



UNIVERSITEIT ANTWERPEN

Faculteit Wetenschappen

Departement Fysica

An ab initio study
of the electronic and dynamical properties
of Si, Ge and Si/Ge nanowires

Een *ab initio* studie
van de elektronische en dynamische eigenschappen
van Si, Ge en Si/Ge nanodraden

Proefschrift voorgelegd tot het behalen van de graad van doctor in de
wetenschappen aan de Universiteit Antwerpen te verdedigen door

Hartwin Peelaers

Promotor: Prof. dr. B. Partoens
Co-promotor: Prof. dr. F. M. Peeters

Antwerpen 2009

Contents

Acknowledgements	iv
List of abbreviations	vii
1 Introduction	1
1.1 Nanowire growth	2
1.1.1 Top-down approach	2
1.1.2 Bottom-up approach	3
1.2 Applications	8
1.2.1 Nanowire FETs	8
1.2.2 Sensors	9
1.2.3 Logic devices, solar cells and nanoelectronic power sources	10
1.3 Organization of the thesis	12
2 Density Functional Theory	15
2.1 Basic equations	15
2.2 The Hohenberg-Kohn theorems	17
2.2.1 The first Hohenberg-Kohn theorem	17
2.2.2 The second Hohenberg-Kohn theorem	19
2.2.3 The Levy-Lieb constrained search formulation	20
2.3 Kohn-Sham equations	20
2.4 The exchange-correlation energy	23
2.4.1 The exchange-correlation hole	23
2.4.2 Local Density Approximation	27
2.4.3 The Generalized Gradient Approximation	29
2.5 Hellmann-Feynman theorem	29
2.6 Practical calculations	30
2.6.1 Crystal lattice and reciprocal space	31
2.6.2 Bloch theorem	32
2.6.3 Basis sets	33
2.6.4 Band structures and \mathbf{k} -point grids	34
2.6.5 Pseudopotentials	35
3 Electronic properties of B and P doped semiconductor nano-wires	39
3.1 Silicon nanowires	39

CONTENTS

3.1.1	Introduction	39
3.1.2	Computational details	41
3.1.3	Results	41
3.2	Germanium nanowires	47
3.2.1	Introduction	47
3.2.2	Results	47
3.3	Experimental verifications	51
3.4	Conclusions	53
4	Density Functional Perturbation Theory	55
4.1	Phonons: Introduction	55
4.2	Phonons from lattice dynamics	56
4.3	Practical phonon calculations	57
4.3.1	Frozen phonons	57
4.3.2	Density functional perturbation theory	57
5	Phonons in semiconductor nanowires	61
5.1	Silicon nanowires	61
5.2	Germanium nanowires	68
5.3	Core-shell nanowires	73
5.4	Conclusions	78
6	GW approximation: theory	81
6.1	The bandgap problem	81
6.2	Green's functions	83
6.3	Dyson equations and selfenergies	87
6.4	Hedin equations	90
6.5	Practical G_0W_0 calculations	98
7	Full quasi-particle corrected band structures of Si and Ge nanowires	103
7.1	Introduction	103
7.2	LDA convergence	104
7.3	GW convergence	105
7.3.1	Convergence parameters	105
7.3.2	Convergence study	107
7.4	Complete band structure	113
7.4.1	Wannier functions	114

7.4.2	GW corrected bands	116
7.5	Conclusions	118
8	Electronic structure of doped and undoped Si nanowires using a hybrid functional	127
8.1	Introduction	127
8.2	Computational details	128
8.3	Results	129
8.4	Conclusions	133
9	Conclusions and outlook	137
10	Samenvatting	141
	Curriculum Vitae	144
	List of publications	145
	Bibliography	147

On this page I wish to express my gratitude to my promotor Prof. B. Partoens and to my co-promotor Prof. F.M. Peeters, as this thesis would not be the thesis you are reading today without their support, their willingness to answer all questions I had, and the fruitful discussions I had with them.

I also want to thank the people I collaborated with: those from Prof. X. Gonze's group in Louvain-la-Neuve and from Prof. Ph. Ghosez's group in Liège.

My thanks also go out to my colleagues from the CMT group in Antwerp, to the visitors there, and especially to my roommates.

Finally, I also want to thank my parents and brothers for supporting me during my studies.

List of abbreviations

Abbreviation or symbol

1D (2D, 3D)	one-(two-, three-) dimensional
BZ	Brillouin zone
CVD	chemical vapor deposition
DB	dangling bond
DFPT	density functional perturbation theory
DFT	density functional theory
DOS	density of states
FET	field effect transistor
GGA	generalized gradient approximation
HOMO	highest occupied molecular orbital
HK	Hohenberg-Kohn
IBZ	irreducible Brillouin zone
IFC	interatomic force constant
KS	Kohn-Sham
LDA	local density approximation
LED	light emitting diode
LSDA	local spin density approximation
LUMO	lowest unoccupied molecular orbital
MBE	molecular beam epitaxy
MOSFET	metal-oxide-semiconductor field effect transistor
MLWF	maximally localized Wannier function
NW	nanowire
PAW	projected augmented wave
RPA	random phase approximation
SEM	scanning electron microscopy
TEM	tunneling electron microscopy
VLS	vapor-liquid-solid
VSS	vapor-solid-solid

1

Introduction

“There is plenty of room at the bottom”, is the title of the famous talk (1959) of Nobel prize laureate Richard Feynman, in which he envisioned a synthesis of nanoscale building blocks, with precisely controlled size and composition, which could be assembled into larger structures with unique functions and properties. At that time this was still fiction, but nowadays we started to explore this world of nanoscale building blocks.

Nanoscale indicates that the typical sizes are of the order of nanometers (which is equal to one billionth of a meter). These sizes are becoming more and more important in the continuous miniaturization of electronics, as smaller means greater performance: more components per chip, faster operation, lower cost, and less power consumption. According to the roadmap for downscaling and the introduction of new technologies in the semiconductor industry (the International Technology Roadmap for Semiconductors [1]), one dimensional structures such as carbon nanotubes and semiconducting wires are explicitly mentioned as realistic additions to the current technology. However, before devices based on nanotubes or nanowires can be used in industry, a lot of work still has to be performed, both experimental and theoretical.

Semiconducting nanowires have some important advantages over carbon nanotubes. They can be created in a predictable and controllable way, as their chemical composition, diameter, length, doping, growth direction and surfaces can be controlled during their growth. This is possible due to a good understanding of their growth mechanisms and the broad range of chemical compositions available (versus simply carbon). Additionally, reliable methods exist that allow for their parallel [2] assembly. Also nanowire building blocks can be combined in ways not possible in conventional electronics, for example side branches can be grown, thus creating a complex network of

nanowires [3]. It is also possible to create axial or radial (e.g. core-shell nanowires) structures using for instance different chemical compositions. On the other hand for carbon nanotubes it is -despite great progress made- difficult to produce uniform semiconducting nanotubes, as this depends on the chirality of the rolled-up graphene nanosheet.

Due to these advantages semiconducting nanowires have already been assembled in various types of electronic devices, such as field-effect transistors [4, 5, 6], p-n diodes [5, 6], light-emitting diodes (LEDs), bipolar junction transistors [4], complementary inverters [6], nanoscale lasers [7], complex logic gates and even computational circuits that could be used as basic digital calculators [8], sensors for all kind of chemical and biological substances [9, 10], solar cells [11], nanoelectronic power sources [12], and nanogenerators [13].

In this introductory chapter I first focus on some often used growth techniques, followed by a highlight of some of the applications.

1.1 Nanowire growth

There exist a lot of techniques that can be used to create nanowires, which can be classified in two main approaches, i.e. the top-down and the bottom-up approach.

1.1.1 Top-down approach

In the top-down approach, small features are patterned in bulk materials by a combination of lithography, etching, and deposition to form the devices. This approach has been very successful, with microelectronics as good example.

Top-down approaches are more complex with respect to bottom-up ones, because the small dimensionality relies on the ability of performing high resolution lithography. On the other hand, contacts, control gates and interconnections can be easily fabricated at the same time as the nanostructures by using standard integrated circuit technology, without problems of positioning and/or alignment. For the creation of semiconducting nanowires several different methods have been used, such as electron beam lithography [14], as is used for example in the device shown in Fig. 1.1, focussed ion beam milling [15] and atomic force local oxidation [16].

Current developments reduce the resolution limits, but these improvements in resolution cause a near exponential increase in costs. But apart from this economical drawback, other problems arise: the need to make structures with near atomic perfection and the incorporation of distinct chemical and functional properties requires new approaches.

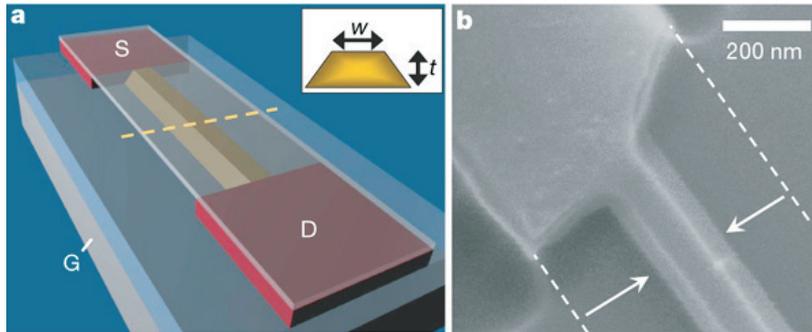


Figure 1.1: (a) Schematic of a top-down generated nanowire by using electron beam lithography, after the etching step. The S, D and G indicate respectively the source, drain and gate. (b) Scanning electron micrograph of the device, showing the trapezoidal form of the nanowire. From Ref. [14].

1.1.2 Bottom-up approach

In the bottom-up approach functional devices are assembled from chemically or physically synthesized nanoscale building blocks. A good analogon for this approach can be found in nature: based on amino acids much more complex proteins can be constructed. This approach has a lot of potential to go beyond the limits and functionality of top-down techniques. In the bottom-up approach new kinds of device concepts can be constructed, as this approach is not limited to a planar geometry as is the case for the top-down approaches.

The bottom-up technique that is currently receiving most attention is the so-called Vapor-Liquid-Solid (VLS) growth. This mechanism was first proposed in the mid-1960s by Wagner and Ellis [18] and recently reused by various research groups. It can be used for a range of different materials. Here we focus on silicon and germanium. This method works by using small metal nanoclusters of e.g. gold, that works as a catalyst. This catalyst is

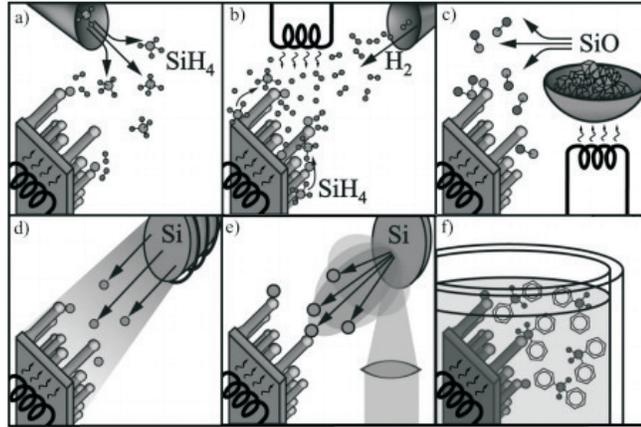


Figure 1.2: Several experimental methods to grow Si nanowires. (a) Chemical Vapor Deposition (CVD), (b) annealing in a reactive atmosphere, (c) evaporation of SiO, (d) Molecular Beam Epitaxy (MBE), (e) laser ablation, and (f) solution-based growth (which uses a growth mechanism similar to VLS). From Ref. [17].

deposited on a substrate and is then heated above the eutectic temperature of the metal and the semiconducting material, that is introduced in the growth chamber in the vapor phase. Currently different experimental methods exist to obtain this vapor phase, which are depicted in Fig. 1.2. It then forms small liquid metal-semiconductor alloys (e.g. Au-Si in our example). The gaseous precursor molecules (e.g. SiCl_4 or SiH_4 for Si) crack on the surface of the droplet and the semiconductor (Si) is incorporated in the droplet. This causes the droplet to become supersaturated with the semiconductor material, that freezes out at the substrate/droplet interface. This process leads to the growth of a wire, with the droplet on top of this wire. The naming of the method is now also clear: the semiconductor in the vapor phase passes through the liquid droplet and ends up as a solid nanowire. This method is schematically depicted in Fig. 1.3. Other methods are possible, such as the Vapor-Solid-Solid (VSS) growth, where the catalyst is a solid particle instead of a liquid droplet. The VLS method can also be observed by in situ transmission electron microscope (TEM), as is shown in Fig. 1.4 for a Ge nanowire with a Au catalyst.

Using this method the growth process can be controlled. The size of the resulting nanowires is largely determined by the size of the droplets. During

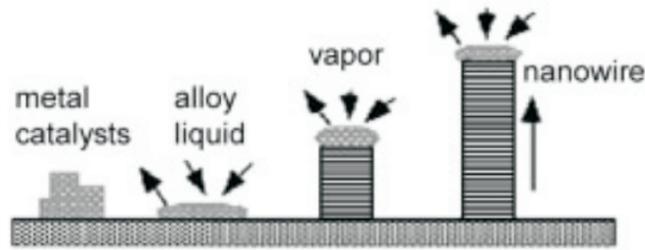


Figure 1.3: Schematic of the VLS growth of a semiconducting nanowire. Starting from a deposited metallic catalyst, a droplet of an alloy of metal-semiconductor is formed above the eutectic temperature when the semiconductor is introduced in the gas phase in the growth chamber. When the liquid becomes supersaturated in the semiconductor material, nanowire growth is achieved. From Ref. [19].

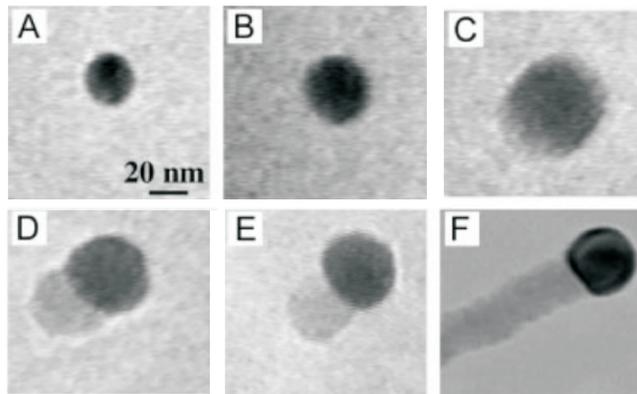


Figure 1.4: The growth of a Ge nanowire, using a Au catalyst, as observed using in-situ TEM. (A) The Au nanocluster in solid state, (B) initiation of alloying at 800°C, (C) a liquid Au/Ge alloy is formed, (D) nucleation of Ge on the alloy surface, (E) further elongation of the Ge nanowire, (F) the obtained Ge nanowire. From Ref. [20].

growth other species can be introduced in the growth chamber, e.g. B_2H_6 , to introduce B dopants in the growing wire. The growing material can also be changed completely, which allows to grow heterostructures, either axial (where sections of different materials with the same diameter are grown along the wire axis) or radial, where core and shell regions can have a different

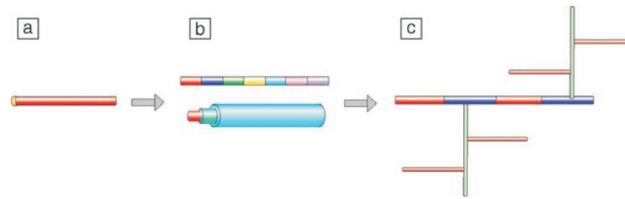


Figure 1.5: Different types of nanowire structures, possible due to the control of the growth conditions. (a) homogeneous material nanowires, (b) axial (top) and radial (bottom) heterostructure nanowires, and (c) branched nanostructures. Different colors indicate different materials or different doping. From Ref. [21].

composition. The interfaces in these structures are abrupt and free of defects. The catalyst determines in the VLS growth where a wire will grow, while the orientation of the wire is determined by the surface lattice of the substrate. Therefore if one decorates a grown radial or axial nanowire with catalyst particles one can obtain side branches, which will grow along the direction of least lattice mismatch between the wire and its side branch. These three different types of structures are depicted in Fig. 1.5.

For the growth of Si nanowires, the most used catalyst material is gold. This has several reasons: it is readily available, it has a high chemical stability (i.e. it will not oxidize), it has low safety requirements as it is not toxic and it has a low vapor pressure, even at high temperatures so that it will not evaporate easily when growing a Si nanowire. There is however one serious drawback, as gold creates deep level defects in Si. These deep levels can act as centers for charge-carrier recombination. The closer the level is to the center of the bandgap, the higher the recombination rate will be. Because of its chemical stability, cleaning of samples and equipment is difficult. Therefore a lot of experimental research was conducted to search for other catalysts. A schematic picture of different catalyst material ionization energies is given in Fig. 1.6. These materials fall roughly in three cases. Those with impurity bands close to the valence or conduction band, so these will cause either p-doping (In, Ga, and Al) or n-doping (Bi, Li, Sb, and Te) and those with levels close to the bandgap middle, e.g. Au, Zn Cu, Fe, Cr, Pb, or Co. If neither of these cases is wanted, Sn, Tl, Ag, Pt, Pd, and Ni remain as candidate catalyst particles. For many of these materials growth of nanowires has been achieved (see e.g. references in Ref. [17]).

1.1. NANOWIRE GROWTH

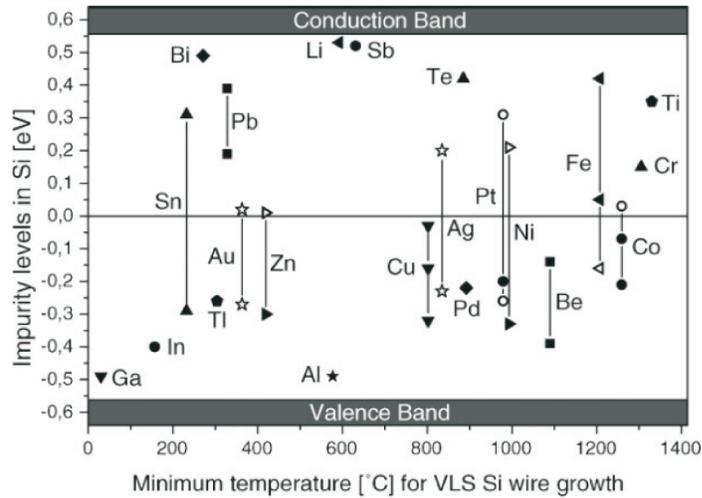


Figure 1.6: The ionization energies of different catalyst materials in Si. These are given with respect to the middle of the bandgap (assuming 1.12 eV for Si) and with their minimum VLS growth temperature. Full symbols above the band gap middle (indicated by the line) indicate donor levels, while open symbols indicate acceptor levels. Below the bandgap middle, full symbols indicate acceptor levels and open symbols donor levels. From Ref. [17].

