. This additional constraint is required to avoid unphysical shifts in the total energy surface. Figure 7.4 also shows this function as solid red line, so the quality of the fitting can be directly seen by comparison.

The factor α has been included to allow a scaling of the whole potential, without essentially modifying its shape. Its inclusion was motivated because of the results obtained with the unscaled pair potential that showed that the Oxygen-Oxygen distance (in α-quartz) was too short by 0.1 Å. The factor $\alpha = 1.1$ was fitted to obtain correct O-O distances in small negatively charged molecules (O_2^{-2}), clusters (Si_3O_6 and Si_3O_4), and α-quartz.

In practice, it will be necessary to supply to the program in the input file an additional line including the atomic species whose distance should be evaluated for applying the pair potential (in this case oxygen-oxygen), the keyword identifying the seventh order polynomial, the cut-off radius and the values of the coefficients in the order listed above. The gradients are corrected by adding the values of the analytical gradients of the polynomial, evaluated along the X, Y and Z directions.

The effect of the pair potential was tested over a supercell of α-quartz, using the set of parameters described in 7.3.2. A structural relaxation using the experimental geometry leads to the same relaxed structure whether the pair potential is used or not. The reason is that in α-quartz the O-O distance is bigger than 2.5 Å, which is the cut-off radius specified for the pair potential, and therefore the correction is not applied. If a perturbation is introduced, making two Oxygen ions come close

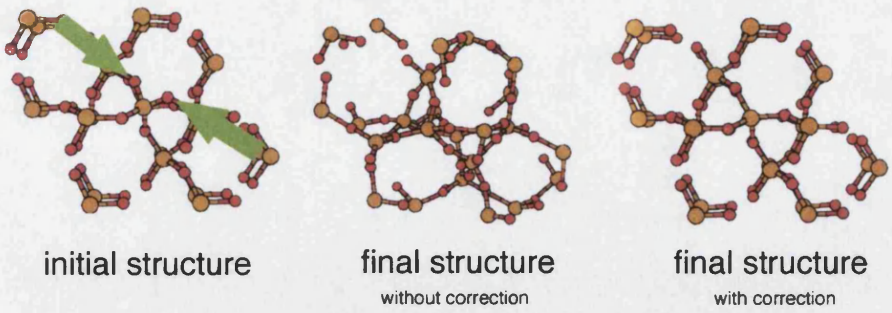


Figure 7.5: **Performance of the pair potential correction.**
The picture shows a supercell of 72 atoms representing the crystalline structure of α-quartz. If a strong perturbation (indicated with green lines) causes two oxygen atoms to come together, the INDO structural relaxation is not able to find the crystalline structure again (bottom left). If, the correction to the total energy and gradients is used, the correct structure is recovered (bottom right).

together (as shown in Figure 7.5), then the results are very dependent on whether the correction is applied or not. If the pair potential is not used, then the Oxygens feel the extra unphysical interaction. This introduces a strong perturbation that eventually causes the whole microstructure to collapse (see Figure 7.5). However, if the correction is applied, the artificial attraction is eliminated and the structure of the crystal is kept.

Managing to obtain the correct microstructure of the system has a strong influence in both the energetics and the electronic structure. As an example, I have performed a structural relaxation of a supercell of amorphous Silicon Oxide (shown in Figure 7.6) generated using classical molecular dynamics. If the pair potential correction is not applied, the relaxed structure differs considerably from the initial one, while the relaxed structure obtained applying the correction is very similar (all three are displayed in Figure 7.6). The relaxation energy involved in the uncorrected calculation is 36 eV, which is surprisingly high. If the correction is applied, the relaxation takes 3.12 eV.

Looking at the electronic structure of the system (also in Figure 7.6), the big relaxation occurring if the correction is not applied leads to a widening of the occupied bands, and therefore the value of the band gap (measured as a difference of one electron energies) reduces from 14.36 eV to 9.80 eV. If the correction is applied the changes in the electronic structure are minimum, and the gap is 14.40 eV.

7.2.4 Charge dependent parameters.

In the first step of the self consistent cycle, the elements of the Fock matrix are obtained by means of a Hückel guess. This initial Fock matrix may be very different from the one obtained at the end of the calculation. In this situation the set of parameters is not flexible enough, and the self consistent cycle is driven to slow convergence or even divergence. On the other hand, in many cases a single set of

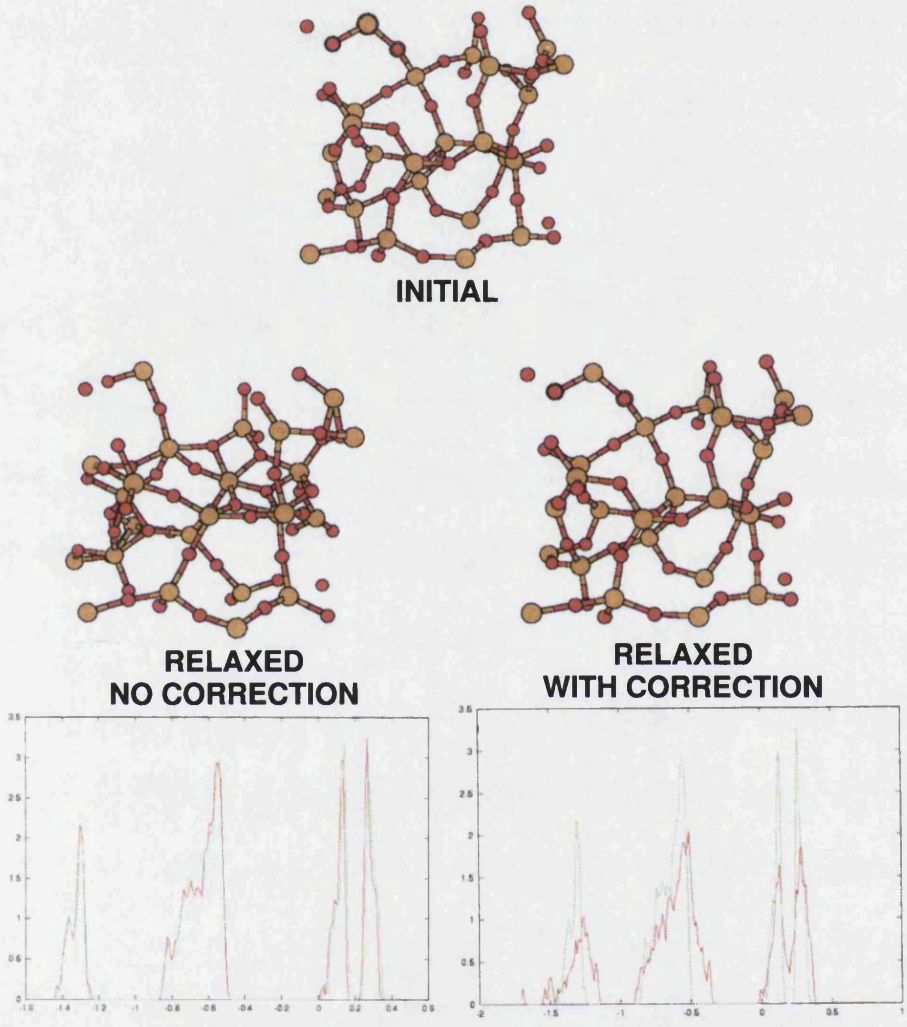


Figure 7.6: **Effect of the correction on the electronic structure.**
The picture shows a supercell of 72 atoms representing an amorphous structure of silica. Once more the relaxation is very dependent on whether the correction is used (bottom right) or not (bottom left). Although the correction is applied to the total energy only, the electronic structure of the system is very affected since it depends on the microstructure of the system. The green plots show the density of states of the initial structure, while the red ones show the Density of states of the final structures.

parameters can also not be enough to describe two atoms that, being from the same species, are present with different charge state in the system. As an example of this, the parameterization that will be shown shortly requires defining two different types of Nitrogen species for obtaining an accurate description of several phases of silicon nitride. This strategy becomes difficult to handle if the systems are more complex, like in the case of an interface where the charge of atoms of a given species changes from one monolayer to the next.

In order to avoid such a problem, I have introduced into the MOSYM code the possibility of using a set of parameters that are defined as functions of the charge. These parameters are recalculated at every cycle of the self-consistent calculation, using the information about the charge state of each ion given by charge population analysis. In the first instance, the only parameters defined in this way have been the electronegativities. The program admits defining them as a second order polynomial of the charge on each ion. I also have implemented a "safety switch" that stops the calculation if the recalculation of the parameters drives the convergence towards an unphysical situation. The program checks at every iteration that the charge of each ion is kept withing some limits that can be defined at input. Finally, another switcher determines when the charge dependent parameters should start to be calculated and used. This allows skipping first iterations of the self-consistent calculation, where the charges may be unreasonable, and affects critically the speed of convergence.

Unfortunately, I only had time for performing very few attempts of fitting a set of charge-dependent parameters, and the results were, so far, not very successful. The strategy followed for the fitting of charge dependent parameters was to build a database of different systems that contained the same atomic species, but in different charge states. For each system, the electronegativity parameter was optimized in order to achieve the best agreement possible with *ab initio* HF calculations. The charge-dependent parameter (which is now a function) was defined as the best fitting to the points given in a plot representing the optimized electronegativity versus the charge population of the ion.

7.3 Development of a parameterization.

7.3.1 The target systems. Gate dielectrics.

The original targets of the collaboration project with Fujitsu Labs have been silica systems and possible substitutes for this material as a gate dielectric. The list of such candidates changes rapidly and is determined by the trends in the industry. At the time of definition of the project, silicon nitride seemed to be an interesting candidate. In fact, this material has been considered for a long time as a potential substitute [10] as a gate dielectric.

In brief, silicon nitride presents a higher dielectric constant (7.8 [151]) than that of silicon dioxide, and therefore it allows the use of a thicker gate dielectric layer without decreasing device performance. Thicker gate layers avoid the problems related to

tunneling of carriers and dielectric breakdown. In addition, silicon nitride displays higher resistance than silicon dioxide to the diffusion of impurities [152]. The main advantage of this material over many high-κ materials is that it is easy to generate over the silicon substrate. A big drawback on the application of silicon nitride as gate oxide has been a high level of defects found, both in the interface with silicon and in the bulk [10]. Some improvement is possible if the nitride is combined with oxygen, to form silicon oxynitride. In this case the flexibility of the network increases, allowing one to reduce the density of defects, at the price of lowering the dielectric constant.

Silicon oxynitride may display very different stoichiometries, and it is therefore important to understand how the electrical properties of the system change as a function of the composition of the material. Since using the oxynitride is intended to reduce the number of defect sites, it is necessary to evaluate the energies of formation of these defects. The task of evaluating all these properties for many different compositions is considerably demanding if first-principles methods are to be used. This is especially true if realistic models are to be used, since these materials display an amorphous phase. We therefore decided to develop a set of parameters that will allow the study at INDO level, as a preliminary stage. The aim of the parameterization was to find a set that will allow the study of models of silicon dioxide, silicon nitride and silicon oxynitride, and of simple defects inside these materials.

7.3.2 The set of parameters.

The INDO parameterization requires defining several types of one-center and two-center parameters. The former ones basically determine the interaction between electrons and nuclei of an specific atomic species, while the second one rules the interaction between electrons of different atomic species. Therefore, each atomic species can be characterized by providing the following parameters, for each occupied orbital i (only valence electrons are considered): the exponent of the orbital, ξ^i, the electronegativity, $E_{neg}(i)$, the interaction between electron densities, β^i, and the strength of the interaction between the electron density and the nucleus, α. A detailed description of the parameters, and its physical meaning has already been given in section 2.5.

The set of parameters that has been fitted during this study is shown in Tables 7.2, 7.3 and 7.4. Many of these parameters are modifications of those fitted and employed previously by A. Sokol, and appear listed also in[149], together with details of the fitting process and results of his tests. This set with minor modifications came out as best possible choice after a long process of selection in which we also tried to develop new sets. The main difference between the sets employed by Sokol and the one shown here is the use of non-diagonal β elements that have been calculated explicitly, as opposed to the conventional INDO parameterization where the non diagonal β are calculated automatically by finding the average between the diagonal elements ($\beta_{ij} = \frac{\beta_{ii}+\beta_{jj}}{2}$). This allows obtaining extra control over the description

Element	Core Charge	Shell	ζ (a.u.[1])	E_{neg}(eV)	$P^{(0)}$(e)	β(eV)
H	+1	$1s$	1.150	5.783	0.800	6.940
O	+6	$2s$	2.270	4.500	1.974	$-16.0^{\dagger}$
		$2p$	1.860	-12.000	1.960	$-16.0^{\dagger}$
Si	+4	$3s$	1.559	42.100	0.000	$-8.0^{\dagger}$
		$3p$	1.834	35.100	0.000	$-24.0^{\dagger}$
N	+5	$2s$	2.200	-4.700	1.974	-26.000
		$2p$	1.900	-18.000	1.667	-28.000

Table 7.2: **Set of one-centre INDO parameters.**
The β parameters are actually matrices, whose diagonal elements are listed here. The rest of the elements of the matrices are calculated automatically, except for the elements marked with † are defined explicitly in table 7.4.

Core type	H($1s$)	Si($3s, 3p$)	O($2s, 2p$)	N1($2s, 2p$)	N2($2s, 2p$)
H	0.000	0.001	0.490	0.490	0.490
Si	0.000	0.310	0.245	0.000	0.100
O	0.000	0.050	0.100	0.100	0.100
N1	0.000	0.030	0.100	0.100	0.100
N2	0.000	0.000	0.100	0.100	0.100

Table 7.3: **Set of two-centre INDO parameters.**
See text for more details about the double set of parameters for Nitrogen (N1 and N2).

of the geometry of silica polymers. Also, a set of parameters for Nitrogen has been fitted from scratch. In this case the peculiarity of the set is the use of two different types of α parameters (named N1 and N2 in table 7.3). This was introduce in order to obtain correct descriptions of the geometries of two different crystalline phases of Silicon Nitride.

7.3.3 Performance of the set of parameters.

We first tested the accuracy of the set of parameters just presented by observing the bulk properties of several dielectric materials. Some examples concerning amorphous and crystalline Silica have already been shown in Section 7.2.3. Given that the set of parameters concerning Nitrogen is completely new, I will focus now in the Nitrogen containing compounds.

The structure of Silicon Nitride (Si_3N_4) has been characterized by several groups, because of its application as a ceramic material. We have taken as a reference the data

	Si(s)	Si(p)		O(s)	O(p)		Si(s)	O(p)
Si(s)	-8.0	-16.0	O (s)	-16.0	-16.0	Si (s)	-20.0	-20.0
Si(p)	-8.0	-24.0	O(p)	-16.0	-16.0	O(p)	-14.0	-20.0

Table 7.4: **Set of β parameters.**

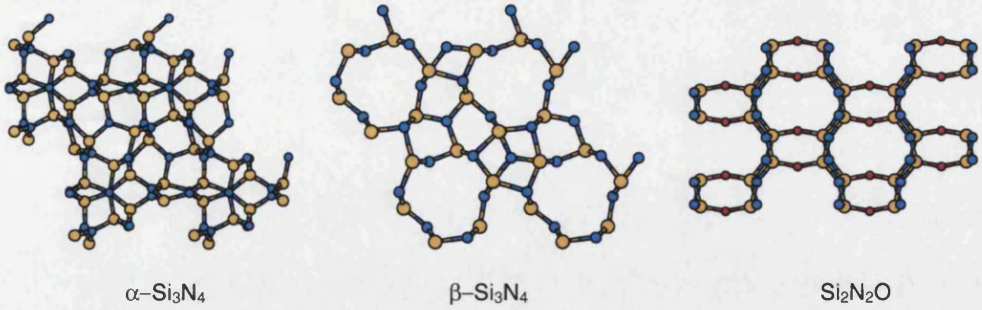


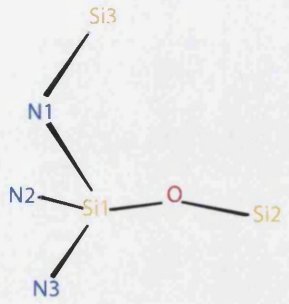
Figure 7.7: **Schematic representation of α-Si_3N_4, β-Si_3N_4 and Si_2N_2O.** The Yellow spheres represent Silicon ions, the blue spheres represent Nitrogen ions, and the red spheres represent Oxygen ions. The structures are shown from a plane perpendicular to the c-axis.

from X-ray powder diffraction experiments obtained by Yand *et al.* Polycrystalline Silicon Nitride displays two different phases (α and β). Both phases are based on tetrahedral SiN_4 units. Silicon Oxynitride (Si_2N_2O) can be considered as a hybrid between SiO_2 and Si_3N_4. Its structure is based on SiN_3O units, and can be seen as planes of SiN_3 units linked by the Oxygens. A graphical representation of all these materials, α- and β-Si_3N_4 and Si_2N_2O, is shown in Figure 7.7. We also have compared our results with those obtained from first-principles calculations (at DFT-LDA level) performed by Xu and Ching [153].