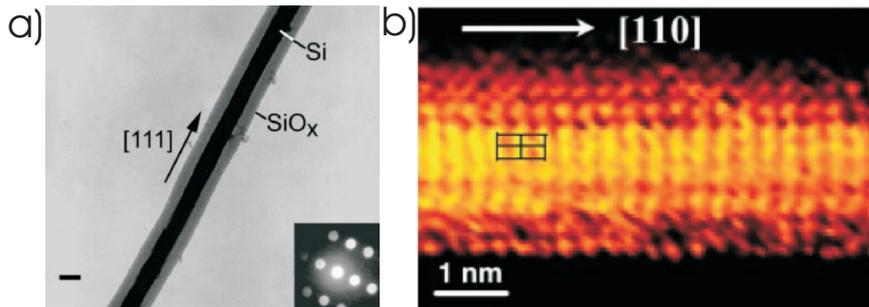


Figure 1.7: (a) Diffraction contrast TEM image of a Si nanowire in the [111] direction. The thick layer of SiO_x is clearly visible. Image from Ref. [22]. (b) STM image of a [110] directed, hydrogen passivated Si nanowire, after HF etching of the SiO_x layer. From Ref. [23].

When growing small diameter Si or Ge nanowires a sheath of SiO_x or GeO_x is formed. This sheath can then be etched away, so that the resulting nanowire diameter is much smaller. The interface between the Si or Ge and the oxide is a smooth interface, as can be seen in Fig. 1.7a. Fig. 1.7b shows a Si nanowire after the etching step. This etching step also allows one to choose another passivation material, for example by using hydrofluoric (HF) acid to create hydrogen passivated wires. The majority of the small diameter Si nanowires are oriented along the [110] [24] direction, while for Ge the majority is oriented along the [111] [25] direction. However the [110] direction is also observed [25, 26].

1.2 Applications

In this section some interesting applications are shown. Please note that this is an exemplary overview in order to obtain an idea of the possibilities of nanowires.

1.2.1 Nanowire FETs

The field effect transistor or FET is the modern workhorse of the semiconductor industry. Therefore a lot of efforts have focused on making FETs from nanowires. Due to the versatile character of the nanowires different approaches are possible. One can use single nanowires placed on a substrate (that often also serves as gate electrode) in a traditional planar device [29]. These devices have performances comparable with state-of-the-art planar devices. Yet another method, which is not possible to achieve in standard planar devices, makes use of a coaxial geometry by utilizing core-multishell wires [28, 30]. In Fig. 1.8(a-d) a possible evolution from a standard metal-oxide-semiconductor FET (MOSFET) towards more advanced geometries is shown: in (b) a FINFET, with a gate surrounding all but one side of the conduction channel, in (c) a nanowire FINFET containing a gate surrounding the conduction channel completely, but non-uniform, and in (d) a vertical nanowire FET, assembled from a core-multishell nanowire, having a uniform wrap-around gate. In (e) a SEM of an actual experimentally realized array of such nanowire FETs is shown. The advantage of these last nanowires is that they can be assembled in parallel, which is important for possible future mass production. The electric properties of these (unoptimized) devices are

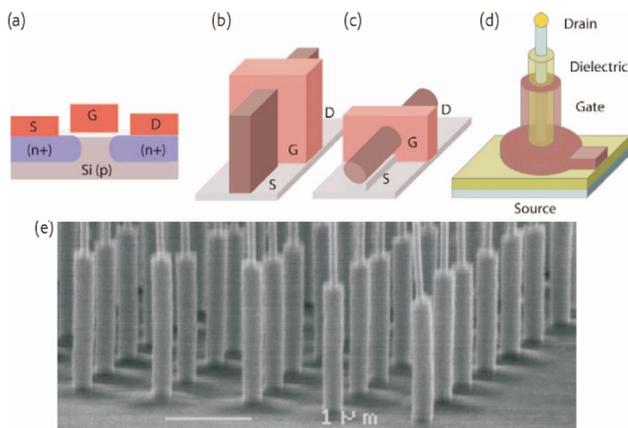


Figure 1.8: Different types of FETs: (a) Schematic of a traditional Metal-oxide-semiconductor FET (MOSFET), where the conduction channel is affected by the planar gate in one direction, (b) FINFET: the gate surrounds the channel from all but one side, (c) Nanowire FINFET: complete, but non-uniform wrap around gate, (d) vertical nanowire FET: complete uniform wrap-around gate, and (e) SEM image of grown vertical nanowire FETs (InAs with a wrap gate of SiN_x). From Ref. [27] and [28].

very good, comparable with results achieved with carbon nanotubes. This good performance comes from the fact that the conduction channels in the wire can be fully squeezed so that no current can pass by the all-around gate.

1.2.2 Sensors

The FETs from the previous section can also be used as sensitive, label-free sensors for chemical and biological species [9, 10, 31]. This application is possible as binding to the nanowire surface is similar to applying a gate voltage, which leads to an accumulation or depletion of charge carriers, which leads to changes in the conductivity. Because of the small diameters and high performance of the nanowire FETs these devices are extremely sensitive, as the binding of a molecule leads to a change in the carrier concentration in the whole wire cross section, while for planar devices this occurs only in the surface region. Examples of this sensitivity are the detection of single virus particles [10], the simultaneous realtime detection of multiple disease marker proteins (e.g. for the detection of cancer proteins) at the femtomolar level (by

using arrays of nanowire FETs) [32], or as a pH [9] or ion measurement [33] device. Fig. 1.9 shows the application of a nanowire for the detection of a single virus (schematically depicted in (a)) or for the simultaneous detection of multiple cancer markers (schematically shown in (b) and measured conductance in (c)). As can be seen only the nanowires with the correct receptors for each protein show a conductance signal and this signal is proportional to the concentration.

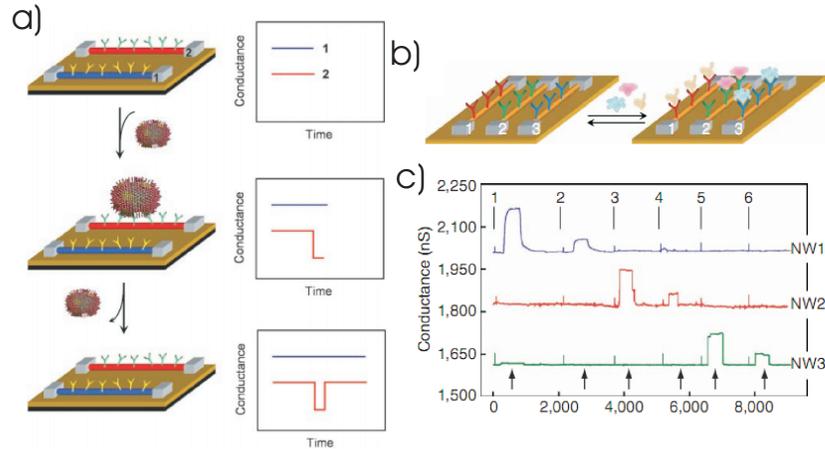


Figure 1.9: Sensor capabilities of nanowires: (a) Schematic view of a single virus attaching to a nanowire with suitable connectors (nanowire 2). This generates a signal in the conductance, while the conductance of the other nanowire, without suitable connectors, does not change. (b) Sensor consisting of three different nanowires, each capable of detecting a specific kind of cancer marker (respectively PSA, CEA, and mucin-1). At the times indicated by numbers the following solutions were injected: 1) 0.9 ng/ml PSA, 2) 1.4 pg/ml PSA, 3) 0.2ng/ml CEA, 4) 2pg/ml CEA, 5) 0.5 ng/ml mucin-1, and 6) 5 pg/ml mucin-1. The black arrows indicate the injection of a buffer solution. From Ref. [10] and [32].

1.2.3 Logic devices, solar cells and nanoelectronic power sources

As explained in the previous subsections, nanowires can be used as FETs. Other devices are also possible. One such device is formed by crossing a p-

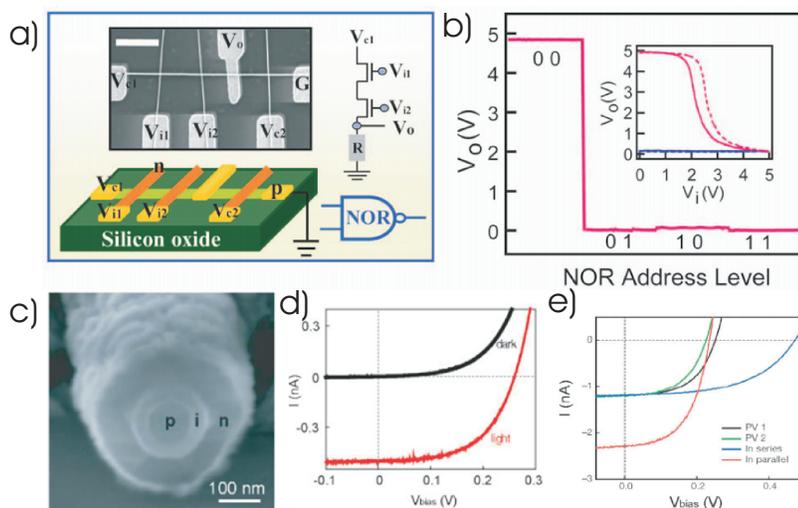


Figure 1.10: (a) Schematic of a logic NOR, constructed from a 1 by 3 crossed nanowire junction array. The inset shows a SEM of an actual device. The scale bar is 1 μm . (b) The output voltage versus the four different input voltages. The inset shows the $V_o - V_i$ relation, where the solid and dashed red (blue) lines correspond to $V_o - V_{i1}$ and $V_o - V_{i2}$ when the other input is 0 (1). (c) A cross section SEM image of a coaxial Si nanowire after wet-chemical etching from the cleaved end. (d) Dark and light I-V curves of the p-i-n device. (e) The I-V curves for two individual (green and black) p-i-n devices, the two devices in series (blue) and in parallel (red). From Ref. [34] and [12].

type Si and an n-type GaN nanowire, which forms a nanoscale p-n junction with a typical turn on voltage of 1V. These results could be reproducibly repeated. Using arrays of these crossed devices, basic logic gates (AND, OR and NOR) can be constructed. These were combined to obtain more complex logic devices such as XOR gates and half-adders [8]. As an example the NOR device is shown in Fig. 1.10(a). Both a schematic picture as well as a SEM image from an actual device are shown. In Fig. 1.10(b) the corresponding measured voltages for different input voltages are shown. The inset shows the output voltage V_o versus the input voltage V_i .

Another area where nanowires show excellent prospects is that of solar cells and photovoltaics. Tian *et al.* [12] used coaxial p-i-n (p-type/intrinsic/n-type) Si nanowires (a cross section SEM is shown in Fig. 1.10(c)) to create

a solar cell, with a maximum power output of 200 pW per nanowire under the standard test environment of solar cells. The difference in I-V curves in dark versus light environment is shown in Fig. 1.10(d), which clearly shows desirable device characteristics. Due to the structure of these wires, the p-i-n interface extends along the wire, while the carrier separation takes place in the radial direction. This yields a carrier collection distance smaller or comparable to the minority carrier diffusion length, so that the photogenerated carriers can reach the p-i-n junction with high efficiency with low bulk recombination rates. Placing these devices in parallel or series allows to scale the output characteristics and to drive larger loads. This is shown in Fig. 1.10(e), which shows the I-V curves for two wires either separately, in parallel, or in series. These devices can be connected to for example the previously discussed devices (e.g. the sensor or the logic gates) to provide power for them.

1.3 Organization of the thesis

This thesis is organized in a layered fashion: chapters containing some of the basic background theory are alternated with chapters containing my own results.

Doping of nanowires plays an important role in several of the applications discussed in the previous section. Experimentally it was observed that wires doped with phosphorus have a higher resistivity than wires doped with boron, for similar concentrations [35, 36]. As *ab initio* calculations allow to study this doping on an atomic level, I performed a detailed study of dopant localization in both Si and Ge nanowires. For thin nanowires the surface to volume ratio is very high, so that surface defects can have a large influence on the properties of the wires. Therefore the electronic properties of doped nanowires in the presence of a dangling bond defect (a missing passivating hydrogen atom, which is a commonly occurring defect) are also studied in **chapter 3**.

Because of the nature of *ab initio* calculations, periodic boundaries were used (see **chapter 2** for a theoretical background on these calculations). The consequence of this is that the dopant concentration is very large if the unit cell is small. A valid question arises in this case: are these structures stable under perturbations. Ground state calculations alone do not provide this information. Therefore we used the *density-functional perturbation theory* in

order to calculate the effects of a perturbation of the atomic positions, which resulted in the calculation of the phonon spectra. After assessing the stability of the previously used structures, these are used to study the effect of the wire diameter on the phonon frequencies. As Si/Ge core-shell structures are also important in applications (as was shown in this introduction), the electronic and dynamical properties of these structures were studied. These results are presented in **chapter 5**.

A known shortcoming of the used *ab initio* techniques is that the band gap is severely underestimated (see **chapter 6** for the theory). By switching to another approximation called GW, based on many-body theory, this problem can be corrected. However in the case of one dimensional structures special care should be taken, as the presence of the vacuum layers surrounding the nanowires causes numerical problems. I discuss the practical application of several available new techniques in **chapter 7**. Also full GW corrected band structures are presented, band masses within GW are calculated and the application of a scissor shift operation is discussed. The methodology used in this chapter was applied in collaboration with the group of X. Gonze of the Université Catholique de Louvain-la-Neuve (UCL).

The usage of a hybrid functional provides a tool to obtain good band structures (including a good prediction of the band gaps), without having the same computational cost as full GW calculations. Therefore our calculations could be extended to large diameter wires, which allowed us to see the direct-to-indirect bandgap transition, as well as for low dopant concentrations to see the localization of the impurity band, which is shown in **chapter 8**. These results were obtained in collaboration with E. Durgun and Ph. Ghosez of the Université de Liège (ULg).

Chapter 2 is the first of the background theory chapters, as it explains the Density Functional Theory (DFT), the theoretical method used throughout this thesis. Also a refresher of basic solid state physics (crystal lattices, reciprocal space, bandstructures, Bloch theorem,...) is given.

Chapter 4 gives a brief introduction to the Density Functional Perturbation Theory (DFPT), which can be used to study perturbations of a system. An important example of these perturbation calculations is the calculation of phonons.

In the last theory chapter, **chapter 6**, I explain the GW approximation, which can be used to obtain a corrected band spectrum, with e.g. more realistic band gaps.

2

Density Functional Theory

In this chapter the basics of the Density Functional Theory (DFT) are explained. First the many-body problem is described, followed by a theoretical proof that the density can be considered as the basic variable (by means of the Hohenberg-Kohn theorems). This is followed by a practical methodology to solve the actual equations by mapping them onto a system of non-interacting particles is given. This leads to the Kohn-Sham equations. In these equations the exchange-correlation energy plays an important role, so in the next section this is studied in more detail. Finally a practical calculation is considered. As such a calculation is done on a crystal, the crystal lattice and reciprocal space is introduced. Due to the application of periodic boundary conditions the Bloch theorem is used. In the last sections band structures, \mathbf{k} -point grids, the utilized basis and pseudopotentials are discussed.

Except for the explicit references made in this chapter, it is also based on Richard M. Martin's book on "Electronic structures" [37], which I suggest for more detailed information. Also several other texts were consulted [38, 39, 40].

2.1 Basic equations

The ground state of a system consisting of atomic cores surrounded by electrons can be found by solving a time independent Schrödinger equation, which has the form

$$\hat{H}\Psi = E\Psi, \tag{2.1}$$

where \hat{H} is the Hamiltonian of the system, E the electronic energy and Ψ the many-body wave function. This Hamiltonian is given by

$$\begin{aligned} \hat{H} = & -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 + \sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \\ & - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}, \end{aligned} \quad (2.2)$$

where capital indices refer to the atomic cores and small indices to the electrons. Z indicates the atomic mass and e indicates the electron charge. The first term corresponds with the kinetic energy of the electrons, the second term with the Coulomb interaction between electrons and cores, the third with the Coulomb interaction between electrons, the fourth with the kinetic energy of the cores, and the last term corresponds with the Coulomb interaction between the cores. In the following the so-called atomic units will be used, so $\hbar = e = m_e = 1$.