We have constructed periodic models following the Large Unit Cell approach. In all cases, the supercells were generated using a $2 \times 2 \times 2$ (diagonal) extension matrix. The supercells of both phases of Silicon Nitride contained 224 each, and that of Silicon Oxynitride contained 160 atoms. For each material, we have performed a geometry optimization using as initial geometry that predicted by experiments, keeping the lattice vectors fixed to the experimental values. No symmetry constraint was imposed over any of the atoms. The self-consistent calculations were converged up to 10^{-9} atomic units in the total energy of the supercell. The criteria for stopping the structural optimization was also not to have changes in the total energy above 10^{-9} atomic units.

The results concerning the micro-structure of the materials have been summarized in Figure 7.8. In general terms, all the bonding distances are reproduced within less than 0.1Å. We have noticed, that the INDO calculation forces the system to have a homogeneous bond-length distribution, so all Si-N bonds tend to have similar values. It should be noticed, however, that the experimental data comes from diffraction experiments, where the data is fitted assuming certain symmetry of the crystal. As a result, the coordinates of the atoms related by symmetry are not independent (and neither the bond-lengths) in opposition to our calculations where the coordinates of each ion have been relaxed independently.

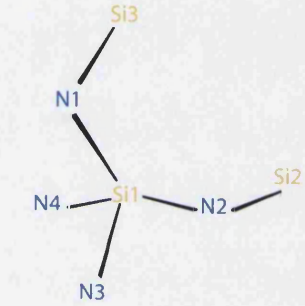
In what concerns the angles, the level of agreement depends considerably on the symmetry of the system. In Silicon Oxynitride, all the N-Si-N and N-Si-O angles



Si_2N_2O

distances (Å)	Calc.	Exp.
N1-Si	1.74	1.72
N2-Si	1.73	1.75
N3-Si	1.74	1.69
O-Si	1.61	1.62

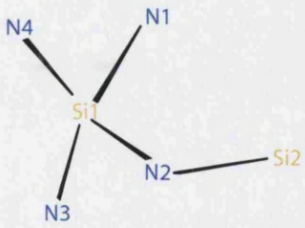
angles (°)	Calc.	Exp.
N1-Si-N2	108.74	111.01
N1-Si-N3	109.78	110.21
N1-Si-O	105.47	107.12
N2-Si-N3	111.40	106.43
N2-Si-O	113.25	113.31
N3-Si-O	108.01	108.74
Si1-O-S2	157.16	145.87
Si1-N1-S3	124.49	123.71



β-Si_3N_4

distances (Å)	Calc.	Exp.
N1-Si	1.67	1.79
N2-Si	1.70	1.73
N3-Si	1.67	1.71
N4-Si	1.67	1.71

angles (°)	Calc.	Exp.
N1-Si-N2	106.10	107.99
N1-Si-N3	108.30	105.13
N1-Si-N4	108.30	105.13
N2-Si-N3	109.70	110.57
N2-Si-N4	109.70	110.57
N3-Si-N4	114.39	116.82
Si1-N2-S2	116.33	122.35
Si1-N1-S3	120.01	120.00



α-Si_3N_4

distances (Å)	Calc.	Exp.
N1-Si	1.67	1.76
N2-Si	1.74	1.69
N3-Si	1.73	1.73
N4-Si	1.74	1.78

angles (°)	Calc.	Exp.
N1-Si-N2	104.0 5	118.00
N1-Si-N3	117.40	109.07
N1-Si-N4	108.71	92.07
N2-Si-N3	104.79	115.17
N2-Si-N4	113.05	117.38
N3-Si-N4	108.34	108.58
Si1-N2-S2	119.26	109.32

Figure 7.8: **Comparison between INDO results and experimental data.**
The schemes represent the local structure inside Si_2N_2O β- and α-Si_3N_4 (from top to bottom). All the geometry optimizations have been performed without any geometry constrain. The experimental data has been taken from [154] and [153] for β- and α-Si_3N_4, and Si_2N_2O respectively.

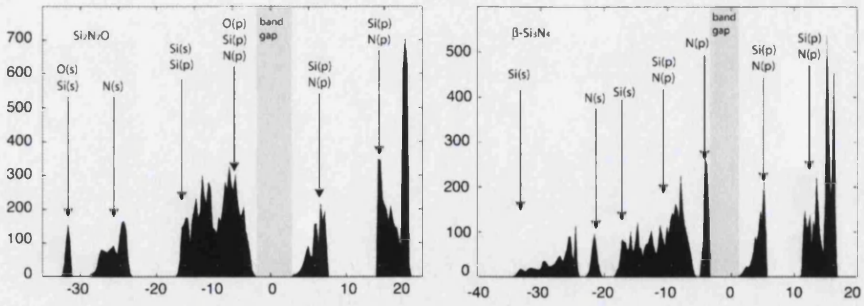


Figure 7.9: **Electronic structure of β-Si_3N_4, and Si_2N_2O.**
The arrows indicate the biggest contributions for each band. All energies are in eV. The density of states is given as states/eV cell.

match within less than 5°. The Si-O-Si is the one in biggest disagreement. It is well known, from the studies performed in Silicon-Oxygen compounds[83, 155], that the total energy of the system changes very little as the Si-O-Si angle is modified[7]. Such subtle changes in the energy are difficult to reproduce with the calculations at INDO level. However, the pair potential correction introduced accounts successfully for it, as it could be seen in the examples in Section 7.2.3. In the case of pure Silicon Nitride compounds, the angles are much better reproduced in the case of β-Si_3N_4, where agreement is within 5°.The α phase has a much more complex structure, and the parameters do not show enough flexibility.

Turning attention towards the electronic structure of these systems, Figure 7.9 shows the Density of States of Silicon Oxynitride and the β phase of Silicon Nitride (the DOS of both Nitrides are very similar). The composition of the DOS (also shown in the Figure) is qualitatively correct, and shows the different hybridizations between the orbital of different atomic species, which results in the covalent bonds that are characteristic of these materials. If we compare our results with those by Xu and Ching [153], obtained by means of a first-principles technique, the only two main differences are: a) that the s levels of Nitrogen appear shifted up in energy with respect to the s levels of Silicon in Silicon Nitride, and that the conduction band in the INDO results appears divided in two regions, separated by a wide gap. This result is unphysical, and is actually due to the absence of d functions in the basis set of Si.

In conclusion, the set of parameters presented here allows a reasonable description of the systems. They provide a qualitatively correct description of the electronic structure of the materials, and therefore the bonding scheme is described correctly. As a consequence, also the micro-structure of the systems is described correctly, with an accuracy of 0.1 Åin the bond-lengths and 5° in the angles. The parameters, however, lack flexibility and fail to describe systems with complicated and/or asymmetric structures, if no external help like a pair potential correction is used. Unfortunately,

[7]This justifies the fact that there is so many Silica polymorphs based in SiO_4 units, that can be arranged in many different ways thanks to the flexibility provided by the Si-O-Si links.

this statement becomes critical in complex systems that contain defects. We have not succeeded, so far, in applying the INDO approach to the study of these systems. Still, we find that the set of parameters, *as it is*, can be of help for certain type of studies, as the one presented next.

As an example of this type of application, I have performed a calculation of the participation functions for a crystalline system (α-quartz) and amorphous silica, using cells containing 576 and 648 atoms respectively. The amorphous structure has been generated by means of a classical molecular dynamics technique. The results are presented in figure 7.10, together with a representation of the structures that allows one to appreciate the difference in symmetry. The graphs display both the density of states of the system (solid line) and the participation function (crosses) for both systems. It is possible to see how the edges of the bands in the amorphous system are composed of states that are not completely delocalized, while the density of states remains essentially unaltered. This is consistent with the predictions by Ching [65]. In this particular example of an amorphous system, there is no unusual localized states. This is not surprising since the system also does not have any particular structural defect (i.e. vacancies or low coordinated atoms). This simple method turns to be really useful for the identification of the region containing a localized state inside a much bigger system. In a further stage, such a region can be modeled more accurately. An example of this kind of work has been presented by Sushko and co-workers in the case of defects in surfaces of MgO [156].