This Hamiltonian can be simplified by taking into account the fact that even for the lightest atom (the hydrogen atom), the mass of the core is already 1800 times the mass of an electron and that for example for a carbon atom this ratio is more than 20000. Therefore one can consider the electrons as moving in the field generated by the solid cores. This is the Born-Oppenheimer approximation [41]. This means that the fourth term in the Hamiltonian (the kinetic energy of the cores) can be neglected and that the fifth term (the core-core Coulomb interaction) is just a classical constant which will be denoted as E_{II} . The Hamiltonian then becomes

$$\hat{H} = \underbrace{-\frac{1}{2} \sum_i \nabla_i^2}_{\hat{T}} + \underbrace{\sum_{i,I} \frac{Z_I}{r_{iI}}}_{\hat{V}_{ext}(\mathbf{r})} + \underbrace{\frac{1}{2} \sum_{i \neq j} \frac{1}{r_{ij}}}_{\hat{V}_{int}} + E_{II}. \quad (2.3)$$

In this expression the kinetic energy operator is denoted \hat{T} , the potential experienced by the electrons and generated by the fixed cores is \hat{V}_{ext} and the internal potential is \hat{V}_{int} . Formally one can write the electronic energy as the expectation value of the Hamiltonian

$$E = \langle \Psi | \hat{H} | \Psi \rangle = \langle \hat{H} \rangle = \langle \hat{T} \rangle + \langle \hat{V}_{int} \rangle + \int d^3r V_{ext}(\mathbf{r}) n(\mathbf{r}) + E_{II}, \quad (2.4)$$

2.2. THE HOHENBERG-KOHN THEOREMS

where $n(\mathbf{r})$ is the density, which is defined as the expectation value of the density operator

$$\hat{n}(\mathbf{r}) = \sum_{i=1}^N \delta(\mathbf{r} - \mathbf{r}_i). \quad (2.5)$$

The wavefunctions itself are many-body wavefunctions and they have to satisfy the normalization condition

$$\int d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 = 1. \quad (2.6)$$

Additionally, they have to be anti-symmetric under the exchange of any 2 electrons, denoted by the indices i and j

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_i, \dots, \mathbf{r}_j, \dots, \mathbf{r}_N) = -\Psi(\mathbf{r}_1, \dots, \mathbf{r}_j, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N). \quad (2.7)$$

To be able to solve equation 2.4 explicitly we use the Hohenberg-Kohn theorems [42] and the Kohn-Sham equations [43], which are derived in the next sections.

2.2 The Hohenberg-Kohn theorems

2.2.1 The first Hohenberg-Kohn theorem

The first Hohenberg-Kohn theorem reads as follows:

For any system of interacting particles in an external potential $V_{ext}(\mathbf{r})$ the potential $V_{ext}(\mathbf{r})$ is determined uniquely, except for a constant, by the ground state density $n_0(\mathbf{r})$.

Please note that the first theorem requires a non-degenerate ground state. This shall be extended later so that degenerate ground states are also allowed. The proof of this theorem is done in two steps.

First we prove that two potentials $V_{ext}^1(\mathbf{r})$ and $V_{ext}^2(\mathbf{r})$ lead to two different wavefunctions, if these potentials are differing more than just a shift. Suppose these two potentials lead to the same wavefunction. One can then write the Schrödinger equation for both Hamiltonians:

$$\begin{aligned} (\hat{T} + \hat{V}_{int} + V_{ext}^1)\Psi &= E_1\Psi \\ (\hat{T} + \hat{V}_{int} + V_{ext}^2)\Psi &= E_2\Psi \\ \Rightarrow (V_{ext}^1 - V_{ext}^2 + \Delta E)\Psi &= 0. \end{aligned}$$

CHAPTER 2. DENSITY FUNCTIONAL THEORY

The last equation is only possible if $V_{ext}^1 - V_{ext}^2 = -\Delta E$, which corresponds to just a shift in the potential, which contradicts our previous assumption. Therefore the wavefunctions should be different.

We are now ready to prove the Hohenberg-Kohn theorem. Suppose one has two potentials V_{ext}^1 and V_{ext}^2 , which differ from each other more than just a shift, which lead to the same ground state density $n_0(\mathbf{r})$. Each external potential leads to a different Hamiltonian and to a different wavefunction as shown before. These are labeled with either a superscript one or two. Schematically this is

$$\begin{aligned} V_{ext}^1 &\rightarrow H^1 \rightarrow \Psi^1 \\ V_{ext}^2 &\rightarrow H^2 \rightarrow \Psi^2. \end{aligned}$$

The energy is then

$$E^1 = \langle \Psi^1 | H^1 | \Psi^1 \rangle < \langle \Psi^2 | H^1 | \Psi^2 \rangle, \quad (2.8)$$

which follows from the variational principle of quantum mechanics. This statement can be rewritten as

$$E^1 < \langle \Psi^2 | H^2 | \Psi^2 \rangle + \langle \Psi^2 | H^1 - H^2 | \Psi^2 \rangle \quad (2.9)$$

$$< E^2 + \int d^3r (V_{ext}^1(\mathbf{r}) - V_{ext}^2(\mathbf{r})) n_0(\mathbf{r}). \quad (2.10)$$

Similarly one can obtain for the second Hamiltonian

$$E^2 < E^1 + \int d^3r (V_{ext}^2(\mathbf{r}) - V_{ext}^1(\mathbf{r})) n_0(\mathbf{r}). \quad (2.11)$$

Summing 2.10 and 2.11 leads to $E^1 + E^2 < E^1 + E^2$, which is impossible. Therefore one can conclude that the external potential is determined uniquely by the density and therefore the density also determines the Hamiltonian H and the corresponding wavefunction Ψ .

Schematically this can be written as

$$\begin{array}{ccc} n_0(\mathbf{r}) & \xRightarrow{HK} & V_{ext}(\mathbf{r}) \\ \uparrow & & \downarrow \\ \Psi_0(\{\mathbf{r}\}) & \leftarrow & \Psi_i(\{\mathbf{r}\}), \end{array}$$

where the arrow HK indicates the application of the first Hohenberg-Kohn theorem. This is an important result, as this indicates that we can consider the potential as being a functional of the ground-state density.

2.2.2 The second Hohenberg-Kohn theorem

The second Hohenberg-Kohn theorem reads as follows:

A universal functional for the energy $E[n]$ in terms of the density $n(\mathbf{r})$ can be defined, valid for any external potential $V_{ext}(\mathbf{r})$. The exact ground state energy of the system is the global minimum of this functional and the density that minimizes the functional is the exact ground state density $n_0(\mathbf{r})$.

This can be shown by fixing the possible shift in the potential (e.g. by adding a condition such as: the potential should go to zero at infinity). By doing this the electronic energy can be considered as being a functional of the density, as all terms are determined when the density is determined. Thus one has

$$E_{HK}[n] = T[n] + E_{int}[n] + \int d^3r V_{ext}(\mathbf{r})n(\mathbf{r}) + E_{II}. \quad (2.12)$$

The first two terms are universal by construction, as they only contain internal energies (kinetic and interaction energy), which are only functionals of the density. This universal functional is called F_{HK} .

Suppose one has a system with ground state density $n^1(\mathbf{r})$ and ground state wavefunction Ψ^1 , corresponding to an external potential $V^1(\mathbf{r})$. The Hohenberg-Kohn energy then becomes

$$E_{HK} = E^1 = \langle \Psi^1 | \hat{H}^1 | \Psi^1 \rangle.$$

If a different density $n^2(\mathbf{r})$ is considered, this corresponds to a different wavefunction Ψ^2 . From the minimal property of the ground state it follows that this density leads to a higher energy.

$$E^1 = \langle \Psi^1 | \hat{H}^1 | \Psi^1 \rangle < \langle \Psi^2 | \hat{H}^1 | \Psi^2 \rangle = E^2$$

This proves the second Hohenberg-Kohn theorem. Please note that we restricted ourselves to densities that correspond to the ground state densities of Hamiltonians with certain external potentials V_{ext} . These densities are called “V-representable” densities. In fact, this restriction is not necessary to prove the theorem. The constrained search formulation of Levy and Lieb does not require this restriction and has the advantage that it has a clearer physical interpretation.

2.2.3 The Levy-Lieb constrained search formulation

This formulation makes use of the extremal principle of quantum mechanics. One can write the ground state energy as

$$E = E_{LL} = \min_{\Psi} \left(\langle \Psi | \hat{H} | \Psi \rangle \right). \quad (2.13)$$

This minimization is performed in two steps: first the minimization over all wavefunctions is performed leading to a certain density, followed by a minimization over all densities. Thus

$$E_{LL} = \min_n \left(\min_{\Psi \rightarrow n} \langle \Psi | \hat{H} | \Psi \rangle \right) \quad (2.14)$$

$$= \min_n \left(\min_{\Psi \rightarrow n} \langle \Psi | \hat{T} + \hat{V}_{int} | \Psi \rangle + \int d^3r n(\mathbf{r}) V_{ext}(\mathbf{r}) \right) \quad (2.15)$$

$$= \min_n \left(\min_{\Psi \rightarrow n} \left(F[n] + \int d^3r n(\mathbf{r}) V_{ext}(\mathbf{r}) \right) \right) \quad (2.16)$$

This functional is defined for all densities that can be derived from wavefunctions with N electrons. Therefore these densities are called “N-representable”. This is a weaker restriction than the “V-representability”, and in practice, since all applications are related to wavefunction techniques, all densities occurring satisfy this condition. The minimum of this functional is the same as the minimum of the Hohenberg-Kohn functional, since the minimizing density is a density that can be generated by an external potential. Additionally this approach can handle degenerate ground states.

2.3 Kohn-Sham equations

We now have a formal method to find the ground state energy of the initial system. Unfortunately we still have to solve an interacting many-body problem. If our problem could be reformulated as a problem of independent particles, it could be solved more easily. This reformulation of the problem leads to the Kohn-Sham equations, as is shown in this section.

The interacting many-body system will be mapped on a system that is non-interacting, but which has the same density. The kinetic energy functional of this system is defined as

$$T_S[n] = \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} | \Psi \rangle. \quad (2.17)$$

2.3. KOHN-SHAM EQUATIONS

Using 2.16 and 2.17 one can define an exchange-correlation functional of the density as follows

$$E_{xc}[n] = F[n] - T_S[n] - \frac{1}{2} \int \int \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2, \quad (2.18)$$

where the last term is the Hartree energy (E_H), which is the classical Coulomb contribution to the energy of the system interacting with itself. The exchange-correlation functional is still not known, but let's assume we do.

The original energy functional (but rewritten using the definition of the exchange-correlation functional 2.18), which is labeled the Kohn-Sham energy (E_{KS}), is given by

$$E_{KS}[n] = T_S[n] + \int d\mathbf{r} V_{ext}(\mathbf{r})n(\mathbf{r}) + E_H[n] + E_{xc}[n]. \quad (2.19)$$

This energy has to be minimized under the constraint of the total electron number N . This constraint is introduced by using the Lagrange multiplier λ . This leads to the following expression

$$\begin{aligned} 0 &= \delta \left(E_{KS}[n] - \lambda \left\{ \int d\mathbf{r} n(\mathbf{r}) - N \right\} \right) \\ &= \int d\mathbf{r} \left(\frac{\delta T_S[n]}{\delta n(\mathbf{r})} + V_{ext}(\mathbf{r}) + \int d\mathbf{r}_1 \frac{n(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}|} + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} - \lambda \right) \delta n(\mathbf{r}). \end{aligned} \quad (2.20)$$

(2.21)

If, on the other hand, the non-interacting system of electrons moving in a Kohn-Sham potential V_{KS} , having the same density and the same number of electrons N , is minimized, one ends up with

$$0 = \int d\mathbf{r} \left(\frac{\delta T_S[n]}{\delta n(\mathbf{r})} + V_{KS} - \lambda \right) \delta n(\mathbf{r}). \quad (2.22)$$

Comparing equations 2.21 and 2.22, one can extract an expression for the Kohn-Sham potential:

$$\begin{aligned} V_{KS}(\mathbf{r}) &= V_{ext}(\mathbf{r}) + \int d\mathbf{r}_1 \frac{n(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}|} + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \\ &= V_{ext}(\mathbf{r}) + \int d\mathbf{r}_1 \frac{n(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}|} + v_{xc}(\mathbf{r}), \end{aligned} \quad (2.23)$$

where the notation $v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}$ was introduced in the last line.

We now have simplified our problem to solving a non-interacting system of electrons, moving in the Kohn-Sham potential. For this system the Schrödinger equation

$$\left(-\frac{1}{2}\nabla^2 + V_{KS}(\mathbf{r})\right)\Psi_i(\mathbf{r}) = \varepsilon_i\Psi_i(\mathbf{r}), \quad (2.24)$$

can be written, which can be solved to obtain the eigenfunctions Ψ_i and the eigenvalues ε_i . The eigenfunctions can be used to construct the density

$$n(\mathbf{r}) = \sum_i \Psi_i^*(\mathbf{r})\Psi_i(\mathbf{r}). \quad (2.25)$$

A problem occurring in this equation is that in order to calculate V_{KS} , one needs to have the density, and this density can only be obtained by solving this equation. So one has to solve this equation self-consistently: one starts with an educated guess for the density (a trial density), calculates the Kohn-Sham potential V_{KS} , uses this in equation 2.24 to obtain the eigenfunctions Ψ_i , uses these to calculate a new density, and start the cycle again with this new density. Of course one needs a criterion to abort this cycle. For example one can demand that the difference between the Kohn-Sham potential from the previous cycle and the newly obtained one is smaller than some predefined convergence value. When this happens self-consistency is achieved and the Kohn-Sham equations are solved.

Schematically the importance of the Kohn-Sham equations can be represented as

$$\begin{array}{ccccc} V_{ext}(\mathbf{r}) & \xleftarrow{HK} & n_0(\mathbf{r}) & \xleftrightarrow{KS} & n_0(\mathbf{r}) & \xrightarrow{HK_0} & V_{KS}(\mathbf{r}) \\ \downarrow & & \uparrow & & \uparrow & & \downarrow \\ \Psi_i(\{\mathbf{r}\}) & \Rightarrow & \Psi_0(\{\mathbf{r}\}) & & \Psi_{i=1, N_e} & \Leftarrow & \Psi_i(\mathbf{r}). \end{array}$$

The left-hand side contains the many-body wavefunctions, where the first Hohenberg-Kohn theorem is applied (see chapter 2.2.1). The right-hand side refers to the non-interacting system, that is chosen so that it has the same groundstate density as the interacting system. HK_0 refers to the application of the first Hohenberg-Kohn theorem to the non-interacting system.

Alternatively the problem can be rewritten as a minimization problem. In that case one has to minimize

$$E_{KS}\{\Psi_i\} = \min_{\Psi_i} \sum_i \langle \Psi_i | \hat{T} + \hat{V}_{ext} | \Psi_i \rangle + E_H[n] + E_{xc}[n], \quad (2.26)$$

under the constraint that

$$\langle \Psi_i | \Psi_j \rangle = \delta_{ij}. \quad (2.27)$$

Both approaches are equivalent, but in the ABINIT program, the self-consistent method of solving the Kohn-Sham equations is chosen.

2.4 The exchange-correlation energy

All we have to do to solve the original many-body problem is solving the Kohn-Sham equation 2.24. While being an exact equation, it is still not possible to solve this as the exchange-correlation functional, which should be a universal functional, is not known. However one can approximate this functional, which is discussed in this section. Before doing that, the exchange-correlation hole is studied.

2.4.1 The exchange-correlation hole

It is instructive to have a closer look at the significance of the exchange-correlation energy. Taken again equation 2.18, which can be written as

$$E_{xc}[n] = F[n] - (T_S[n] - E_H). \quad (2.28)$$

This equation can be interpreted as the difference between the true internal energy ($F[n]$) and an approximation, in which the kinetic energy is that of the non-interacting system ($T_S[n]$) and the potential energy is that of a non-quantized charge distribution (E_H). Further splitting this equation gives

$$\begin{aligned} E_{xc} &= T + V_{int} - T_S - E_H \\ &= (T - T_S) + (V_{int} - E_H) \\ &= (T - T_S) + V_{xc}, \end{aligned} \quad (2.29)$$

where the notation $[n]$, to indicate the functional dependency on the density n , is dropped and where V_{xc} , the exchange-correlation potential, is defined.

The Hartree energy, defined as

$$E_H = \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}', \quad (2.30)$$

can be rewritten as

$$E_H = \int d\mathbf{r} v_H(\mathbf{r}) n(\mathbf{r}), \quad (2.31)$$

where $v_H(\mathbf{r})$ is the Hartree energy per particle given by

$$v_H(\mathbf{r}) = \frac{1}{2} \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (2.32)$$

So, if a particle is located at \mathbf{r} in a non-quantized charge density, it can be considered to have a potential energy $v_H(\mathbf{r})$ due to its interaction with the charge density. This is consistent with the total Hartree energy (equation 2.31) being the integral of v_H times the electron density.

A similar explanation can be given to the internal potential V_{int} so that

$$V_{int} = \int d\mathbf{r} v_{int}(\mathbf{r}) n(\mathbf{r}), \quad (2.33)$$

where v_{int} is the internal potential energy per electron. This is the energy resulting from the interaction of an electron at \mathbf{r} with a charge distribution. However, a complication occurs in this case, as the charge distribution will be modified due to the presence of the electron. Thus for every position \mathbf{r} a conditional density $n(\mathbf{r}'|\mathbf{r})$ will be seen by the electron. The internal potential energy per electron thus becomes

$$v_{int}(\mathbf{r}) = \frac{1}{2} \int d\mathbf{r}' \frac{n(\mathbf{r}'|\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|}. \quad (2.34)$$

The conditional density can also be written as

$$n(\mathbf{r}'|\mathbf{r}) = \frac{n(\mathbf{r}, \mathbf{r}')}{n(\mathbf{r})}, \quad (2.35)$$

where $n(\mathbf{r}, \mathbf{r}')$ is the electron pair density, which is given by

$$n(\mathbf{r}, \mathbf{r}') = n(\mathbf{r}) n(\mathbf{r}') g(\mathbf{r}, \mathbf{r}'). \quad (2.36)$$

The function $g(\mathbf{r}, \mathbf{r}')$ in this formula is called the pair-correlation function. If the two electrons were totally independent, the electron pair density would be equal to the product of the individual electron densities and the pair-correlation function g would be equal to 1, which also corresponds to its

2.4. THE EXCHANGE-CORRELATION ENERGY

maximum value. The larger the correlation, the smaller the pair-correlation will be, thus it can be seen as a measure of the correlation. Combining equations 2.34, 2.35 and 2.36 gives

$$v_{int}(\mathbf{r}) = \frac{1}{2} \int d\mathbf{r}' \frac{n(\mathbf{r}')g(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (2.37)$$

Similarly, the exchange-correlation energy per particle can be defined, using the definitions of the Hartree energy per particle (equation 2.32) and the internal potential energy per particle (equation 2.37) as

$$\begin{aligned} v_{xc}(\mathbf{r}) &= v_{int}(\mathbf{r}) - v_H(\mathbf{r}) \\ &= \frac{1}{2} \int d\mathbf{r}' \frac{n(\mathbf{r}')g(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{2} \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\ &= \frac{1}{2} \int d\mathbf{r}' \frac{n(\mathbf{r}') (g(\mathbf{r}, \mathbf{r}') - 1)}{|\mathbf{r} - \mathbf{r}'|} \\ &= \frac{1}{2} \int d\mathbf{r}' \frac{n_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \end{aligned} \quad (2.38)$$

where in the last line the exchange-correlation density n_{xc} is defined as $n_{xc}(\mathbf{r}, \mathbf{r}') = n(\mathbf{r}') (g(\mathbf{r}, \mathbf{r}') - 1)$. The total exchange-correlation energy is then given by

$$\begin{aligned} E_{xc} &= \int d\mathbf{r} v_{xc}(\mathbf{r})n(\mathbf{r}) \\ &= \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{n_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} n(\mathbf{r}) \end{aligned} \quad (2.39)$$

Thus the exchange-correlation energy can be seen as caused by the Coulomb interaction between the electron density and the exchange-correlation density. It can be shown that the integral over all space of this density is given by

$$\int d\mathbf{r}' n_{xc}(\mathbf{r}, \mathbf{r}') = -1. \quad (2.40)$$

If the spins are taken into account, this integral becomes equal to $-\delta_{\sigma, \sigma'}$, where σ and σ' are the spins of both electrons. This result is the so-called “exchange-correlation hole” which surrounds every electron. It integrates to

negative unity for equal spins, which indicates a lack of electrons of the same spin around each electron as can be expected for fermions, while for different spins it becomes zero.

In the previously used definition for the exchange correlation energy 2.29 we considered the non-interacting system and the true many-body interacting system as two different systems, having the same ground-state density. It is however possible to consider both systems as two different instances of a continuous set of systems. Each instance of this set will be characterized by a parameter λ , called the electron-electron coupling constant. This can be done by considering the interaction $\lambda/|\mathbf{r} - \mathbf{r}'|$, where λ equal to zero corresponds to the non-interacting system and a value of one to the interacting system. The continuous path between these two systems is also called the adiabatic connection. To constrain the system so that the density of each system is equal, an external potential V_λ is added. The Hamiltonian then becomes

$$\hat{H}_\lambda = \hat{T} + \lambda V_{int} + \sum_i^N V_\lambda. \quad (2.41)$$

Defining the functional $F_\lambda[n]$ as

$$\begin{aligned} F_\lambda[n] &= \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} + \lambda \hat{V}_{int} | \Psi \rangle \\ &= \langle \Psi^\lambda | \hat{T} + \lambda \hat{V}_{int} | \Psi^\lambda \rangle, \end{aligned} \quad (2.42)$$

we obtain

$$\begin{aligned} F_1[n] &= F[n] = T[n] + V_{int}[n] \\ F_0[n] &= \tilde{T}_S[n]. \end{aligned} \quad (2.43)$$

Using these in an equation similar to 2.29 gives

$$\begin{aligned} \tilde{E}_{xc} &= (T - \tilde{T}_S) + (V_{int} - E_H) \\ &= F_1 - F_0 - E_H \\ &= \int_0^1 d\lambda \frac{\partial F_\lambda[n]}{\partial \lambda} - E_H. \end{aligned} \quad (2.44)$$

\tilde{E}_{xc} will be equal to E_{xc} (from equation 2.29) if the density is non-interacting V-representable (as $T_S = \tilde{T}_S$ in that case). Further rewriting (using 2.42)

gives

$$\begin{aligned}\tilde{E}_{xc} &= \int_0^1 d\lambda \langle \Psi^\lambda | \hat{V}_{int} | \Psi^\lambda \rangle - E_H \\ &= \int \int d\mathbf{r} d\mathbf{r}' \frac{\tilde{n}_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} n(\mathbf{r}),\end{aligned}\tag{2.45}$$

where $\tilde{n}_{xc}(\mathbf{r}, \mathbf{r}')$ is the coupling constant averaged exchange-correlation hole given by

$$\tilde{n}_{xc}(\mathbf{r}, \mathbf{r}') = \int_0^1 d\lambda n_{xc}(\mathbf{r}, \mathbf{r}').\tag{2.46}$$

This coupling constant averaged exchange-correlation hole is often used to examine new exchange-correlation functionals. Please note that the exchange-correlation energy only depends on the spherical average of the exchange-correlation hole.

2.4.2 Local Density Approximation

The simplest physical way to approximate the exchange-correlation energy is the Local Density Approximation (LDA). In this approximation two assumptions are made: i) the local exchange-correlation energy per particle only depends on the local density (hence the name of the approximation) and ii) is equal to the exchange-correlation energy per particle of a homogeneous electron gas, that has the same density, in a neutralizing positive background (“jellium” background). The total exchange-correlation energy E_{xc} is then given by the sum of the contributions of each point in space, where it is assumed that the contribution of one point only depends on the density of that particular point, independent of the other points. So

$$E_{xc}[n] = \int d^3r n(\mathbf{r}) \varepsilon_{xc}(n(\mathbf{r})),\tag{2.47}$$

where $\varepsilon_{xc}(n(\mathbf{r}))$ is the exchange-correlation energy per particle. In the LDA approximation this is equal to $\varepsilon_{xc} = \varepsilon_{xc}^{LDA}$. This approximation is more accurate for systems with slowly varying densities, as it is assumed that the density is locally a constant.

The value of this exchange-correlation energy is calculated by quantum Monte Carlo calculations [44], which are tabulated and parameterized (in

fact, the exchange part can be given exactly). For the calculations presented in this thesis we used the parametrization of Teter [45], which has the form of a Padé approximant of the form

$$\varepsilon_{xc}^{LDA} = \frac{a_0 + a_1 r_s + a_2 r_s^2 + a_3 r_s^3}{b_0 + b_1 r_s + b_2 r_s^2 + b_3 r_s^3 + b_4 r_s^4}, \quad (2.48)$$

where r_s is the Wigner-Seitz radius, given by

$$r_s = \left(\frac{3}{4\pi n} \right)^{1/3} \quad (2.49)$$

and where a_i and b_j are parameters, which can be found in the work of Teter.

This approximation can be extended to treat magnetic materials, by taking into account the spin-polarization of the material. In order to do this, the electron density is divided in two parts: one for spin-up and one for spin-down electrons. These two densities are treated as two separate LDA calculations. The total density is then given by the sum

$$n(\mathbf{r}) = n_{\uparrow}(\mathbf{r}) + n_{\downarrow}(\mathbf{r}), \quad (2.50)$$

where n_{\uparrow} and n_{\downarrow} are respectively the spin-up and spin-down density. The magnetization (or the spin-polarization) ζ is then given by

$$\zeta = \frac{n_{\uparrow} - n_{\downarrow}}{n_{\uparrow} + n_{\downarrow}}. \quad (2.51)$$

This is called the “local spin-density approximation” (LSDA).

While being a simple approximation, the results of this approximation (either LDA or LSDA) are surprisingly good, while the densities in solids or molecules are not uniform (the density looks more like a superposition of atomic densities) and even have cusps at the nuclei. Part of this success can be attributed due the fact that LDA fulfills the sum rule 2.40 (while the exchange-correlation hole is qualitatively wrong, its spherical average is a very good approximation).

In general, LDA almost always leads to a correct picture of binding trends (across the periodic table). Also structures, bond lengths, vibrational energies, phonon spectra and other properties are predicted correctly, or with a systematic deviation. Binding energies of solids and molecules are usually overestimated, which leads to an underestimation of the bond lengths. Bandgaps are also underestimated (a notorious example of this is the bandgap of bulk Ge, which is predicted to be metallic).

2.4.3 The Generalized Gradient Approximation

In LDA one uses the knowledge of the density in a point \mathbf{r} . In real systems the density varies in space. A logical improvement of the LDA approximation would be to include also information of this rate of change in the functional. This can be done by adding gradient terms. This approach is called the “gradient-expansion approximation”. There are two serious drawbacks to this approach: the higher order corrections are difficult to calculate (if known at all) and the inclusion of the lower order gradient corrections does not improve the result much, and often even makes the result worse. For example for free atoms the energy becomes infinite [46], indicating that the sum rule 2.40 is not satisfied.

After this failure it was found that more general functionals of the density $n(\mathbf{r})$ and the gradients $\nabla n(\mathbf{r})$ could be constructed, not necessarily utilizing all orders, that were able to obtain more accurate results. In general these functionals have the form

$$E_{xc}^{GGA} = \int d^3r \varepsilon_{xc}(n, |\nabla n|, \nabla^2 n), \quad (2.52)$$

and such semilocal approximations are called “generalized gradient approximations” (GGA). Many different forms of GGAs exist, which can be roughly divided in two groups: those starting from physical principles, e.g. the much used PBE [47] form, and those that use parameters that are fitted to experimental data, e.g. the B3LYP functionals [48, 49].

2.5 Hellmann-Feynman theorem

In order to be able to calculate the equilibrium geometry of a system, one needs to be able to calculate the forces on the nuclei, as the equilibrium position is found when all forces vanish. Therefore one needs to calculate the derivative of the energy with respect to the ionic positions and this should be zero for the equilibrium position

$$\mathbf{F}_I = -\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}_I} = 0. \quad (2.53)$$

To calculate these forces, one can make use of the Hellmann-Feynman theorem [50, 51]. This theorem states that the first derivative of the eigenvalues

of a Hamiltonian H_λ , that depends on a parameter λ , is given by the expectation value of the derivative of the Hamiltonian. Symbolically:

$$\frac{\partial E_\lambda}{\partial \lambda} = \langle \Psi_\lambda | \frac{\partial H_\lambda}{\partial \lambda} | \Psi_\lambda \rangle, \quad (2.54)$$

where Ψ_λ is the eigenfunction of the Hamiltonian H_λ and E_λ the corresponding eigenvalue. This can be easily proven:

$$\frac{\partial E_\lambda}{\partial \lambda} = \frac{\partial (\langle \Psi_\lambda | H_\lambda | \Psi_\lambda \rangle)}{\partial \lambda} \quad (2.55)$$

$$= \langle \frac{\partial \Psi_\lambda}{\partial \lambda} | H_\lambda | \Psi_\lambda \rangle + \langle \Psi_\lambda | \frac{\partial H_\lambda}{\partial \lambda} | \Psi_\lambda \rangle + \langle \Psi_\lambda | H_\lambda | \frac{\partial \Psi_\lambda}{\partial \lambda} \rangle \quad (2.56)$$

$$= \langle \Psi_\lambda | \frac{\partial H_\lambda}{\partial \lambda} | \Psi_\lambda \rangle + E_\lambda \frac{\partial}{\partial \lambda} \langle \Psi_\lambda | \Psi_\lambda \rangle \quad (2.57)$$

$$= \langle \Psi_\lambda | \frac{\partial H_\lambda}{\partial \lambda} | \Psi_\lambda \rangle. \quad (2.58)$$

We can now apply it to calculate the derivative of the total energy (equation 2.4) with respect to the ionic positions

$$\mathbf{F}_I = -\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}_I} = -\langle \Psi | \frac{\partial H}{\partial \mathbf{R}_I} | \Psi \rangle \quad (2.59)$$

$$= -\int n_{\mathbf{R}}(\mathbf{r}) \frac{\partial V_{ext}(\mathbf{r})}{\partial \mathbf{R}_I} d\mathbf{r} - \frac{\partial E_{II}}{\partial \mathbf{R}_I}. \quad (2.60)$$

The same method can be used to obtain the Hessian matrix

$$\frac{\partial^2 E(\mathbf{R})}{\partial \mathbf{R}_I \partial \mathbf{R}_J} = -\frac{\partial \mathbf{F}_I}{\partial \mathbf{R}_J} \quad (2.61)$$

$$= \int \frac{\partial n_{\mathbf{R}}(\mathbf{r})}{\partial \mathbf{R}_J} \frac{\partial V_{eff}(\mathbf{r})}{\partial \mathbf{R}_I} d\mathbf{r} + \int n_{\mathbf{R}}(\mathbf{r}) \frac{\partial^2 V_{eff}(\mathbf{r})}{\partial \mathbf{R}_I \partial \mathbf{R}_J} d\mathbf{r} + \frac{\partial^2 E_{II}(\mathbf{R})}{\partial \mathbf{R}_I \partial \mathbf{R}_J}. \quad (2.62)$$

2.6 Practical calculations

In this section we discuss the practical application of the equations introduced in previous section. Crystal lattices exhibit a periodical symmetry, which can be utilized to reduce the amount of atoms one needs to consider in a calculation, by only using the unit cell. Of course nanowires are not periodic

in three dimensions, so a supercell approach should be used. The previously derived equations need to be represented in a basis and this basis should be of finite size to be able to do a computer calculation. Band structures and \mathbf{k} -point grids are also introduced. In atoms not all electrons contribute to form a bond. This observation leads to the use of pseudopotentials.

First we start with some basic crystal definitions and properties.

2.6.1 Crystal lattice and reciprocal space

A solid material consists of roughly 10^{24} electrons and ionic cores per cm^3 . In principle all these positions are needed in order to construct the Kohn-Sham Hamiltonian. But the periodic symmetry of the crystal lattice allows to reduce the problem to only those electrons and ionic cores that are contained in the unit cell, which is the smallest repeat unit. The crystal is determined when the positions in this cell are given and the set of rules to repeat this cell (by performing translations). The positions and types of the atoms in the unit cell is called the basis, while the set of translations that generate the complete crystal is called the Bravais lattice. This set of translations forms a group, as the sum of two translations is again a translation. Additional symmetries can exist (e.g. a rotation) that leave the crystal the same. So the space group of a crystal is given by the sum of the translation group and the point group.

The set of all translations forms a lattice in space, in which each translation can be written as a sum of integral multiples of primitive vectors

$$\mathbf{T}(\mathbf{i}) = \mathbf{T}(i_1, i_2, i_3) = i_1 \mathbf{a}_1 + i_2 \mathbf{a}_2 + i_3 \mathbf{a}_3. \quad (2.63)$$

In this definition (for a 3-dimensional crystal), the \mathbf{a}_j , $j = 1, 2, 3$, are the primitive translation vectors. The cell which is the most symmetric and compact is called the Wigner-Seitz unit cell. The positions of the atoms in the unit cell can then be described with respect to the primitive translation vectors.

Because of the periodicity of the lattice all periodic functions (such as the density) can be Fourier transformed. The Fourier transformed space is also called the reciprocal space. The set of reciprocal vectors \mathbf{b}_i of the primitive translations \mathbf{a}_j , which thus satisfy

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi \delta_{ij}, \quad (2.64)$$

is called the reciprocal lattice. A vector in reciprocal space is usually denoted G and is given by

$$\mathbf{G}(\mathbf{i}) = \mathbf{G}(i_1, i_2, i_3) = i_1 \mathbf{b}_1 + i_2 \mathbf{b}_2 + i_3 \mathbf{b}_3. \quad (2.65)$$

Using this reciprocal lattice the first Brillouin zone can be defined as the Wigner-Seitz cell of the reciprocal lattice.

2.6.2 Bloch theorem

The Hamiltonian \hat{H} of the Schrödinger equation of a system of independent particles operator is invariant to any lattice translation $\mathbf{T}(\mathbf{n})$, as the effective potential has the periodicity of the crystal and the derivative operator is invariant under translations. This Schrödinger equation reads

$$\hat{H}\Psi(\mathbf{r}) = \left(-\frac{1}{2}\nabla^2 + V(\mathbf{r}) \right) \Psi_i(\mathbf{r}) = \varepsilon_i \Psi_i(\mathbf{r}). \quad (2.66)$$

We can also define translation operators $\hat{T}_{\mathbf{n}}$ that act on functions by displacing the arguments,

$$\hat{T}_{\mathbf{n}}\Psi(\mathbf{r}) = \Psi(\mathbf{r} + \mathbf{T}_{\mathbf{n}}) = \Psi(\mathbf{r} + n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3). \quad (2.67)$$

Due to the invariance of the Hamiltonian to any translation $\mathbf{T}(\mathbf{n})$, the Hamiltonian commutes with the translation operators

$$\hat{H}\hat{T}_{\mathbf{n}} = \hat{T}_{\mathbf{n}}\hat{H}. \quad (2.68)$$

Therefore, the eigenstates of \hat{H} can be chosen to be eigenstates of all $\hat{T}_{\mathbf{n}}$ simultaneously. The eigenstates of the translation operator can be found independent of the details of the crystal, so they can block diagonalize the Hamiltonian, thereby classifying the states by their eigenvalues of the translation operators.

As mentioned earlier, the translations form a group, so

$$\hat{T}_{\mathbf{n}_1}\hat{T}_{\mathbf{n}_2} = \hat{T}_{\mathbf{n}_1+\mathbf{n}_2}. \quad (2.69)$$

A similar condition is valid for the eigenvalues $t_{\mathbf{n}}$ and eigenstates $\Psi(\mathbf{r})$ of the operators $\hat{T}_{\mathbf{n}}$, thus

$$\hat{T}_{\mathbf{n}_1}\hat{T}_{\mathbf{n}_2}\Psi(\mathbf{r}) = t_{(\mathbf{n}_1+\mathbf{n}_2)}\Psi(\mathbf{r}) = t_{\mathbf{n}_1}t_{\mathbf{n}_2}\Psi(\mathbf{r}). \quad (2.70)$$

Each translation can be composed of a product of primitive translations, thus $t_{\mathbf{n}}$ can be written as a product of $t(\mathbf{a}_i)$

$$t_{\mathbf{n}} = [t(\mathbf{a}_1)]^{n_1} [t(\mathbf{a}_2)]^{n_2} \dots, \quad (2.71)$$

and since the modulus of each $t(\mathbf{a}_i)$ should be unity

$$t(\mathbf{a}_1) = e^{i2\pi y_i}. \quad (2.72)$$

Because there are periodic boundary conditions, $(t(\mathbf{a}_1))^{N_i}$ is equal to 1, so that $y_i = 1/N_i$. Using equation 2.64 this becomes

$$t_{\mathbf{n}} = e^{i\mathbf{k}\cdot\mathbf{T}_{\mathbf{n}}}, \quad (2.73)$$

where \mathbf{k} is a vector in reciprocal space

$$\mathbf{k} = \frac{n_1}{N_1}\mathbf{b}_1 + \frac{n_2}{N_2}\mathbf{b}_2 + \dots \quad (2.74)$$

\mathbf{k} can be restricted to the primitive cell and there are exactly the same number of values of \mathbf{k} as there are cells. Combining equation 2.70 and 2.73 gives

$$\hat{T}_{\mathbf{n}}\Psi(\mathbf{r}) = \Psi(\mathbf{r} + \mathbf{T}_{\mathbf{n}}) = e^{i\mathbf{k}\cdot\mathbf{T}_{\mathbf{n}}}\Psi(\mathbf{r}), \quad (2.75)$$

which is the Bloch theorem, that states that eigenstates of the translation operator vary from one cell to another with the phase factor given by equation 2.75. Applied to the eigenfunctions of the Kohn-Sham Hamiltonian this gives

$$\Psi_{i,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_i(\mathbf{r}), \quad (2.76)$$

where $u_i(\mathbf{r})$ is periodic $u_i(\mathbf{r} + \mathbf{T}_{\mathbf{n}}) = u_i(\mathbf{r})$ and where \mathbf{k} can be used to label the states.