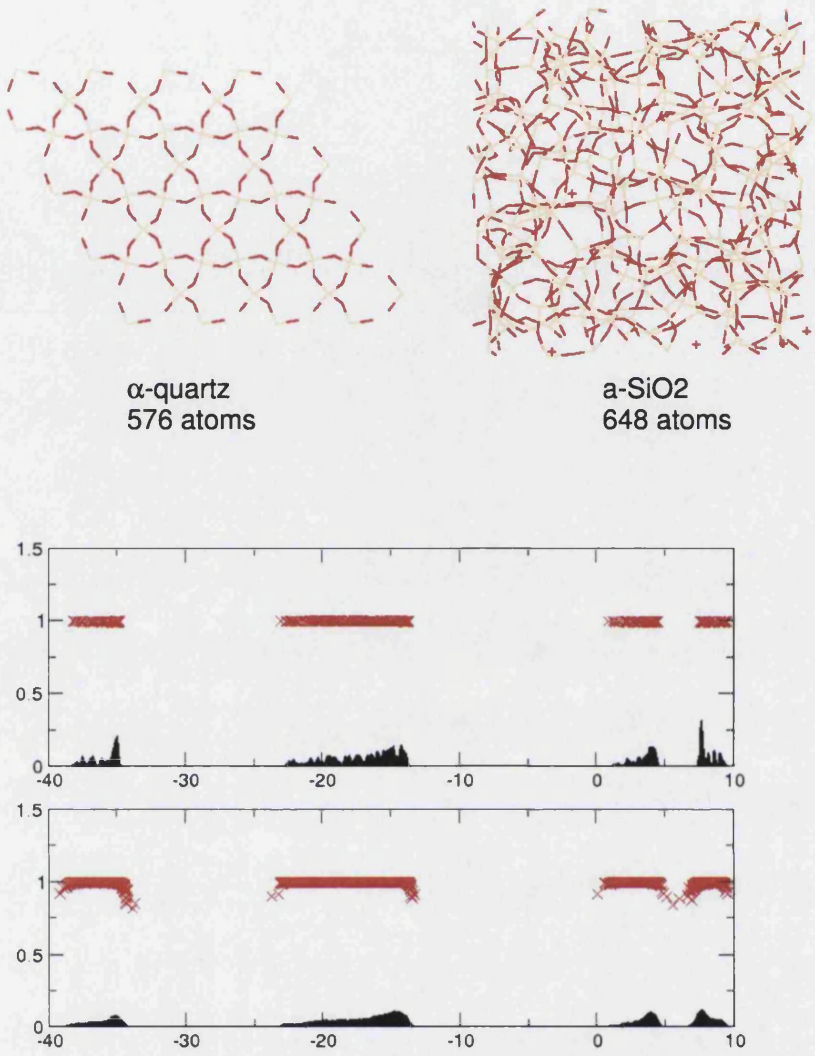


Figure 7.10: **Participation functions for α-quartz (top graph) and amorphous silica (bottom graph).**
The red crosses represent the values of the participation function for each one of the eigenvalues of the system. The function range goes from 1 (completely delocalized) to 0 (completely localized). The normalized density of states is also projected in black under the function, and allows one to see how in amorphous systems the states on the edges tent to be localized. A schematic representation of the unit cells used in the calculation are also shown on top. All energies are given in eV.

7.4 Conclusions.

The work presented in this chapter concerns two different main subjects a) the update and addition of new features to the code MOSYM, that performs semi-empirical LCAO electronic structure calculations at INDO level, and b) the development and testing of a parameterization for the study of systems formed of silicon dioxide, silicon nitride and silicon oxynitride.

The upgrading of the code consisted mainly in a substitution of the diagonalization routines, which are employed during the self-consistent solution of the equations of the INDO method. The increase in the speed of the computations was significant.

The addition of the new features was focused on two different points: a correction for the total energy and gradients based on a pair potential scheme and the modification of some of the INDO parameters, redefining them as functions of the charge of the ions, instead of keeping them constant along the calculation. The pair potential correction was tested against a system of silicon dioxide, in which an unphysical attraction between the oxygen ions caused the structure of the system to collapse. The problem was successfully solved by the pair potential correction. We showed how the fitting of the potential can be deduced directly as the difference between two adiabatic curves as calculated at INDO level and with a higher order method. The correction was fitted using molecular systems as a reference, and successfully transferred towards extended systems. The modifications of the code that enable the use of charge dependent parameters have been included, but further tests will be required for understanding and validating the new parameterization.

In what concerns the sets of parameters generated for the description of compounds containing Si, O and N, our tests show that crystalline and even amorphous materials are described with an accuracy of 0.1Å in bond-lengths and 5° in the angles. The transferability of the set, however, was limited to cases in which there is no strong redistribution of the charge density during the self-consistent calculation. Examples of systems where this happens are some point defects (like vacancies) and also some interfaces.

The ability to deal with systems of many atoms with this particular code and method has been shown in an example in which a cell containing 648 ions have been employed. The calculation was done in a single processor machine, in a few hours of CPU time. The equivalent calculation using standard *ab initio* techniques is currently unaffordable, even if supercomputing facilities are available. Therefore, semi-empirical methods could be considered as useful tools for the qualitative analysis of systems beyond the first-principle methods. At present, however, new methods allow one to maintain the *ab initio* approach, with a remarkable low cost of the calculations (for instance, the SIESTA code, at the time of publication of this work, was able to deal with systems of the size of those compared here in a very similar time scale, using a DFT approach [157]). These new methods may be then preferred to the approach presented here.

Chapter 8

Final conclusions.

The research presented in this Thesis intended to explore the limits of applicability of some of the most widely employed electronic structure calculation techniques, for the modelling and predictive simulation of some dielectric materials. These materials are employed as gate dielectrics in the MOS structures, and the evolution of their physical and chemical properties as the devices are scaled down is a main concern for the developers of microelectronics.

We have choosen some problems that, either have been in the scene for many years without getting any final answer, or are becoming very important for the near future of the technology. The former group includes the presence of hydrogen inside silicon dioxide, and the problem of band alignment at the Si/SiO_2 interface. The latter includes the study of the properties of point defects in zirconia and hafnia, which are being considered as possible substitutes of silica as gate dielectric. The conclusions achieved on each case are the following:

- **Band alignment at the Si/SiO_2 interface.** Usual band alignment techniques use information corresponding to bulk calculations of the materials that constitute the junction. They require the (non-trivial) definition of an absolute reference level whose position should be determined inside the band structure of each material. Instead, we have employed a model of a true interface, and studied directly the evolution of the bands accross it. This avoids the need of the reference level, and also allows exploring the dependence of the band alignment on particular features of the interface (thickness of the oxide, presence of defects, roughness). We studied how different parameters of the calculation affect the results: the B3LYP Hamiltonian provided the best accuracy, and the basis sets should contain polarisation functions. The thickness of the oxide layer of some of the models studied is close to that of some of the thinnest devices. It can be seen that there is a region inside the oxide, around 4Å thick, which contains interface states. The dimensions of this region do not change as the thickness of the oxide layer changes. Since carriers can be able to tunnel from these states, this possibly establishes a limit for the minimum thickness of the oxide.

- **Defects in hafnia and zirconia.** Any alternative to silica as a gate dielectric should first prove to be reliable. Part of this reliability is determined by how easily certain types of defects can be created and by whether these defects can trap charge carriers or not. Hafnia and zirconia are among the possible substitutes of silica. We have performed a systematic study of the charge trapping ability of several types of simple defects, using a density functional technique (GGA) and plane wave basis set. The main drawback of this approach is the underestimation of the band gap caused by the inaccurate treatment of the unoccupied states. We applied an empirical correction to the gap width, in order to be able to express the electron affinities, ionization potentials and formation energies as ground state properties. Oxygen vacancies can be created at much lower cost in energy than the cation vacancies, and are able to act as traps when positibely charged. In a MOS system, they should be able to capture electrons coming from the conduction band of the silicon layer. Some complex pairs like impurities near the vacancy do not seem to affect significantly this behavior. In conclusion, the charge trapping ability of even simple defects in these materials should be taken into account when considering their use as gate dielectrics.