2.6.3 Basis sets

Before being able to solve the Kohn-Sham equations numerically using a computer program (ABINIT in our case), we must be able to represent the wavefunction. For that we need a basis that is both computational efficient and that allows to control the accuracy. Such a basis is formed by plane waves, which have the advantage that it is very convenient to switch between

the real space and the reciprocal space during the calculations, by performing fast Fourier transforms. There exists alternatives that use localized basis functions e.g. by using Gaussian wave functions or numerical orbitals, but these approaches have the draw back that these bases are not complete so that it is not clear if a certain calculation is converged with respect to the basis size. Their advantage is that they contain much less basis functions, so that calculations will be performed faster. Another advantage is that the introduction of a vacuum layer is much less computational demanding (due to the localized nature of the basis functions) than in the plane wave case.

The cell periodic part of equation 2.76 $u_i(\mathbf{r})$ can be expanded in the basis of plane waves

$$u_i(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{k}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}, \quad (2.77)$$

where the \mathbf{G} vectors are vectors in the reciprocal lattice. The Kohn-Sham orbitals then become

$$\Psi_{i,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_i(\mathbf{k} + \mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}. \quad (2.78)$$

Thus, the basis for a given \mathbf{k} is discrete, but in principle still infinite. However, the coefficients for the plane waves $u_{i,\mathbf{k}}(\mathbf{r})$, have a kinetic energy of $(\mathbf{k} + \mathbf{G})^2/2$. Typically the plane waves with lower energies play a more important role than those with higher energies. Therefore the number of plane waves can be restricted by placing an upper boundary to the kinetic energy of the plane waves. This boundary is called the cutoff energy E_{cut} and the restriction is thus given by

$$\frac{|\mathbf{k} + \mathbf{G}|^2}{2} < E_{cut}. \quad (2.79)$$

2.6.4 Band structures and \mathbf{k} -point grids

In the case of an infinite crystal, the spacing of the \mathbf{k} -points goes to zero, so that \mathbf{k} can be considered a continuous variable. For each of these values a discrete spectrum of eigenvalues of the Kohn-Sham equation exists. Due to the translational symmetry, it is sufficient to only consider the first Brillouin zone. Due to the continuous nature of \mathbf{k} , this leads to bands of eigenvalues and energy gaps where there can be no eigenstates. The plot of these eigenvalues is called a bandplot.

To calculate many properties, such as the electronic density, it is necessary to integrate over all \mathbf{k} -points in the Brillouin zone. As electronic states of \mathbf{k} -vectors close to each other are very similar, it is possible to replace the integral by a discrete sum over special chosen \mathbf{k} -points. The number of these \mathbf{k} -points necessarily depends on the material: for isolators only a few points are needed as all bands are filled, while for metals more points are needed for the bands that cross the Fermi level. The presence of additional symmetries, such as rotations or mirror reflections, allows to consider only a part of the Brillouin zone. The smallest possible part, that by applying all symmetries can be mapped to the complete Brillouin zone is called the irreducible Brillouin zone (IBZ).

An efficient method to choose the grids of \mathbf{k} points to be used to do the summations was proposed by Monkhorst and Pack [52]. They presented a simple formula that gives uniform grids, valid for any crystal:

$$\mathbf{k}_{n_1, n_2, n_3} = \sum_i^3 \frac{2n_i - N_i - 1}{2N_i} \mathbf{G}_i. \quad (2.80)$$

This grid will exactly integrate a periodic function that has only Fourier components that extend only to $N_i \mathbf{T}_i$ in each direction. For many crystal types it is best to set N_i even, so that the resulting grid does not contain the highest symmetry points (such as $\mathbf{k}=0$, also called the Γ point). The advantage is that the same sampling of the Brillouin zone is achieved, using less \mathbf{k} points.

2.6.5 Pseudopotentials

Now all ingredients are present to solve the Kohn-Sham equations. But a full electron calculations will still be computationally very demanding. This is caused by the localized features, such as the tightly bound core orbitals and the highly oscillating nature of the valence electrons. To describe these features a huge number of plane waves are required as real space details are inversely proportional to the largest wavevector norm. However, since most physical properties of solids only depend on the valence electrons, one can distinguish between the core and valence electrons. This is the basis of the pseudopotential approach. In this approach the core electrons and the ionic core are replaced by a pseudopotential, that acts on a set of pseudo wavefunctions. This pseudopotential will be constructed so that the pseudo wavefunction has no radial nodes within the core region and so that after a certain

radius r_{cut} the pseudo wavefunction becomes equal to the real wavefunction. Of course the pseudopotential should preserve the atomic properties. Because the phase shifts occurring when scattering across the core depend on the angular momentum state, the pseudopotential must be non-local. It is also a requirement that the pseudopotential at r_{cut} is continuous, also in its first and second derivative (sometimes higher order continuity is imposed).

In order to obtain a pseudopotential, the all-electron Schrödinger equation has to be solved:

$$\left(-\frac{1}{2}\nabla^2 + V\right)\Psi_l^{AE} = \varepsilon\Psi_l^{AE}, \quad (2.81)$$

where Ψ_l^{AE} is the all-electron wave function with angular momentum l . Once the valence eigenvalues are obtained with this equation, they are inserted again in the Schrödinger equation, but with the pseudized wave functions. Inverting then yields the pseudopotential. As there is a lot of freedom when generating the pseudo wavefunctions, different methods exist to create a pseudopotential.

In this thesis we make use of the Troullier-Martins [53] form of pseudopotentials. Fig. 2.1 shows the pseudized Troullier-Martins wavefunctions generated using the Opium package ¹. A Si atom contains 14 electrons, which are ordered in the following orbitals: $1s^2 2s^2 2p^6 3s^2 3p^2$. As only the outer electrons are important for many of the properties of the Si atom, the pseudopotential only contains explicitly the $3s$, $3p$ and $3d$ (which is empty in the ground state configuration). The vertical lines indicate the values of the cutoff radius r_{cut} for each of these orbitals. As can be seen from the figure, the all-electron wavefunctions match the pseudized ones for larger radii. One can also see another advantage of using pseudopotentials, as the nodes present in the $3s$ and $3p$ orbitals are not included in the pseudopotential. As the Troullier-Martins pseudopotential is norm conserving, the integrated charge density of the core is equivalent to the integrated all-electron density.

¹The Opium package can be freely downloaded from <http://opium.sourceforge.net/>.

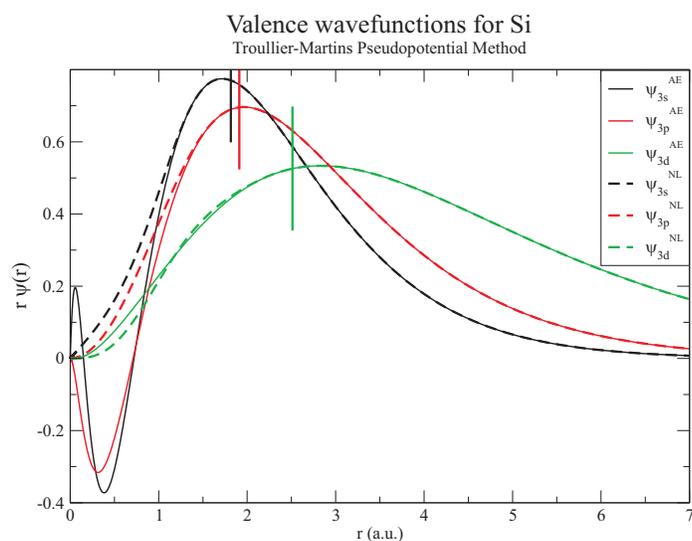


Figure 2.1: The Si all electron wavefunctions Ψ^{AE} and the pseudized wave functions Ψ^{NL} (using the Troullier-Martins form) for the s , p and d orbitals. The graph is created using the Opium package.

3

Electronic properties of B and P doped semiconductor nanowires

3.1 Silicon nanowires

3.1.1 Introduction

In the first chapter we showed that silicon nanowires (SiNWs) have already shown applications as electronic devices, including FETs [4], p-n diodes [5, 6], bipolar junction transistors [6] and as chemical and biological sensors [9]. This wide range of applications can be achieved because the growth direction, diameter, passivation (oxidation or hydrogenation) and doping can be controlled during the synthesis of SiNWs [35, 54, 55]. Current growth methods like the vapor-liquid-solid growth method [22, 56] (see also introduction) and the oxide-assisted growth method [57] can fabricate SiNWs with diameters as small as 1.3 nm and up to several tens of nanometers. Detailed analysis shows that the smaller diameter wires grow along the [110] direction, while larger diameter wires are oriented along the [112] or [111] directions [24].

Previous *ab initio* theoretical investigations concentrated mainly on the structural and electronic properties of pristine [58] and H-passivated wires [59]. There are many ways that the electronic structure of the wire can be changed. In Ref. [58] it was shown how the bandgap depends on the surface layer, while in Ref. [60] the size of the band gap was shown to increase with decreasing wire diameter. Recently, in Ref. [61] different examples of the influence of the morphology on the band gap were demonstrated.

Alternatively, one can change the electronic properties through doping of the SiNW, as shown in Refs. [35, 54, 55, 62]. Being able to selectively control the bandgap and the Fermi level allows to control properties like photolu-

CHAPTER 3. ELECTRONIC PROPERTIES OF B AND P DOPED SEMICONDUCTOR NANOWIRES

minescence and conductivity is of primary importance for the development of new applications. Therefore a detailed understanding of the electronic properties of doped SiNWs and the localization of the dopants is important. In the first part of this chapter we report on an *ab initio* study of thin (1.2 and 1.6 nm) B, P and BP codoped Si wires. We focus on: i) the formation energy to create doped wires, ii) the localization of the dopants in the wire, and iii) the changes in the band gap due to doping. Because the studied wires are very thin, they have a large surface to volume ratio, and therefore the number of dopants can be even smaller or equal to the number of dangling bonds (corresponding to Si atoms at the surface which are not saturated with a hydrogen atom). These dangling bonds are chemical very reactive. So they will have an impact on the electronic properties. The effect of these dangling bonds on the doped wires is also investigated.

We considered doped H-passivated wires grown in the [110] direction and with a diameter of 1.2 and 1.6 nm, containing 28 and 62 atoms respectively in a single unit cell. A single unit cell of both wires is shown in Fig. 3.1. As dopants boron and phosphorus are considered, as these are also experimentally used. Boron will act as an electron acceptor, as it lacks one electron, while phosphorus has an additional electron and will therefore act as an electron donor.

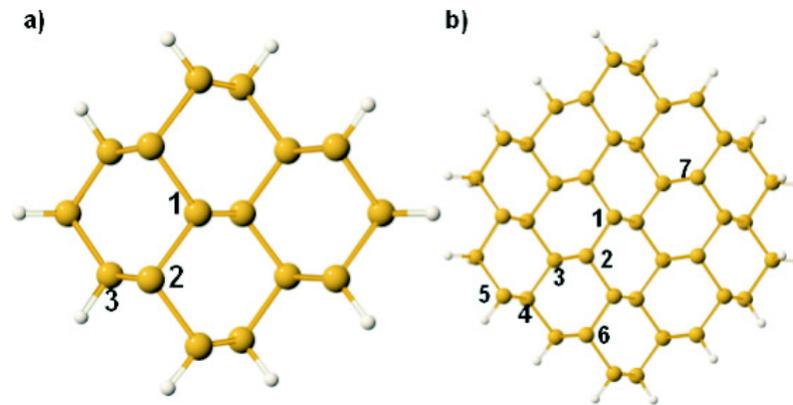


Figure 3.1: A single unit cell of the considered wires. a) The 1.2 nm [110] wire consists of 16 Si and 12 H atoms. b) The 1.6 nm wire consists of 42 Si and 20 H atoms.

3.1.2 Computational details

All calculations are performed using density-functional theory (DFT) within the local density approximation (LDA) for the exchange-correlation functional. The present results have been obtained through the use of the ABINIT code [63]. Troullier-Martins [53] pseudopotentials are used. A kinetic energy cutoff of 20 Hartree was used for the undoped and P-doped SiNWs, and of 30 Hartree for the B-doped SiNWs. A 1x1x4 Monkhorst-Pack [52] grid is used and the position of the atoms are relaxed without any symmetry constraint. In the x and y direction a layer of vacuum space is used to create an infinite one-dimensional system.

In order to study the localization of the dopants in the wires, the formation and the segregation energy are calculated. The formation energy E_f of the impurity B and/or P can be defined as the energy needed to insert an atom B and/or P (taken from a reservoir) into the wire after removing one (or more) Si atoms (to a reservoir, assumed to be bulk Si)

$$E_f = E(\text{doped SiNW}) - E(\text{SiNW}) + (n + m)\mu_{Si} - n\mu_B - m\mu_P, \quad (3.1)$$

where n and m are the number of respectively B and P dopants, μ_{Si} the chemical potential of Si (here set equal to the total energy per atom of bulk Si) and $\mu_{B/P}$ the chemical potential of an atom of the impurity. For these chemical potentials we use the total energy per atom in the tetragonal B_{50} crystal for B and the orthorhombic black phosphorous for P.

The segregation energy E_s is defined as the energy difference between a SiNW doped in the center and one doped at the edge of the wire

$$E_s = E(\text{centered doped SiNW}) - E(\text{edge doped SiNW}). \quad (3.2)$$

3.1.3 Results

Dopant localization

First the preferred positions of the dopants (B or P) in the smallest SiNW ($d=1.2$ nm) are determined and the effect of a change in concentration (achieved by enlarging the unit cell) is studied. Each structure is allowed to relax without any symmetry constraint. After the structural relaxation we calculated the formation and segregation energy. The labels corresponding to the positions are shown in Fig. 3.1(a). The results of this calculation

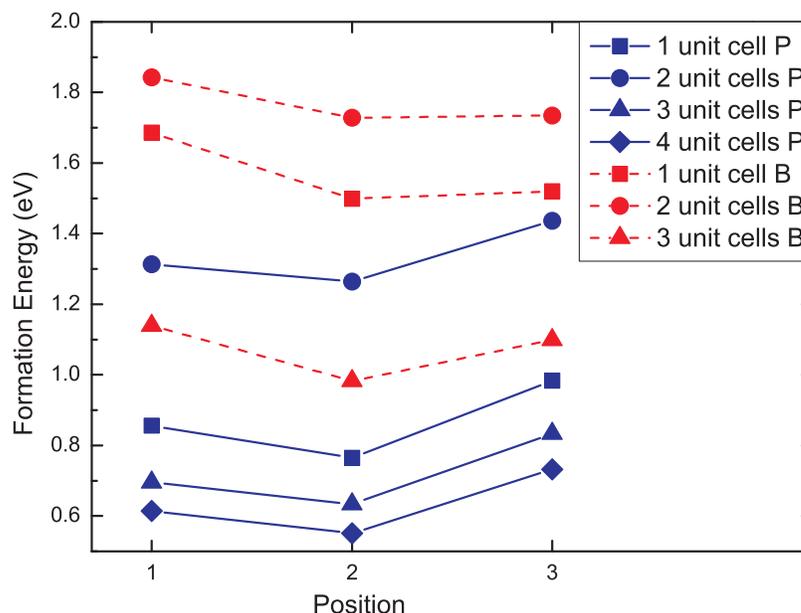


Figure 3.2: The formation energy as function of the position (labeled according to Fig. 3.1 (a)) and as function of the size of the unit cell (and thus of concentration) for the 1.2 nm wire, for both B and P doping.

are shown in Fig. 3.2. It is clear that the position labeled 2 (on the edge, only binding Si atoms) is the preferential position for the dopant in both cases. The formation energy for P doping is smaller than for B doping. One can also see that the formation energy for a double unit cell, but only containing one dopant (resulting in a lower dopant concentration) is much higher than the formation energy for the single unit cell or the triple unit cell. This indicates that at least two competing effects occur. One is an interaction between the dopants, which enlarges the formation energy. The other is the possibility to form a resonance structure (like in benzene) of electrons (in case the dopant is P) or holes (in case the dopant is B) when the dopants are close to each other (as is the case in the single unit cell). The last effect is dominant in the case of the single unit cell, effectively lowering the formation energy. For a double unit cell the importance of this effect diminishes and the first effect becomes dominant. For the triple unit cell the interaction diminishes (and the resonance does not influence the system anymore), and the formation energy becomes smaller in comparison to the double unit cell case.

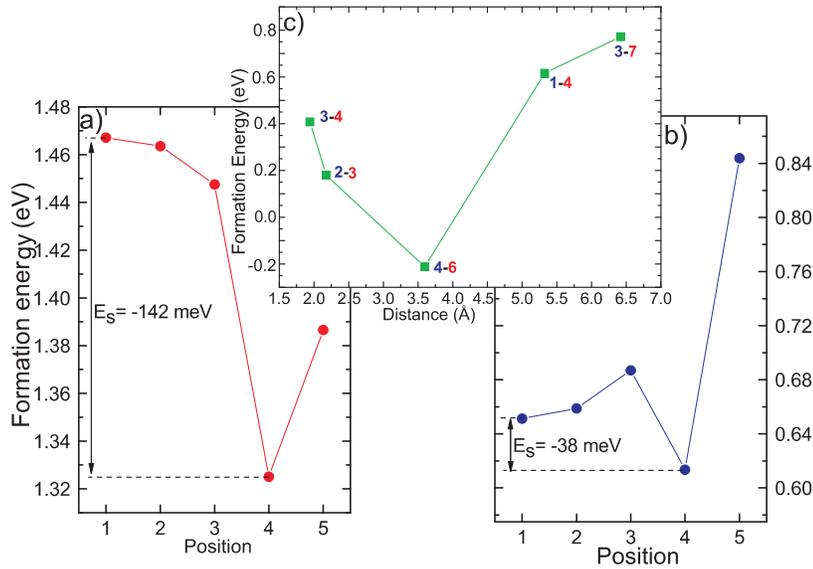


Figure 3.3: The formation energy as function of the position (labeled according to Fig. 3.1) for the 1.6 nm wire for (a) B doping, (b) P doping and (c) BP codoping.