- **H inside α-quartz.** Understanding the role played by hydrogen inside the MOS systems has been an issue as old as microelectronics itself. Several studies have addressed different aspects of the problem. Employing the same DFT-GGA approach as for the study of defects inside hafnia and zirconia, we explored the stability and diffusion of atomic neutral hydrogen inside quartz. Hydrogen does not seem to establish any chemical bond with the host, although the electron becomes more delocalized than when H is isolated. This causes the proton to lose some of the screening and therefore an electrostatic interaction between the host and the impurity appears. The minimum found are, however, very shallow (0.2 eV). Because of this, some approximations adopted in the method may constitute big sources of inaccuracy. One of these sources is the approximate treatment of the exchange and correlation energies. Another is the classical treatment of the hydrogen nuclei, which actually is a very light particle. An estimation of the zero-point energy indicated that the depth of the minima found by hydrogen is even smaller than that predicted by the electronic structure calculation. Finally, we attempted to calculate the isotropic constants for direct comparison with EPR experiments, but the results displayed a very strong dependence on the basis set. Given the accuracy achieved with these calculations, we concluded that this problem is actually on the edge of the range of applicability of conventional DFT schemes.

There is an additional part of the research, that has focused on the development of methodology rather than in facing a particular problem. We have updated a semi-empirical code, in order to apply it to the study of systems with big number

of atoms. The tool can be used for a preliminary analysis of a complex structure, before a more detailed and accurate study of a particular region. We have been able to handle systems as big as 648 atoms in single processor workstations. We fitted a parameterisation that allowed us to reproduce the crystalline structures of several phases of silicon dioxide, silicon nitride and silicon nitroxide with errors of 0.1Å in the the bond-lengths and 5° in the angles, with respect to experimental results. We did not succeed, however, on porting this parameterisation towards compelx systems with strong charge redistributions like some defects or interfaces.

In overall conclusion, computational disciplines have been largely employed in the research vinculated to the progress of microelectronics. The miniaturisation of the devices and the progress of the computational disciplines has placed us in the edge of the era in which the direct simulation of real systems will be possible, as it is now in many areas of chemistry. Meanwhile, modelling of the systems constitutes a fundamental step. And in any case, a careful analysis of the limitations of the techniques is needed, if the predictive capability of the tools is to be preserved.

Bibliography

[1] G. E. Moore. *Int. Elec. Devices Mtg (IEDM) Technical Digest.* IEEE, Piscataway, New Jersey, 1975.

[2] International technology roadmap for semiconductors. Technical report, The Semiconductor Industry Association., 1999-2000. Updated versions available online at http://public.itrs.net/.

[3] M.L. Green, E. P. Gusev, R. Degraeve, and E. L. Garfunkel. Ultrathin (<4nm) SiO_2 and Si-O-N gate dielectric layers for silicon microelectronics: understanding the processing, structure, and physical and electrical limits. *J. Appl. Phys.*, 90(5):2058–2110, 2001.

[4] M. Städelle, B. R. Tuttle, and K. Hess. Tunneling through iltrathin SiO_2 gate oxides from microscopic models. *J. Appl. Phys.*, 89(1):348–363, 2001.

[5] D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp. The electronic structure of the atomic scale of ultrathin gate oxides. *Nature*, 399:758–761, 1999.

[6] T. Yamasaki, C. Kaneta, T. Uchiyama, T. Uda, and K. Terakura. Electronic structures of SiO_2/Si(001) interfaces. In H. Z. Massoud, I. J. R. Baumvol, M. Hirose, and E. H. Poindexter, editors, *The physics and chemistry of SiO_2 and the Si-SiO_2 interface-4.*, volume 2000-2, page 295, New Jersey, 2000. The Electrochemical Society, Pennington.

[7] J.R. Hook. *Solid State Physics.* The Manchester Physics Series. Willey, Chichester, New York, 2nd edition, 1991.

[8] S. Tang, R. M. Wallace, A. Seabaugh, and D. King-Smith. Evaluating the minimum thickness of gate oxide on silicon using first-principles method. *Appl. Surf. Sci.*, 135:137, 1998.

[9] D. A. Buchanan. Scaling gate dielectric: Materials, integration, and reliability. *IBM J. Res. Develop.*, 43(3):245–260, 1999.

[10] P. Balk, editor. *The Si-SiO_2 system.*, volume 32 of *Materials Science Monographs.* Elsevier Science Publishers, New Y,ork, 1988.

[11] A. I. Kingon, J. P. Maria, and S. K. Streiffer. Alternative dielectrics to silicon dioxide for memory and logic devices. *Nature*, 406:1032–1038, 2000.

[12] J. Robertson. Band offsets of wide-band-gap oxides and implications for future electronic devices. *J. Vac. Sc. Technol. B*, 18(3):1785–1791, 2000.

[13] A.H. Edwards, W. M. Shedd, and R. D. Pugh. Theory of H^- in SiO_2. *L. Non-crystalline Solids*, 289(1-3):42–52, 2001.

[14] A. H. Edwards. Dissociation of H_2 at silicon dangling orbitals in a-SiO_2: a quantum mechanical treatment of nuclear motion. *J. Non-Crystalline Sol.*, 187:232–243, 1995.

[15] G. Pacchioni and M. Vitiello. EPR and ir spectral properties of hydrogen-associated bulk and surface defects in SiO_2: *Ab initio* calculations. *Phys. Rev. B*, 58(12):7745–7752, 1998.

[16] G. Pacchioni and R. Ferrerio. Optical transitions and EPR properties of two-coordinated Si, Ge, Sn and related H(i), H(ii), and H(iii) centers in pure and doped silica from *ab initio* calculations. *Phys. Rev B*, 58(10):6090–6096, 1998.

[17] P. E. Blochl. First-principles calculations of defects in oxygen-deficient silica exposed to hydrogen. *Phys. Rev. B*, 62(10):6158, 2000.

[18] A. Yokozawa. First-principles calculations for charged states of hydrogen atoms in SiO_2. *Phys. Rev. B*, 55(20):13783, 1997.

[19] B. Tuttle. Energetics and diffusion of hydrogen in SiO_2. *Phys. Rev. B.*, 61(7):4417, 2000.

[20] S. Goshen, M. Friedman, R. Thieberger, and J. Weil. Models for the hydrogen atom confined within crystalline quartz. *J. Chem. Phys.*, 79(9):4363–4366, 1983.

[21] W. Y. Ching. Theory of amorphous SiO_2 and SiO_x. I. atomic structural models. *Phys. Rev. B*, 26(12):6610, December 1982.

[22] W. Y. Ching. Theory of amorphous SiO_2 and SiO_x. II. electron states in an intrinsic glass. *Phys. Rev. B*, 26(12):6622, December 1982.

[23] W. Y. Ching. Theory of amorphous SiO_2 and SiO_x. III. electronic structures of SiO_x. *Phys. Rev. B*, 26(12):6633, December 1982.

[24] M. A. Szimanski, A. L. Shluger, and A. M. Stoneham. Role of disorder in incorporation energies of oxygen atoms in amorphous silica. *Phys. Rev. B*, 63:224207, 2001.

[25] M. Springborg. *Methods of electronic-structure calculations.*, chapter 8.2, pages 75–79. Wiley Series in Theoretical Chemistry. John Willey & sons, Ltd., Baffins Lane, Chichester, West Sussex PO 19 1UD, England., 1st edition, 2000.

[26] K. Ohno, K. Esfarjani, and Y. Kawazoe. *Computational materials science.*, chapter 2.6, pages 100–105. Solid-State sciences. Springer-Verlag, Berlin, Heidelberg, 1st edition, 1999.

[27] C. Pisani. *Quantum-mechanical ab-initio calculation of the properties of crystalline materials*, chapter 3, Section 4.2. Number 67 in Lecture Notes in Chemistry. Springer-Verlag, Berlin-Heidelberg, 1st edition. edition, 1996.

[28] J. Muscat, A. Wander, and N. M. Harrison. On the prediction of band gaps from hybrid functional theory. *Chem. Phys. Lett.*, 342:397, 2001.

[29] A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry.* Dover Publications, Inc., 31 East 2nd Street, Mineola, N. Y. 11501, 1st edition, 1996. First published by McGraw-Hill, 1989.

[30] V. R. Saunders, R. Dovesi, C. Roetti, M. Causa, N. M. Harrison, R. Orlando, and C. M. Zicovich-Wilson. *CRYSTAL 98 1.0 User's Manual.*, March 1999.