For the thicker nanowire we observe that the edge position remains the preferred position. This is shown in Fig. 3.3. The formation energy for B doping is also in this case larger than the formation energy for P doping, as is the segregation energy. Another effect that can be seen is that for the B doped wire the formation energy decreases when the dopant is located closer to the surface. For the P doped wire the opposite is observed (when ignoring the lower formation energy for the preferential position). For BP codoping (see the formation energy in Fig. 3.3c as function of the distance between both dopants) we found that both dopants prefer to sit on the preferential edge sites and close together.

The reason why these near edge positions are energetically favorable is a relaxation effect: in the more centrally located positions the allowed relaxation is smaller due to the stability of the Si structure. Fig. 3.4 compares the formation energy for the unrelaxed and relaxed single and codoped 1.2 nm wire. It shows that the relaxation has a large effect for the B doped wire, and a much smaller effect on the P doped wire. A similar behavior was found recently in Si nanoclusters [64]. This also explains the tendency of B atoms to prefer positions closer to the edge as these positions allow more

relaxation, and thus a lowering of the formation energy. Another remarkable effect, similar to what was found in doped Si nanoclusters [65], that can be seen from Fig. 3.4 is the fact that BP codoping has a lower formation energy than single B or P doping.

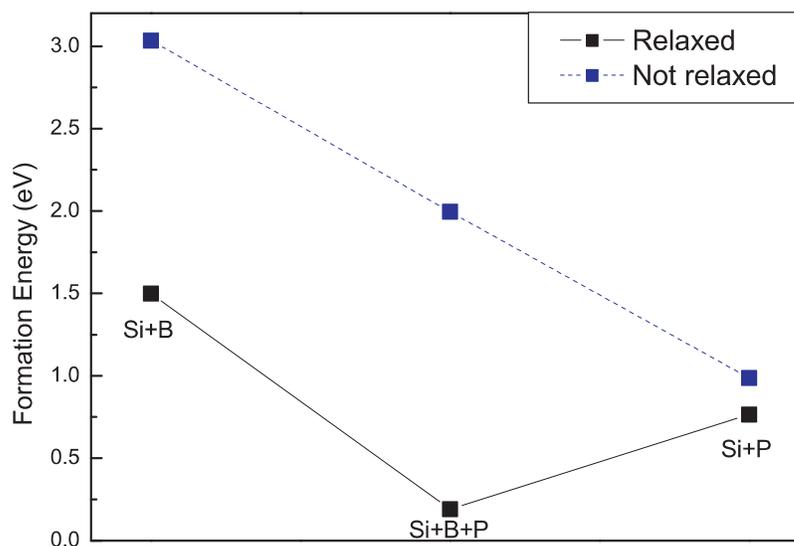


Figure 3.4: Formation energy for the unrelaxed and relaxed doped 1.2 nm wire. The dopants are placed on the preferred positions.

From studies on doped Si nanoclusters [65] one may expect that the bandgap will change when doping the SiNW. Fig. 3.5 shows the electron density of states (DOS) for the smallest wire (calculated for a single unit cell using the tetrahedron method), for an undoped, P doped, B doped and BP codoped wire. The zero energy point is taken at the calculated Fermi energy (highest occupied energy level). Doping reduces the bandgap due to an additional electron (donor) band (in case of P doping) or hole (acceptor) band (in case of B doping). We found that codoping leads to the largest reduction of the bandgap.

Dopant localization in the presence of a dangling bond defect

Up to now, fully H-passivated SiNWs were studied. If one considers a typical dopant concentration of $5 \cdot 10^{18} \text{ cm}^{-3}$ and an estimate of 10^{12} cm^{-2} [66] dangling bonds at the surface, one finds that for wires with a radius smaller than

3.1. SILICON NANOWIRES

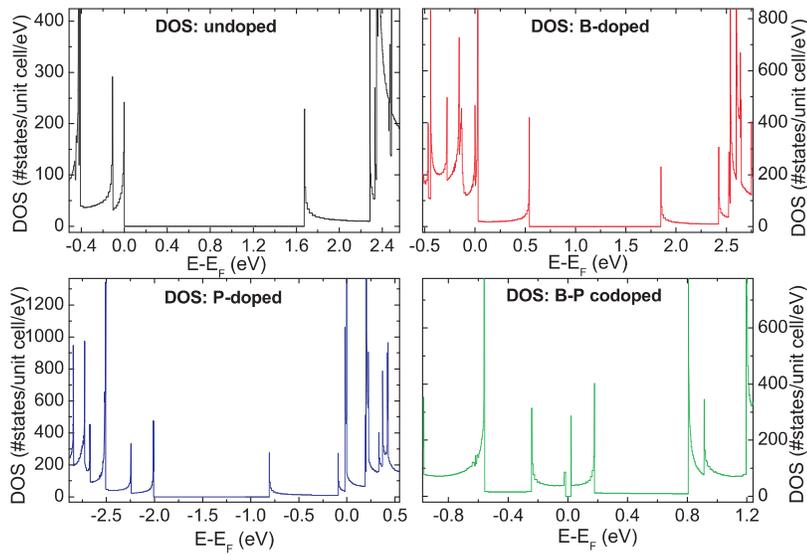


Figure 3.5: Density of states for a single unit cell of the undoped, B doped, P doped and BP codoped 1.2 nm [110] wire.

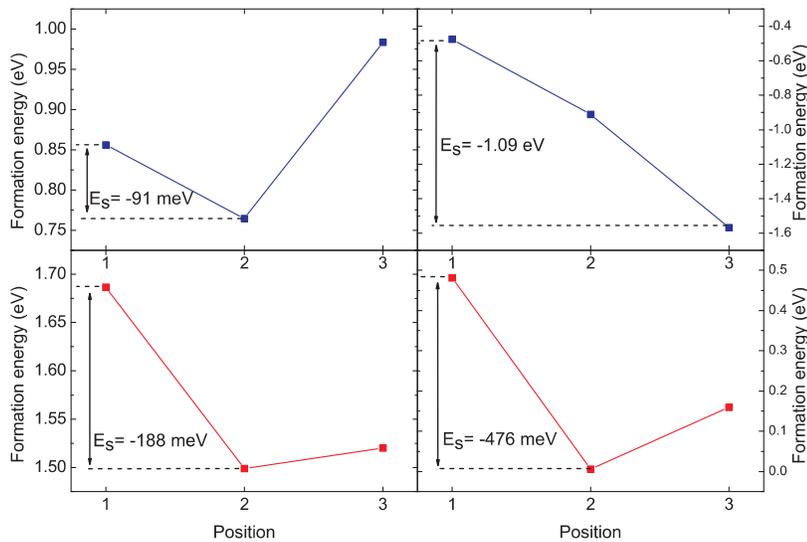


Figure 3.6: Formation energies for the B and P doped 1.2 nm wire (left) and these of the same wire, but with a dangling bond (right). The upper figures are those for the P doped wire, the lower for the B doped wire.

CHAPTER 3. ELECTRONIC PROPERTIES OF B AND P DOPED SEMICONDUCTOR NANOWIRES

4 nm, there are more dangling bonds than dopant atoms. This illustrates the importance of studying the effect of dangling bonds on the formation and segregation energy for thin SiNWs. We create a dangling bond by removing a H atom and the resulting formation energy diagram is shown in Fig. 3.6. The presence of a dangling bond lowers the formation energy in both cases, thus making it easier to dope the wire. The preferred position for a P dopant changes to the position near the dangling bond. The segregation energy to this position is -1.09 eV. This large segregation energy indicates that the dangling bond will attract the P-type dopants. This “trapping” of the dopant renders it electronically inactive as can be seen from Fig. 3.7 (with label P3): the wire will have a semiconductor character with a large bandgap. Now the question arises if the shallow donor electron of P would not always passivate the dangling bond, even if the P atom was located in the center of the wire. In Fig. 3.7 (with label P1) one can see however that if the dopant would remain in the center, the bandgap becomes much smaller. In case of doping with a B atom, the preferred position remains position 2. The segregation

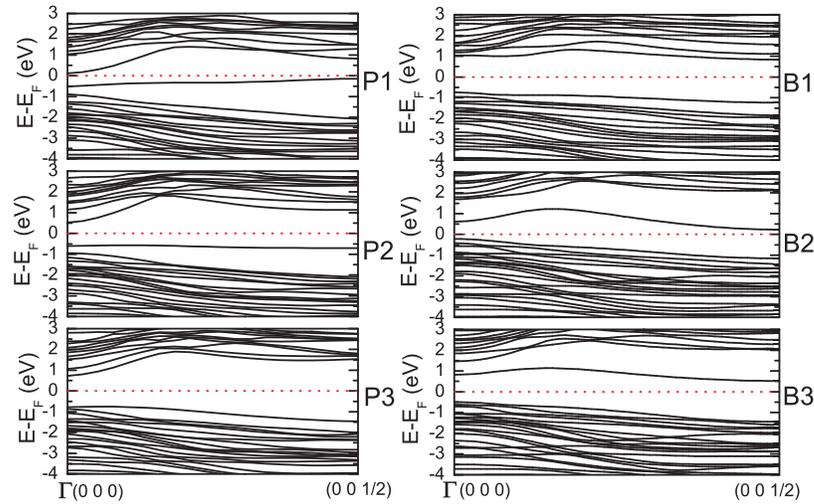


Figure 3.7: Bandplots of the B (right) and P (left) doped 1.2nm wire, containing 1 DB. The plots are labeled according to Fig. 3.1(a).

energy is equal to 476 meV. This is only half of the segregation energy for a P dopant. So for the same concentration of dopants, the fraction of P atoms that will be trapped is larger than the fraction of B atoms. So the concentration of free carriers will be larger for B doping than for P doping. This

is in agreement with experimental conductance measurements which show a higher resistivity for P doped Si wires than for B doped Si wires, for the same level of concentration [35, 36] and with the recent theoretical results of Ref. [62].

3.2 Germanium nanowires

3.2.1 Introduction

As explained in the introduction, also nanowires composed of germanium can have important applications. Germanium NWs are particularly interesting because of its compatibility with Si and their higher carrier mobilities. Also for this kind of wires knowledge of the dopant properties is essential to understand its possible applications.

Previous theoretical research focused mainly on the electronic [67] and optical properties [68, 69] of free standing Ge NWs and the effect of quantum confinement on these properties. Doped wires have not yet been theoretically explored, while it is well known that dopants can alter the electronic properties substantially [70]. Here we focus on the properties of doped Ge nanowires. As dopants we consider boron (p-type NW) and phosphorus (n-type NW).

We consider hydrogen passivated wires, which can be obtained experimentally using the vapor-liquid-solid growth method or the oxide-assisted growth method [22, 56, 57]. These wires grow dominantly along the [111] direction [25]. Growth along the [110] direction is also observed [25, 26].

We study: i) the formation energy needed to create the doped wires, ii) the preferred position of the dopants, and iii) the changes that occur when a dangling bond is introduced. These results are compared with the results on doped Si nanowires, obtained in the previous section.

3.2.2 Results

Dopant localization

For our numerical study we considered wires in the [110] direction with a diameter of 1.2 and 1.6 nm, containing 28 and 62 atoms in a unit cell, respectively. These unit cells are shown in Fig. 3.8. We first study the most

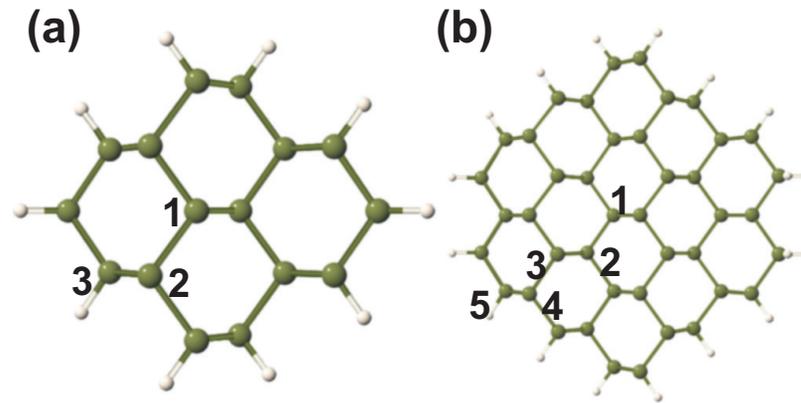


Figure 3.8: Relaxed unit cell of NW: (a) the 1.2 nm [110] wire consists of 16 Ge and 12 H atoms, and (b) the 1.6 nm wire consists of 42 Ge and 20 H atoms.

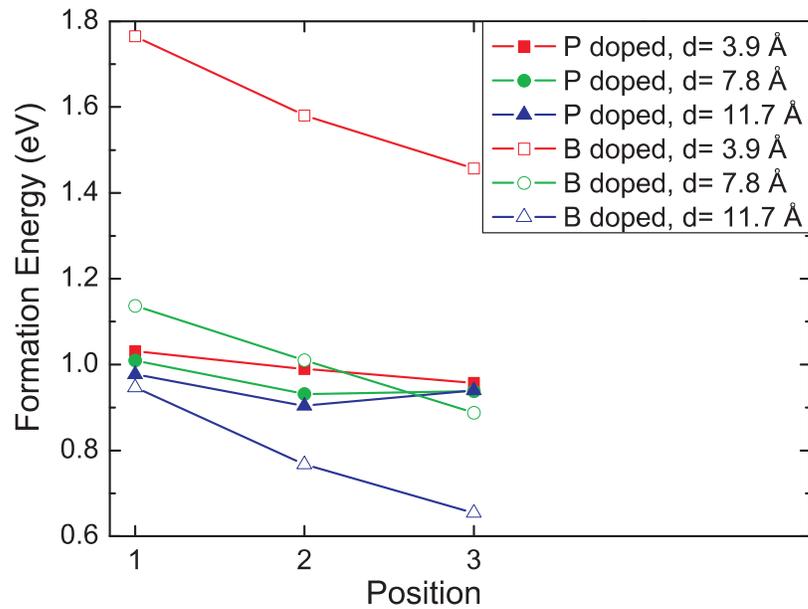


Figure 3.9: The formation energy of the 1.2 nm wire as function of the dopant position (see Fig. 3.1(a)) and the dopant separation d along the wire axes.

likely location of B and P dopants in these wires. This location can be found by comparing the formation energies. In Fig. 3.9 the formation energy is

3.2. GERMANIUM NANOWIRES

shown for the different possible locations in the 1.2 nm wire, for different distances (in the wire direction) between the dopants (here 3.9 Å, 7.8 Å and 11.7 Å). The positions are labeled as shown in Fig. 3.1(a). This figure shows that the formation energy decreases when a B dopant is located closer to the edge of the wire. The corresponding segregation energy is 0.3 eV. A similar trend is found for P dopants, but much smaller ($E_s = 70$ meV). This effect occurs due to relaxation. Indeed the Ge cage-like structure (zinc blende) is more robust in the center regions, so that the amount of allowed relaxation is limited. For the impurity in the edge position the structure can be deformed more easily, which allows for an increased relaxation lowering the total energy. This is illustrated in Fig. 3.10, which shows the formation energy for B and P doped 1.2 nm nanowire for an unrelaxed structure (using the structure of the relaxed undoped wire as starting atom positions) and a fully relaxed structure. The effect of relaxation is much larger (1.3 eV for the center location and 1.7 eV for the edge location) for B dopants than for P dopants (0.17 and 0.3 eV respectively for the center and the edge). This difference is related to the relative atom size: B atoms are smaller than P atoms, which are smaller than Ge atoms. Thus B dopants can benefit more from lattice relaxation. When the distance in the wire direction between the dopants increases, the formation energy decreases, because interactions between dopants also decrease with distance. Please note that, while the dopant concentrations are very high, this doesn't influence our conclusions, because the trends are the same for all considered dopant concentrations. Fig. 3.11 shows the formation energy for the different dopants in case of the thicker (1.6 nm) wire (full lines). The qualitative behavior is similar to that of the 1.2 nm wire.

Here we notice differences with equivalent Si wires. For Si the preferred position was always the near edge position (position 2 or 4, respectively for the 1.2 and 1.6 nm wire). For Si two competing interactions were observed when the axial distance between the dopants was increased: a repulsive and an attractive one (for very small distances), while for Ge there is only a repulsive interaction.