[31] R. G. Parr and W. Yang. *Density functional theory of atoms and molecules.* Number 16 in International series of monographs on chemistry. Oxford University Press, Oxford, Clarendon, 1989.

[32] G. Kresse and J. Furthmüller. *VASP the GUIDE.* Institut für Materialphysik, Universität Wien, Sensengasse 8, A-1130 Wien, Austria, 4.4.5 edition, 2001. Also online at http://cms.mpi.univie.ac.at/vasp/.

[33] A. D. Becke. A new mixing of Hartree-Fock and local density functional theories. *J. Chem. Phys.*, 98(2):1372–1277, 1993.

[34] G. Pacchioni, F. Frigoli, and D. Ricci. Theoretical description of hole localization in α-quartz Al center: The importance of exact electron exchange. *Phys. Rev. B*, 63:054102, 2000.

[35] A. L. Shluger, P. V. Sushko, and L. N. Kantorovich. Spectroscopy of low-coordinated surface sites: Theoretical study of MgO. *Phys. Rev. B*, 59:2417–2430, 1999.

[36] C. Pisani, U. Birkenheuer, S. Casassa, and F. Corà. *EMBED01; An ab initio LCAO-HF Program for local defects in Crystals.* Dipartimento di Chimica Inorganica, Chimica Fisica e Chimica dei Materiali; Università di Torino, via P. Giuria 5; I-10125 Torino, Italy., 2001.

[37] G. Makov and M. C. Payne. Periodic boundary conditions in *ab initio* calculations. *Phys. Rev. B*, 51:4014, 1995.

[38] L. N. Kantorovich. Elimination of the long-range dipole intraction in calculations with periodic boundary conditions. *Phys. Rev. B.*, 60(23):15476, 1999.

[39] T. Yamasaki, C. Kaneta, T. Uchiyama, and K. Terakura T. Uda. Geometry and electronic structures of SiO_2/Si(001) interfaces. *Phys. Rev. B*, 63(11):115314, 2001.

[40] A. Franciosi and C. G. Van de Walle. Heterojunction band offset engineering. *Surface Science Reports*, 25:1–140, 1996.

[41] R. L. Anderson. *Solid-State Electron.*, 5:341, 1962.

[42] Frensley, W. R., and H. Kroemer. *Phys. Rev. B*, 16:2642, 1977.

[43] W. A. Harrison. *J. Vac. Sci. Technol.*, 14:1016, 1977.

[44] Tersoff J. Theory of semiconductor heterojunctions: the role of quantum dipoles. *Phys. Rev. B*, 30(8):4874, 1984.

[45] C. G. Van de Walle and R. M. Martin. Theoretical study of band offsets at semiconductor interfaces. *Phys. Rev. B*, 35(15):8154, 1987.

[46] A. A. Demkov and O. F. Sankey. Growth study and theoretical investigation of the ultrathin oxide SiO_2-Si heterojunction. *Phys. Rev. Lett.*, 83:2038–2041, 1999.

[47] P. P. Edwald. *Ann. Phys.*, 64:253, 1921.

[48] E. R: Smith. Electrostatic energy in ionic crystals. *Proc. Royal Society of London A*, 375:475–505, 1981.

[49] C. Pisani, R. Dovesi, and C. Roetti. *Hartree-Fock treatment of crystalline systems.* Number 48 in Lecture notes in chemistry. Springer-Verlag, Berlin-Heidelberg, 1st edition, 1988.

[50] R. Dovesi. *The languaje of band theory*, volume Lecture notes in Chemistry. of *67*, chapter 2, page 31. Springer-Verlag, Berling-Heidelberg, 1st edition. edition, 1996.

[51] E. R. Davidson and D. Feller. *Chem. Rev.*, 86:681, 1986.

[52] G. Kresse and J. Hafner. Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. *J. Phys.: Condens. Matter.*, 6(40):8245–8257, 1994.

[53] D. Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41(11):7892–2895, 1990.

[54] J.A. Pople and D.L. Beveridge. *Approximate molecular orbital theory.* McGraw-Hill, New York., 1970.

[55] M. J. Frisch, G. W. Trucks, H. B.Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann Jr., J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, K. Morokuma Q. Cui, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople. Gaussian 98 (revision A.11), 2002.

[56] J. A. Pople, D. P. Santry, and G. A. Segal. Approximate self-consistent molecular orbital theory. I. Invariant procedures. *J. Chem. Phys.*, 43(10):S129–S135, 1965.

[57] J. A. Pople and G. A. Segal. Approximate self-consistent molecular orbital theory. II. Calculations with complete neglect of differential overlap. *J. Chem. Phys.*, 43(10):S136–S151, 1965.

[58] J. A. Pople and G. A. Segal. Approximate self-consistent molecular orbitals theory. III. CNDO results for AB_2 and AB_3 systems. *J. Chem. Phys.*, 44(9):3289–3296, 1966.

[59] J.A. Pople, D. L. Beveridge, and P. A. Dobosh. Approximate self-consistent molecular-orbital theory. V. Intermediate neglect of differential overlap. *J. Chem. Phys.*, 47(6):2026–2033, 1967.

[60] A. L. Shluger. Application of molecular-models to electronic-structure calculations of defects in oxide crystals. 1. parameterization of the modified INDO method. *Thoretical Chimia Acta*, 66(6):355–363, 1985.

[61] C. Pisani, R. Dovesi, and C. Roetti. Chapter III Band structure and density of states (DOS). Lecture notes in Chemistry. Springer-Verlag, Berling Heidelberg, 1988.

[62] J.A. Weil, J. R. Bolton, and J. E. Wertz. *Electron Paramagnetic Resonance; Elementary theory and practical applications.* John Wiley and Sons., New York, 1994.

[63] P.W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109(5):1492, 1958.

[64] D. L. Griscom G. Pacchioni, L. Skuja, editor. *Defects in SiO_2 and related dielectrics: Science and Technology.*, volume 2 of *Mathematics, Physics and Chemistry.* Kluwer Academic Publishers, Dordrecht (Nl), 2000.

[65] W. Y. Ching. Microscopic calculation of localized electron states in an intrinsic glass. *Phys. Rev. B*, 46:607, March 1981.

[66] M. Schulz. The end of the road for silicon? *Nature*, 399:729–730, 1999.

[67] P. Balk. 40 years mos technology-from empiricism to science. *Microelectronic Eng.*, 48:3–6, 1999.

[68] C. R. Helms and E. H. Poindexter. The silicon-silicon-dioxide system: its microstructure and imperfections. *Rep. Prog. Phys.*, 57:791–852, 1997.

[69] A. Pasquarello, M. S. Hybertsen, and R. Car. Theory of Si $2p$ core-level shifts at the Si(001)-SiO_2 interface. *Phys. Rev. B*, 53(16):10942, April 1996.

[70] S. T. Pantelides, editor. *The physics of SiO_2 and its Interfaces.* Pergamon, New York, 1978.

[71] K. Ng and D. Vanderbilt. Structure and kinetics of the Si(100)-SiO_2 interface. *Phys. Rev. B*, 59:10132, 1999.

[72] Y. Tu and J. Tersoff. Microscopic dynamics of silicon oxidation. *Phys. Rev. Lett.*, 89:086102, 2002.

[73] Y. Tu and J. Tersoff. Structure of the silicon-oxide interface. *Thin Sol. Films*, 400:95, 2001.

[74] R. L. McGreevy and L. Putzai. Reverse monte carlo simulation: a new technique for the determination of disordered structures. *Mol. Simulation*, 1, 1988.

[75] T. Watanabe and I. Ohdomari. Large-scale modeling of SiO_2/Si(001) interface structures by using a novel inter-atomic interaction model. In H. Z. Massoud, I. J. R. Baumvol, M. Hirose, and E. H. Poindexter, editors, *The physics and chemistry of SiO_2 and the Si-SiO_2 interface-4.*, volume 2000-2, page 319, New jersey, 2000. The Electrochemical Society, Pennington.

[76] T. Watanabe and I. Ohdomari. Modeling of a SiO_2/Si(001) structure including step and terrace configurations. *Appl. Surf. Sci,*, 162-163:116, 2000.