Dangling bond defects

A dangling bond (DB) is a common defect occurring in hydrogenated nanowires. This defect occurs when one of the passivating hydrogen atoms is missing. A reasonable estimation for the concentration of these DB defects at a sur-

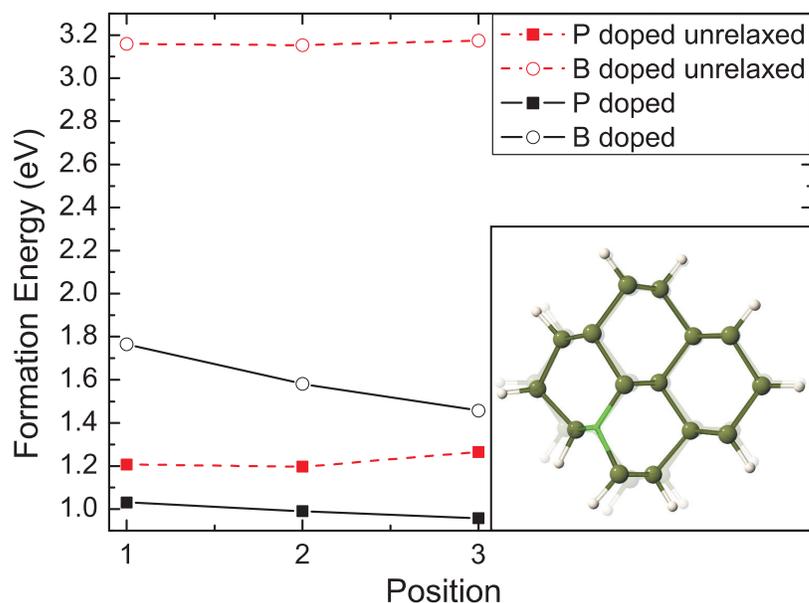


Figure 3.10: The formation energy of the smallest wire (1.2 nm) for a relaxed and unrelaxed structure. The effect of the relaxation lowers the formation energy and this effect is larger for a P doped wire than it is for a B doped one. The effect is also larger for position 3 (the edge location). The inset shows the relaxed structured for a wire with a B dopant on position 2. The transparent image shows the unrelaxed structure.

face, based on this concentration for Si surfaces, is 10^{12} cm^{-2} [66]. When considering a typical dopant concentration $5 \cdot 10^{18} \text{ cm}^{-3}$, the number of DB can be larger than the number of dopants for wires with a diameter smaller than 4 nm. We will now investigate the effect that such a DB has on the preferred positions of the dopants. Therefore we calculated the formation energies for the dopants at different positions in the presence of a missing hydrogen atom. The results of these calculations are shown for the 1.2 nm wire in Fig. 3.12 and for the 1.6 nm wire in Fig. 3.11 (dashed lines).

This figure shows that the effect of a DB is very pronounced for P dopants: the formation energy of these dopants is decreased considerable near the edge in comparison with the center position. The formation energy for B dopants is the lowest in a near edge position (position 2 in the 1.2 nm wire or position 4 in the 1.6 nm wire). As can be seen from the inset of Fig. 3.12, which shows the relaxed unit cell in case of a B dopant on position 2, the difference

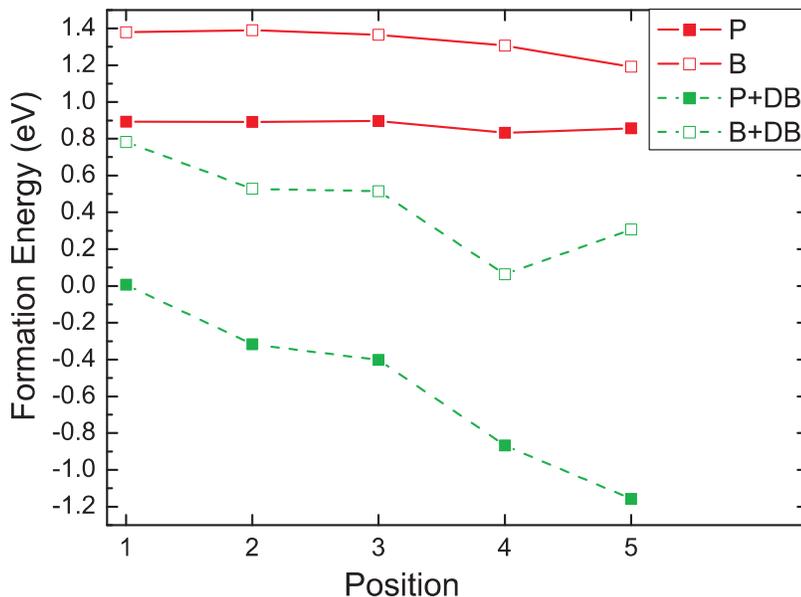


Figure 3.11: The formation energy for the 1.6 nm wire in the presence and absence of a dangling bond as function of the dopant position (see Fig. 3.1(b)).

between position 2 and position 3 is very small, so we will use this position to calculate the segregation energy. When the dopant separation in the wire direction increases, the formation energy decreases, as was the case in the wires without DB, but the segregation energy remains of the same order of magnitude. The value for this segregation energy (for a dopant separation of 7.8 \AA) for a P dopant is of the order of 1 eV, while for B dopants it is of the order of 0.5 eV. This indicates that for similar dopant concentrations the fraction of P dopants near the edge will be larger than the fraction of B dopants. When a dopant is located near the DB, this DB will “trap” the additional electron (for P) or hole (for B dopants), thus reducing the number of free carriers, increasing the resistivity. This is similar to the behavior of Si nanowires, where this effect was calculated [62] and measured [35, 36].

3.3 Experimental verifications

The fact that dopant positions near the edge of the nanowires are more favorable than positions in the center of the wire is recently also measured

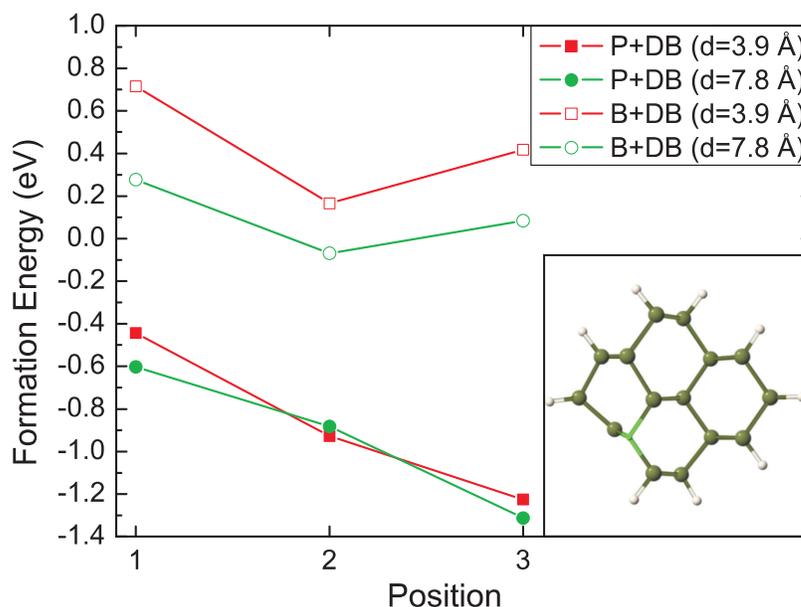


Figure 3.12: The formation energy of the 1.2 nm wire, in the presence of a DB defect, given as function of the dopant position for two different dopant separations along the wire axis ($d=3.9$ and 7.8 nm). The inset shows the relaxed unit cell for a B dopant placed at position 2.

experimentally. These experiments should be conducted carefully to avoid deposition of the dopants on the sides of the wires using a vapor/solid type of growth. This results in tapered nanowires, where the concentration of dopants is larger at the edges than in the center due to the different growth proces. An indication of this different growth proces is that the concentration of dopants is higher near the base of the nanowire than close to the catalyst, which was measured by Raman scattering studies [71, 72]. These kind of structures are not relevant to our calculations. Furthermore some of the measurement methods used (e.g. high resolution secondary ion mass spectroscopy (for gold impurities [73]) or atom probe tomography [74, 75]), that provide nanometer scale distribution information, place additional limitations to the growth proces, due to the vertical geometry needed for these measurements. However in a recent work Xie *et al.* [76] performed measurements on n- and p-type Si, and n-type Ge nanowires grown using the standard VLS method. Transport measurements of these nanowires configured as FETs in combination with repeatedly low temperature surface oxidation

and selectively surface oxide etching were used in this study. Because an oxidation/etching cycle induces only small changes in FET properties for uniformly doped nanowires, while it will induce large changes for surface dopants, it can be used to distinguish between these two profiles. It was shown that the dopants are indeed preferentially located at the edges for wires with diameters smaller than 22 nm. For larger diameters doping is similar to bulk doping, as the surface to volume ratio becomes much smaller and not all dopants can be incorporated at the surface.

3.4 Conclusions

To conclude, in this chapter we studied the formation energies, segregation energies and the density of states of B, P and BP doped hydrogen passivated silicon and germanium nanowires using ab initio calculations. We showed that in general the preferred position for a dopant atom is near the edge of the wire (the most external lattice site with only Si or Ge nearest neighbors), as this position allows a larger relaxation of the lattice. This decrease in formation energy due to relaxation is the largest in case of B dopants. We also pointed out some peculiarities due to the high amount of doping in the considered wires. Selective doping of the wires allows to change the bandgap. When a dangling bond is present, P dopants will segregate to the edge of the wire, thereby trapping its additional electron and thus lowering the conductance. For B dopants this effect is smaller, which can explain the experimental measured lower conductivity of P doped Si nanowires for similar concentrations of dopants. Recent experiments confirm the result that edge positions are more favorable than center positions.

4

Density Functional Perturbation Theory

4.1 Phonons: Introduction

Phonons are quantized modes of vibration, occurring in crystals and solids. They are the quantum mechanical analogon of normal modes. The name phonon is coming from the Greek $\varphi\omega\nu\eta$ (phonē), meaning voice, as the long wavelength phonons give rise to sound in solids. Phonons are considered to be quasi particles with wave vector \mathbf{q} and angular frequency ω . Their energy is given by $\hbar\omega$ and their momentum by $\hbar\mathbf{q}$. The angular frequency can be given as $\omega = \omega(\mathbf{q})$, which gives rise to a phonon band spectrum. The phonon modes that have zero frequency at $\mathbf{q} = 0$ are called the acoustic modes, which correspond in a typical 3 dimensional solid to collective translations of the atoms in the three directions. Phonons that have non zero frequency at $\mathbf{q} = 0$ are called optical phonons. This naming is coming from the fact that the collective modes carry sound waves, thus they are called acoustic, while the optical modes are modes that can be excited, for example by electromagnetic radiation (e.g. infrared radiation) in ionic lattices, thus they are called optical. In these modes, the atoms of sublattices typically move in opposite directions. An additional distinction of the modes that can be made is if they are longitudinal, which indicates that all atoms oscillate in the wave direction, or transversal, in which the atoms oscillate perpendicular to the wave direction. The total number of modes is equal to the number of atoms N multiplied by the number of degrees of freedom (3).

4.2 Phonons from lattice dynamics

The phonon frequencies can be obtained from lattice dynamics. The position of nucleus I is given by

$$\mathbf{R}_I = \mathbf{R}_l + \boldsymbol{\tau}_s, \quad (4.1)$$

where \mathbf{R}_l is the position of the l th unit cell in the Bravais lattice and $\boldsymbol{\tau}_s$ the equilibrium position of the atom in the unit cell. In the Born-Oppenheimer approximation the nuclei can be considered as being classical particles. For each of the nuclear positions a Newton equation can be written

$$M_I \frac{\partial^2 \mathbf{R}_I}{\partial t^2} = - \frac{\partial}{\partial \mathbf{R}_I} E(\mathbf{R}_I), \quad (4.2)$$

where we are still using the convention that capital letters indicate the atomic nuclei and $E(\mathbf{R}_I)$ is the Born-Oppenheimer energy surface. At a time t we displace the atoms from their equilibrium position. This displacement is given by the vector $\mathbf{u}_I(t)$. If the displacement is small, we can expand the Born-Oppenheimer energy surface by using a Taylor expansion:

$$E(\mathbf{R}_I + \mathbf{u}_I) = E(\mathbf{R}_I) + \sum_{I\alpha} \frac{\partial E}{\partial u_{I\alpha}} u_{I\alpha} + \frac{1}{2} \sum_{I\alpha, J\beta} \frac{\partial^2 E}{\partial u_{I\alpha} \partial u_{J\beta}} u_{I\alpha} u_{J\beta} + \dots \quad (4.3)$$

α and β indicate the three cartesian coordinates. Because the system was in its equilibrium position, the second term vanishes. The interatomic force constants (IFC) matrix $C_{s\alpha, s'\beta}(l, m)$, where the indices I and J are written as $I = (l, s)$, $J = (m, s')$, is defined by

$$C_{s\alpha, s'\beta}(l, m) = C_{s\alpha, s'\beta}(\mathbf{R}_l - \mathbf{R}_m) = \frac{\partial^2 E}{\partial u_{l s \alpha} \partial u_{m s' \beta}}. \quad (4.4)$$

Using this definition, together with the Newton equation 4.2 gives

$$M_{ls} \ddot{u}_{ls\alpha} = - \sum_{ms'\beta} C_{s\alpha, s'\beta}(\mathbf{R}_l - \mathbf{R}_m) u_{ms'\beta}. \quad (4.5)$$

For this equation we search for a solution in the form of a phonon. For each \mathbf{q} -vector in the first Brillouin zone we can write the displacements in the

4.3. PRACTICAL PHONON CALCULATIONS

harmonic approximation as

$$\mathbf{u}_{ls\alpha}(t) = \frac{1}{\sqrt{M_s}} \mathbf{u}_{s\alpha}(\mathbf{q}) e^{i\omega t - \mathbf{q} \cdot \mathbf{R}_l}, \quad (4.6)$$

so that equation 4.5 becomes

$$\omega_{\mathbf{q}}^2 \mathbf{u}_{s\alpha}(\mathbf{q}) = \sum_{s'\beta} \tilde{C}_{s\alpha, s'\beta}(\mathbf{q}) \mathbf{u}_{s'\beta}(\mathbf{q}), \quad (4.7)$$

where the indices I and J are written as $I = (l, s), J = (m, s')$ and the dynamical matrix \tilde{C} is given by

$$\tilde{C}_{s\alpha, s'\beta}(\mathbf{q}) = \frac{1}{\sqrt{M_s M_{s'}}} \sum_m \frac{\partial^2 E}{\partial \mathbf{u}_{ls\alpha} \partial \mathbf{u}_{ms'\beta}} e^{i\mathbf{q}(\mathbf{R}_m - \mathbf{R}_l)}. \quad (4.8)$$

This forms an eigenvalue problem for the $3 \times N_{atoms}$ variables $\mathbf{u}_{s\alpha}(\mathbf{q})$.

4.3 Practical phonon calculations

4.3.1 Frozen phonons

In order to obtain the phonon frequencies one needs to obtain the second order derivatives of the energy with respect to the ionic positions. Equivalently, one can calculate the first derivative of the forces (equation 2.61), which can be easily obtained using the Hellmann-Feynman theorem. The first derivatives can then be obtained by calculating the forces for two different perturbed ionic positions, and utilizing the finite difference method to obtain the derivative. This is the so-called ‘‘frozen phonon’’ method. A drawback of this method is that in order to have the phonon response to an arbitrary phonon vector, this vector should be commensurate with the unit cell, as for a general $q \neq 0$ the translation symmetry is lost. Therefore one will need a super cell approach, thereby increasing the computational cost. This drawback can be avoided by performing a perturbation calculation using the Density Functional Perturbation Theory (DFPT) [77, 78, 79].

4.3.2 Density functional perturbation theory

In section 4.2, it was shown that to obtain the phonon frequencies, the dynamical matrix should be calculated, and in particular the second derivative

CHAPTER 4. DENSITY FUNCTIONAL PERTURBATION THEORY

of the energy with respect to the perturbations (equation 4.8). This can be obtained by the application of the Hellmann-Feynman theorem, followed by another derivative. This is similarly to what was done in section 2.5, but here this has to be done with respect to \mathbf{u} . This gives

$$\frac{\partial^2 E}{\partial \mathbf{u}_{l\alpha} \partial \mathbf{u}_{ms'\beta}} = \int d^3r \frac{\partial^2 V_{eff}(\mathbf{r})}{\partial \mathbf{u}_{l\alpha} \partial \mathbf{u}_{ms'\beta}} n(\mathbf{r}) + \int d^3r \frac{\partial V_{eff}}{\partial \mathbf{u}_{l\alpha}} \frac{\partial n(\mathbf{r})}{\partial \mathbf{u}_{ms'\beta}} + \frac{\partial^2 E_{II}}{\partial \mathbf{u}_{l\alpha} \partial \mathbf{u}_{ms'\beta}}. \quad (4.9)$$

The quantity that still needs to be calculated here is the charge density induced by the perturbation

$$\frac{\partial n(\mathbf{r})}{\partial \mathbf{u}_{ms'\beta}} = \sum_i \left(\frac{\partial \Psi_i^*(\mathbf{r})}{\partial \mathbf{u}_{ms'\beta}} \Psi_i(\mathbf{r}) + \Psi_i^*(\mathbf{r}) \frac{\partial \Psi_i(\mathbf{r})}{\partial \mathbf{u}_{ms'\beta}} \right). \quad (4.10)$$

These quantities can be found by inserting the Taylor expansions in \mathbf{u} of the Kohn-Sham potential V_{KS} , the wavefunction Ψ_i and the eigenvalues ϵ_i

$$\begin{aligned} V_{KS}(\mathbf{r}, \mathbf{u}) &= V_{KS}(\mathbf{r}, \mathbf{u} = 0) + \mathbf{u} \frac{\partial V_{KS}(\mathbf{r})}{\partial \mathbf{u}} + \dots \\ \Psi_i(\mathbf{r}, \mathbf{u}) &= \Psi_i(\mathbf{r}, \mathbf{u} = 0) + \mathbf{u} \frac{\partial \Psi_i(\mathbf{r})}{\partial \mathbf{u}} + \dots \\ \epsilon_i(\mathbf{r}, \mathbf{u}) &= \epsilon_i(\mathbf{r}, \mathbf{u} = 0) + \mathbf{u} \frac{\partial \epsilon_i(\mathbf{r})}{\partial \mathbf{u}} + \dots, \end{aligned} \quad (4.11)$$

$$(4.12)$$

in the Kohn-Sham equation

$$\left(-\frac{1}{2} \nabla^2 + V_{KS}(\mathbf{r}) \right) \Psi_i(\mathbf{r}) = \epsilon_i \Psi_i(\mathbf{r}). \quad (4.13)$$

Keeping only terms in first order in \mathbf{u} gives the Sternheimer equation [80]

$$\left(-\frac{1}{2} \nabla^2 + V_{KS}(\mathbf{r}) - \epsilon_i \right) \frac{\partial \Psi_i(\mathbf{r})}{\partial \mathbf{u}_{ms'\beta}} = \left(-\frac{\partial V_{KS}}{\partial \mathbf{u}_{ms'\beta}} + \frac{\partial \epsilon_i}{\partial \mathbf{u}_{ms'\beta}} \right) \Psi_i(\mathbf{r}), \quad (4.14)$$

where

$$\begin{aligned} \frac{\partial V_{KS}}{\partial \mathbf{u}_{ms'\beta}} &= \frac{\partial V_{ext}}{\partial \mathbf{u}_{ms'\beta}} + \frac{\partial V_H}{\partial \mathbf{u}_{ms'\beta}} + \frac{\partial V_{xc}}{\partial \mathbf{u}_{ms'\beta}} \\ &= \frac{\partial V_{ext}}{\partial \mathbf{u}_{ms'\beta}} + \int d^3r' \frac{1}{|\mathbf{r} - \mathbf{r}'|} \frac{\partial n(\mathbf{r}')}{\partial \mathbf{u}_{ms'\beta}} + \frac{dV_{xc}}{dn} \frac{\partial n(\mathbf{r})}{\partial \mathbf{u}_{ms'\beta}}. \end{aligned} \quad (4.15)$$

4.3. PRACTICAL PHONON CALCULATIONS

This system of equations can then be solved selfconsistently for the $3 \times N_{atoms}$ perturbations at fixed \mathbf{q} -vector.