[77] A. Pasquarello and M. S. Hybertsen. Model interface between silicon and disordered SiO_2. In H. Z. Massoud, I. J. R. Baumvol, M. Hirose, and E. H. Poindexter, editors, *The physics and chemistry of SiO_2 and the Si-SiO_2 interface-4.*, volume 2000-2, page 271, New Jersey, 2000. The Electrochemical Society, Pennington.

[78] A. Pasquarello, M. S. Hybertsen, and R. Car. Interface structure between silicon and its oxide by first-priniples molecular dynamics. *Nature*, 396:58, November 1998.

[79] A. Pasquarello, M. S. Hyberten, and R. Car. Structurally relaxed models of the Si(001)-SiO_2 interface. *Appl. Phys. Lett.*, 68(5):625, January 1996.

[80] J.L. Alay, M. Fukuda, C.H. Bjorkman, K. Nakagawa, S. Yokoyama, S. Sasaki, and M. Hirose. Determination of valence-band alignment at ultrathin SiO_2/Si interfaces by high-resolution X-ray photoelectron-spectroscopy. *Japanese J. of Appl. Phys. Part 2-Letters.*, 34(L653-L656), 1995.

[81] R. Tsu, A. Filios, C. Lofgren, D. Cahill, J. Vannostrand, and C. G. Wang. An epitaxial Si/SiO_2 superlattice barrier. *Solid State Electronics*, 40(1-8):221, 1996.

[82] B. R. Tuttle. *Ab initio* valence band offsets between Si(100) and SiO_2 from microscopic models. *Phys. Rev. B*, 67:155324, 2003.

[83] B. Civalleri, C.M. Zicovich-Wilson, P. Ugliengo, V.R. Saunders, and R. Dovesi. A periodic *ab initio* study of the structure and relative stability of silica polymorphs. *Chem. Phys. Lett.*, 292:394–402, 1998.

[84] EMSL web library of gaussian basis sets. http://www.emsl.pnl.gov:2080/forms/basisform.html.

[85] University of Torino (Italy). Theoretical Chemistry Group. CRYSTAL basis sets. http://www.ch.unito.it/ifm/teorica/.

[86] J. E. Ortega and F. J. Himpsel. Inverse-photoemission study of Ge(100), Si(100) and GaAs(100): Bulk bands and surface states. *Phys. Rev. B*, 47:2130, 1993.

[87] L.W. Hobbs and X. Yuan. *Topology and topological disorder in silica.*, volume 2 of *NATO Science Series, II.Mathematics Physics and Chemistry.*, chapter 2, pages 37–71. Kluwer Academic Publishers., 2000.

[88] P. V. Sushko, A. L. Shluger, and C. R. A. Catlow. Relative energies of surface and defect states: *ab initio* calculations for the MgO(001) surface. *Surface Sci.*, 450:153–170, 2000.

[89] J.W. Keister, J.E. Rowe, J.J. Kolodziej, H. Niimi, H.S. Tao, T.E. Madey, and G. Lucovsky. Structure of ultrathin SiO_2/Si(111) interfaces studied by photoemission spectroscopy. *J. of Vacuum Sci. and Technol. A*, 17:1250–1257, 1999.

[90] A. Nohira, A. Omura, and T. Katayama, M. ans Hattori. Valence band edge of ultra-thin silicon oxide near the interface. *Appl. Surf. Sci.*, 123:546–549, 1998.

[91] S. Horiguchi and H. Yoshino. Evaluation of interface potential barrier heights between ultrathin silicon-oxides and silicon. *J. of Appl. Phys.*, 58:1597–1600, 1985.

[92] D. Babić and E. H. Nicolliah. Laplace transform method of measuring the distribution of Si-SiO_2 barrier heights: Basic principles. *J. Appl. Phys.*, 78(7):4516–4522, 1995.

[93] S. Heike, Y. Wada, S. Kondo, and M. Lutwyche. Scanning tunneling microscope measurement of insulator surfaces. *Appl. Phys. Lett.*, 64(9):1100–1102, 1994.

[94] T. Hattori, T. Aiba, E. Iijima, Y. Okube, H. Nohira, N. Tate, and M. Katayama. Initial stage of oxidation of hydrogen-terminated silicon surfaces. *Appl. Surf. Sci.*, 104/105:323–328, 1996.

[95] H. Nohira, H. Sekikawa, M. Matsuda, and T. Hattori. Effect of chemical preoxidation treatment on the structure of SiO_2/Si interfaces. *Appl. Surf. Sci.*, 104/105:359–363, 1996.

[96] P. Perfetti, C. Quaresima, C. Coluzza, C. Fortunato, and Margaritondo G. Dipole-induced changes of the band discontinuities at the SiO_2-Si interface. *Phys. Rev. Lett.*, 57(16):2065–2068, 1986.

[97] H. Kageshima and K. Shraishi. Theoretical study of the band offset at silicon-oxide/silicon interfaces with interfacial defects. *Surf. Sci.*, 407:133–139, 1998.

[98] A. S. Foster, V. B. Sulimov, F. Lopez-Gejo, A. L. Shluger, and R. M. Nieminen. Structure and electrical levels of point defects in monoclinic zirconia. *Phys. Rev B*, 2001. Submitted.

[99] A. S. Foster, V. B. Sulimov, F. Lopez-Gejo, A. L. Shluger, and R. M. Nieminem. Modelling of point defects in monoclinic zirconia. *J. Non-Crystalline Sol.*, 3003:101–107, 2002.

[100] A. S. Foster, F. López Gejo, A. L. Shluger, and R. M. Nieminen. Vacancy and interstitial defects in hafnia. *Phys. Rev. B*, 65:174117, 2002.

[101] Y. M Sun, J. Lozano, H. Ho, H. J. Park, S. Veldman, and J. M. White. Interfacial silicon oxide formation during synthesis of ZrO_2 on Si(100). *Appl. Surf. Sci.*, 161:115–122, 2000.

[102] A. Kawamoto, J. Jameson, P. Griffin, K. Cho, and R. Dutton. Atomic scale effects of zirconium and hafmiun incorporation at a model silicon/silicate interface by first principles calculations. *IEEE Electron. Dev. Lett.*, 22(1):14–16, 2001.

[103] V. V. Brodskii, E. A. Rykova, A. A. Bagatur'yants, and A. A. Korkin. Modelling of ZrO_2 deposition from $ZrCl_4$ and H_2O the Si(100) surface: initial reactions and surface structures. *Computational Materials Sci.*, 24:278–283, 2002.

[104] G. D. Wilk, R. M. Wallace, and J. M. Anthony. High-κ gate dielectrics: current status anbd materials properties considerations. *J. Appl. Phys.*

[105] G. D. Wilk, R. M. Wallace, and J. M. Anthony. Hafnium and zirconium silicates for advanced gate dielectrics. *J. Appl. Phys.*, 87(1):484–492, 2000.

[106] G. Lucovsky and Jr. G. B. Rayner. Microscopic model for enhanced dielectric constants in low concentration SiO_2-rich noncrystalline Zr and Hf silicate alloys. *App. Phys. Lett.*, 77(18):2912–2914, 2000.

[107] R. Orlando, C. Pisani, C. Roetti, and E. Stefanovich. *Ab initio* hartree-fock study of tetragonal and cubic phases of zirconium dioxide. *Phys. Rev. B*, 45(2):592–601, 1992.

[108] E. V. Stefanovich and A. L. Shluger. Theoretical study of the stabilization of cubic-phase ZrO_2 by impurities. *Phys. Rev. B*, 49(17):11560–11571, 1994.

[109] R: H. French, S. J. Glass, F. S: Ohuchi, Y. N. Xu, and W. Y. Ching. Experimental and theoretical determination of the electronic structure and optical properties of 3 phases of ZrO_2. *Phys. Rev. B.*, 49(8):5133–5141, 1994.

[110] M. W. Finnis, A. T. Paxton, M. Methfessel, and M. van Schilfgaarde. Crystal structures of zirconia from first principles and self-consistent tight binding. *Phys. Rev. Lett.*, 81:5149, 1998.

[111] B. Kralik, E. K. Chang, and S. G. Louie. Structural properties and quasiparticle band structure of zirconia. *Phys. Rev. B*, 57(12):7027–7036, 1998.

[112] E. H. Kisi and C. J. Howard. Crystal structures of zirconia phases and their inter-relation. *Key Eng. Mat.*, 153-154:1–36, 1998.

[113] J. Wang, H. P. Li, and R. Stevens. Hafnia and hafnia-toughened ceramics. *J. Mater. Sci.*, 27(20):5397–5430, 1992.

[114] S. Desgreniers and K. Lagarec. High-density ZrO_2 and HfO_2: Crystalline structures and equations of state. *Phys. Rev. B*, 59:8467–8472, 1999.

[115] J. E. Lowther, J. K. Dewhurst, J. M. Leger, and J. Haines. Relative stability of ZrO_2 and HfO_2 structural phases. *Phys. Rev. B*, 60(21):14 485, 1999.

[116] G. Jomard, T. Petit, A. Pasturel, L. Magaud, G. Kresse, and J. Hafner. First-principles calculations to describe zirconia pseudopolymorphs. *Phys. Rev. B*, 59(6):4044–4052, 1999.

[117] R. Ben-Michael, D. S. Tannhauser, and J. Genossar. ESR centers in reduced stabilized zirconia. *Phys. Rev. B*, 43(10):7395–7404, 1991.

[118] P. Aldebert and J. P. Traverse. Structure and ionic mobility of zirconia at high-temperature. *J. Am. Ceram. Soc.*, 68(1):34–40, 1985.

[119] C.J. Howard, R. J. Hill, and B. E. Reichert. Structures of the ZrO_2 polymorphs at room-temperature by high-resolution neutron power difraction. *Acta Crystallographica B*, 44:116–120, 1988.

[120] G. Kresse and J Furthmüller. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. sci.*, 6:15–50, 1996.

[121] G. Kresse and J Furthmüller. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54(16):11169–11186, 1996.

[122] D. Murnaghan. *Proc. Natl. Acad. Sci. U.S.A.*, 30:244, 1944.

[123] D. M. Adams, S. Leonard, D. R. Russell, and R. J. Cernik. X-ray-diffraction study of hafnia under high-pressure using synchrotron radiation. *J. Phys. Chem. Sol.*, 52:1181–1186, 1991.

[124] David R. Lide, editor. *Handbook of chemistry and physics.* CRC Press, 78th edition, 1997.

[125] M. Winter. www.webelements.com.

[126] D. W. McComb. Bonding and electronic structure in zirconia pseudopolymorphs investigated by electron energy-loss spectroscopy. *Phys. Rev. B*, 54(10):7094–7102, 1996.

[127] M. Houssa, M. Tuominen, M. Naili, V. Afanas'ev, A. Stesmans, S. Haukka, and M. M. Heyns. Trap-assisted tunneling in high permittivity gate dielectric stacks. *J. Appl. Phys.*, 87(12):8615–8620, 2000.

[128] M. Balog, M. Scieber, M. Michiman, and S. Patai. *Thin Solid Films*, 41:247, 1977.

[129] M. Busso, S. Cassasa, C. Pisani, and V.B. Sulimov. *Ab initio* simulation of the oxygen vacancy bistability in pure and ge-doped α-quartz. *Modelling Simul. Mater. Sci. Eng.*, 10:21–33, 2001.

[130] P. A. Packan. Pushing the limits. *Science*, 285(5436):2079–2081, 1999.

[131] H.J. Queisser and E. E. Haller. Defects in semiconductors: some fatal, some vital. *Science*, 281(5379):945, 1998.

[132] P. Danesh and A. Szekeres. Electrical properties of hydrogen-rich Si/SiO_2 structures. *J. Non-cryst. Solids*, 187:270, 1995.

[133] E. Cartier, D.A. Buchanan, J.H. Stathis, and D.J. DiMaria. Atomic hydrogen-indiced degradation of thin SiO_2 gate oxides. *J. Non-Cryst. Sol.*, 187:244–247, 1995.

[134] E. H. Poindexter. Chemical reactions of hydrogenous species in the Si/SiO_2 system. *J. Non-Crystalline Sol.*, 187:257–263, 1995.

[135] J. H. Stathis and D. J. DiMaria. *Defects Generation and Reliability of Ultra-Thin SiO_2 at low voltage.*, pages 33–44. Toronto., 2000.

[136] C. G. Van de Walle and B. R. Tuttle. Microscopic theory of hydrogen in silicon devices. *IEEE Transactions Electron Dev.*, 47:1779–1786, 2000.

[137] T. Y. Luo, M. Laughery, G. A. Brown, H. N. Al-Shareef, V. H. C. Watt, A. Karamcheti, M. D. Jaackson, and H. R. Huff. Effect of H_2 content on reliability of ultrathin *in situ* steam generated (ISSG) SiO_2. *IEEE Electron Device Lett.*, 21:430–432, 2000.

[138] J. Isoya, J.A. Weil, and P.H. Davis. Electron paramagnetic resonance of atomic hydrogen ^1H and ^2H in α-quartz. *J. Phys. Chem. Solids*, 44(4):335–343, 1983.

[139] G.V. Gadiyak. Hydrogen redistribution in thin silicon dioxide films under electron injection in high fields. *J. Appl. Phys.*, 82(11):5573, 1997.

[140] A. Bongiorno, L. Colombo, and F. Cargnoni. Hydrogen diffusion in crystalline SiO_2. *Chem. Phys. Lett.*, 264:435–440, 1997.

[141] L. Levien, C.T. Previtt, and D.J. Weidner. *Am. Mineral*, 65:920, 1980.

[142] P. E. Bl¨ ochl, C. G. Van de Walle, and S. T. Pantelides. First-principles calculations of diffusion coefficients: Hydrogen in silicon. *Phys. Rev. Lett.*, 64(12):1401–1404, 1990.

[143] C. Tuma, A. D. Boese, and N. C. Handy. Predicting the binding energies of H-bonded complexes: A comparative dft study. *Phys. Chem. Chem. Phys.*, 1:3939–3947, 1999.

[144] W. Koch and M. C. Holthausen. *A chemist's guide to Density Functional Theory*, chapter 12, page 213. Wiliey-VCH, Verlag GmbH, D-69469 Weinheim, Germany., 2000.

[145] Atkins P. W. *Physical Chemistry.* Oxford University Press, Oxford, 5th edition, 1994.

[146] W. H. Press, S. A. Teukolcky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran 77: The art of scientific computing.* Cambridge University Press, 2^{nd} edition, 1992.

[147] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK User's Guide.* Society for Industrial and Applied Mathematics SIAM, SIAM Customer Service, P.O. Box 7260, Philadelphia, PA 1910, 3^{rd} edition, 1999. LAPACK libraries and guide can be found online at *http : //www.netlib.org/lapack.*

[148] University of Tennesee. *Basic Linear Algebra Subprograms Technical Forum (BLASTF) Standard.*, 2001.

[149] A. A. Sokol. *Defect structures in zeolite crystals.* PhD thesis, The Royal Institution of Great Britain, UCL, University of London, 1998. Chapter 3 (pp 95-127) is devoted to the use of the modified INDO scheme.

[150] J. D. Gale. GULP - a computer program for the symmetry adapted simulation of solids. *JCS Faraday Trans.*, 93:629, 1997.

[151] Y. C. Yeo, W. C. Lu, Q. amd Lee, T. King, C. Hu, X. Wang, X. Guo, and T. P. Ma. Direct tunneling gate leakage current in transistors with ultrathin silicon nitride gate dielectric. *IEEE Electron Dev. Lett.*, 21:540–542, 2000.

[152] T. P. Ma. Making silicon nitride film a viable gate dielectric. *IEEE Transactions Electron Dev.*, 45:680–691, 1998.

[153] Y. Xu and Ching W.Y. Electronic structure and optical properties of $\alpha-$ and $\beta-$ phases of silicon nitride, silicon oxynitride and with comparison to silicon dioxide. *Phys. Rev B*, 51:17 379 –17 389, 1995.

[154] P. Yand, H. K. Fun, I. A. Rahman, and M. I. Saleh. 2 phase refinements of the structures of α-Si_3N_4 and β-Si_3N_4 made from rice husk by rietveld analysis. *Ceramics International*, 21:137–142, 1995.

[155] M. R. Bär and J. Sauer. *Chem. Phys. Lett.*, 226:405, 1994.

[156] P.V. Sushko, J.L. Gavartin, and A. L. Shluger. Electronic properties of structural defects at the MgO(001) surface. *J. Phys. Chem. B*, 106:2269, 2002.

[157] J.M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal. The SIESTA method for *ab initio* order-N materials simulation. *J. Phys. Cond. Matt.*, 14:2745, 2002.