To obtain $\omega(\mathbf{q})$ this should be repeated for different \mathbf{q} -vectors. But this can be performed more efficiently. When looking at the dynamical matrix $\tilde{C}_{s\alpha,s'\beta}(\mathbf{q})$ 4.8, one observes that this is a periodic function of \mathbf{q} , as $\tilde{C}_{s\alpha,s'\beta}(\mathbf{q}) = \tilde{C}_{s\alpha,s'\beta}(\mathbf{q} + \mathbf{G})$, for all reciprocal lattice vectors \mathbf{G} . As all Bravais lattice points are equivalent in a solid, it does not depend on l . It can therefore be seen as a Fourier expansion of a three dimensional function, where we only have Fourier components at discrete values \mathbf{R}_m . So we can also write

$$\frac{1}{\sqrt{M_s M_{s'}}} \sum_m \frac{\partial^2 E}{\partial \mathbf{u}_{l s \alpha} \partial \mathbf{u}_{m s' \beta}} = \frac{\Omega}{(2\pi)^3} \int d^3 q \tilde{C}_{s\alpha,s'\beta}(\mathbf{q}) e^{-i\mathbf{q} \cdot (\mathbf{R}_m - \mathbf{R}_l)}. \quad (4.16)$$

This allows to use a Fourier interpolation by sampling this integral in a discrete set of \mathbf{q} points, which will give the IFCs for a set of vectors \mathbf{R}_m . If the dynamical matrix is a smooth function of \mathbf{q} and the IFCs decay rapidly in real space, equation 4.8 can be used to calculate the dynamical matrix at arbitrary \mathbf{q} -points, limiting the sum to only a few values of \mathbf{R}_m .

Special care should be taken in metals due to Kohn anomalies, as $\tilde{C}_{s\alpha,s'\beta}(\mathbf{q})$ is no longer a smooth function of \mathbf{q} in that case and the IFC are long range. Also in polar materials the atomic displacements generate long range electrostatic interactions and the dynamical matrix is no longer analytic for the limit $\mathbf{q} \rightarrow 0$. Methods exist for dealing with these problems (e.g. Born effective charges), which can be found for example in the review of Baroni *et al.* in Ref. [77].

5

Phonons in semiconductor nanowires

5.1 Silicon nanowires

Knowledge of the phonon spectrum is important for several applications, e.g., the thermal conductance of semiconducting systems is governed by phonons and their dispersion and interactions. Another example is the low-field mobility, which depends strongly on scattering with acoustic phonons, surface roughness, and impurities. The phonon spectrum can also tell us something about the stability of the system. For example the occurrence of imaginary eigenfrequencies indicates a structural instability, which, e.g., was recently found in small one-dimensional (1D) metallic nanowires [81]. We use this property to study the stability of Si wires grown in the [110] direction.

Early theoretical studies of phonons in nanowires used an elastic continuum model to obtain the phonon dispersion of cylindrical and rectangular nanowires. A qualitative description was given, mainly focusing on the dilatational modes (see [82] for an extended discussion of this model). More complex geometries like coated cylindrical nanowires were also studied using similar models [83]. More recent studies used an atomistic Green function approach to investigate thermal properties [84] or used an adiabatic bond charge method [85] to obtain the full phonon spectrum of pristine wires. The main focus was on the lowest and highest frequency optical modes. In a recent paper [86] a combination of tight-binding and a continuum elastic model was used to calculate the acoustic phonon limited hole mobility, and it was shown that at room temperature this mobility is almost three times larger than the bulk mobility. This is an extra motivation to use SiNWs in future electronic applications.

In this chapter we perform full *ab initio* calculations of the phonon dispersion of hydrogen-passivated SiNWs. We investigate the stability of SiNWs

and determine under which conditions the phonon frequencies of SiNWs are real. We found that structural instabilities can be caused by computational limitations, e.g., the small size of the unit cell used in *ab initio* calculations, or by not sufficiently relaxing the system. Ground-state calculations alone are not able to provide this information. Experimentally it was observed that for small wires the majority of SiNWs are oriented along the [110] direction [24]. Therefore we restrict our study to such wires. We present results for different sizes of SiNWs, for both undoped and B- or P- doped wires. We also study the effect of the size of the wire on the phonon confinement and the group velocity of acoustic modes.

Our calculations are performed within the density-functional theory (DFT) using the local density approximation (LDA) for the exchange-correlation functional. The present results have been obtained using the ABINIT code [63] and Troullier-Martins [53] pseudopotentials. A kinetic energy cutoff of 20 hartree was used for the undoped and P-doped SiNWs and of 30 hartree for the B-doped SiNWs. A $1 \times 1 \times 16$ Monkhorst-Pack [52] grid was used, and all structures were relaxed without imposing any symmetry constraint. The nanowire was considered to be relaxed if the largest force component is smaller than $5 \cdot 10^{-5}$ hartree/bohr and all stress components are smaller

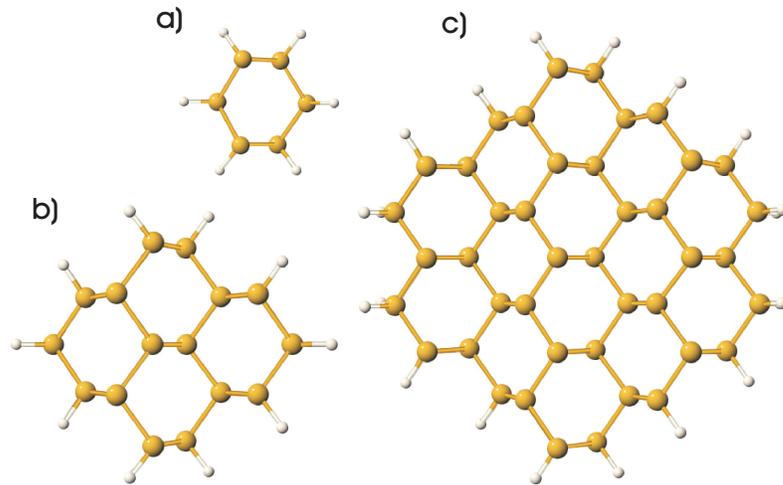


Figure 5.1: Relaxed unit cell of a NW: a) the 0.5 nm [110] wire consists of 6 Si and 8 H atoms, b) the 1.2 nm wire consists of 16 Si and 12 H atoms, and c) the 1.6 nm wire has 42 Si and 20 H atoms.

than 5.10^{-7} hartree/bohr³. Due to the periodic boundary conditions, a layer of vacuum (of 6 Å) was used in the x and y directions to create an infinite one-dimensional system. The phonon band structure was calculated using density functional perturbation theory (DFPT) explained in the previous chapter. All performed DFPT calculations were started from the fully relaxed structures.

First, we calculate the phonon spectrum of a very thin SiNW (diameter of 0.5 nm) in the [110] direction. Fig. 5.1(a) shows the relaxed structure of this wire. The calculated phonon spectrum is shown in Fig. 5.2(a). Following convention, imaginary eigenfrequencies of unstable phonon modes will be shown as a negative frequency. It must be stressed that we only show the spectrum up to a frequency of 1000 cm^{-1} . The higher modes will be discussed later. As there are no imaginary frequencies in this spectrum, the wire can be considered structurally stable. Fig. 5.3 zooms in on the acoustic branches of the undoped 0.5 nm wire near the Γ point. Four zero-frequency modes can be observed, three acoustic modes that are related to the translational symmetry and one extra acoustic mode (not present in bulk) that is related to the rotational invariance of the wire around the wire axis. Two of these modes are linear in q_z , and correspond to displacements along the wire axis, i.e., a translational mode along the wire axis and the other is a rotational mode, i.e., torsion of the wire. The two other modes are proportional to q_z^2 . These softer modes are eigenmodes with displacements orthogonal to the wire axis and correspond to a bending of the wire. The eigenvectors of the rotational mode can be seen in inset (a) of Fig. 5.3, which shows the calculated eigenvectors of this mode for a larger (1.6 nm) wire. Inset (b) shows the longitudinal translational mode (also linear), and insets (c) and (d) show the quadratic orthogonal modes. The shown eigenvectors are normalized with respect to the atomic mass and multiplied by a factor for displaying purposes.

When we dope this wire with B or P (by replacing one of the Si atoms), the wire becomes unstable, as can be seen from the occurrence of imaginary frequencies in the phonon spectrum, as shown in Fig. 5.2(b) and Fig. 5.2(c). But we found that this structure can be stabilized by increasing the unit cell in the growth direction. This increases the degrees of freedom and allows a larger relaxation. This is illustrated in Fig. 5.2(d), which shows the phonon spectrum of the same wire, where the unit cell in the [110] direction is doubled, so that the effective dopant concentration is halved. It should be noted that the lowest acoustic modes still have a very small imaginary frequency.

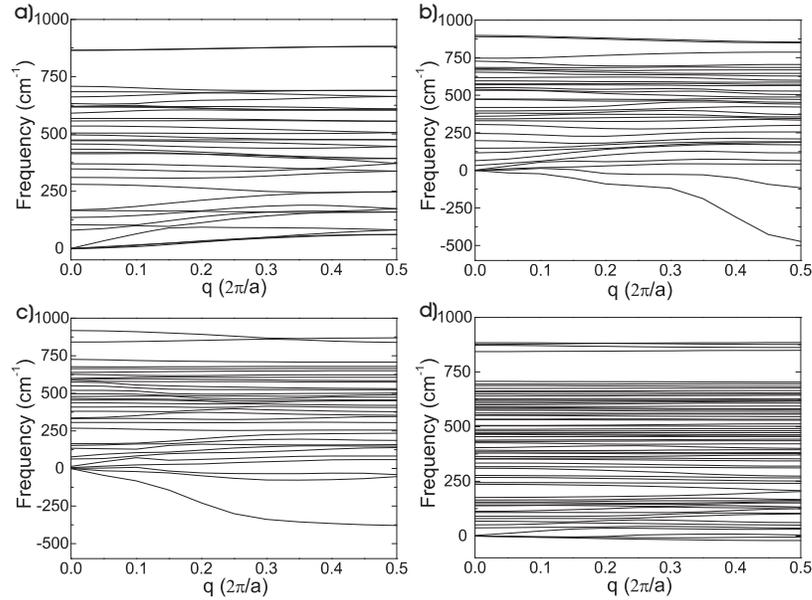


Figure 5.2: Phonon dispersion relation for the 0.5 nm undoped (a) nanowire (see Fig. 5.1(a)), doped with B (b) and P (c). When the unit cell is doubled in the z direction ([110] direction), the P doped 0.5 nm nanowire becomes more stable (d).

Further increasing the unit cell in the growth direction stabilizes the wire, so that the resulting phonon spectra will be completely real. When the passivating hydrogen layer is removed (and after the wire is fully relaxed), we again found imaginary frequencies and thus instabilities in the structure, as could be expected considering the presence of unpassivated bonds.

When a thicker wire with diameter of 1.2 nm (relaxed structure shown in Fig. 5.1(b)) is doped with either B or P, the wire remains stable. The dopants are placed on the preferential dopant locations, which are the near edge locations, as reported in chapter 3. The calculated spectra are shown in Fig. 5.4. In Fig. 5.4(a) the spectrum of an undoped wire is shown, while Fig. 5.4(b) and Fig. 5.4(c) show the spectrum of the B- and P-doped wire, respectively. If the wire diameter is further increased (see Fig. 5.4(d) for the phonon spectrum of an undoped 1.6 nm wire), the same stability is observed.

To have a better understanding of the evolution of the phonon spectrum as function of the wire diameter, we also calculated the corresponding phonon

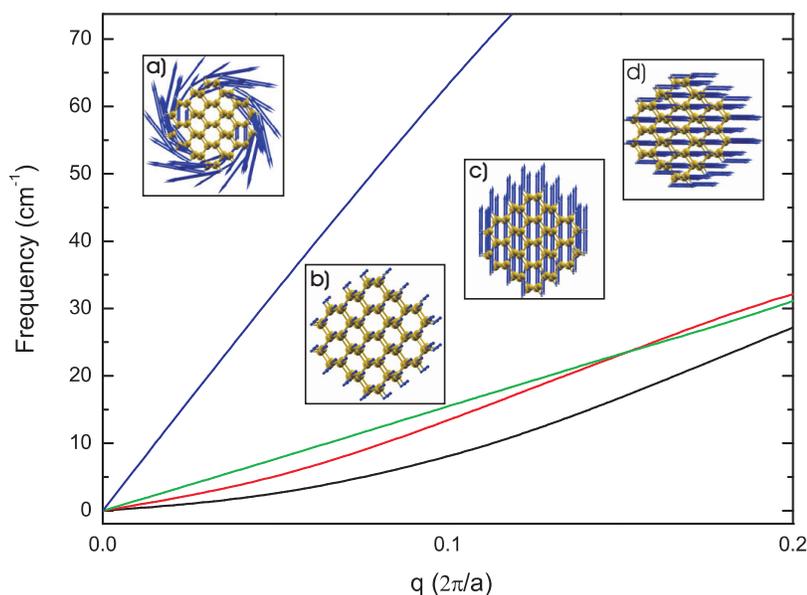


Figure 5.3: A close-up of the region of small frequencies around Γ for the 0.5 nm undoped nanowire. The two linear and the two quadratic modes can be clearly distinguished. The linear modes correspond to a torsion mode (eigenmodes shown in inset (a)) and to the longitudinal acoustic mode (inset (b)). The two softer quadratic modes (insets (c) and (d)) correspond to a bending of the wire.

density of states (DOS), by using a smeared histogram counting method (Gaussian smearing with a full width at half-maximum of 10 cm^{-1}). The resulting DOS is normalized with respect to the number of bands, so that a direct comparison is possible. This DOS is shown in Fig. 5.5 and several features can be seen, especially when also looking at the corresponding phonon band spectra (Fig. 5.2(a), Fig. 5.4(a) and Fig. 5.4(d)). Notice that the curvature of the lowest lying bands decreases when the wire diameter decreases, which can also be observed in the low-frequency DOS (the DOS increases with decreasing wire diameter). This softening of the phonon modes is as expected in 1D-confined structures. Surprisingly, on the other hand there is a hardening of the optical modes: the lowest lying optical mode (denoted by ω_σ and shown for the 1.6 nm wire in inset (a) of Fig. 5.5) is shifted toward higher frequencies when the diameter decreases. This is in agreement with previous calculations within the adiabatic bond charge method [85], but in

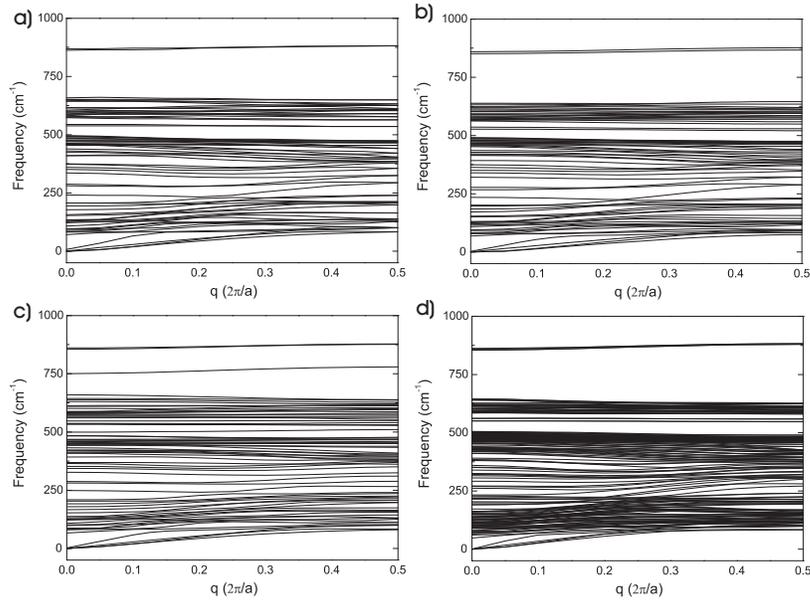


Figure 5.4: Phonon dispersion relation for the 1.2 nm nanowire: a) undoped wire, b) P-doped wire, and c) B-doped wire. In d) the phonon dispersion relation for the 1.6 nm wire is shown.

our case the behavior of this mode is slightly different due to the passivating hydrogen layer. It can also be observed that the optical modes around 700 cm^{-1} (ω_1) are shifted toward higher frequencies and that this shift is more substantial than the shift of ω_σ . A similar upward shift can be observed for the breathing mode (ω_b , shown in inset (b) of Fig. 5.5), which shifts from a value of 111 cm^{-1} for the 1.6 nm wire to a value of 310 cm^{-1} for the smallest wire. Similar upward shifts for the breathing mode have also been observed for carbon nanotubes [87].

Table 5.1 shows the numerical values of these optical phonon modes for the different wires. The phonon velocities (for the longitudinal, i.e., the acoustic mode along the wire direction, and the transversal mode, i.e., the rotational mode) are also shown in this table. Both sound velocities increase with increasing wire diameter and converge toward the velocities in bulk silicon.

As can be seen in the DOS plot, there are also modes with frequencies around 2100 cm^{-1} . These can be attributed to the stretching of the Si-H bonds [88]. For all considered wires the number of such hydrogen modes

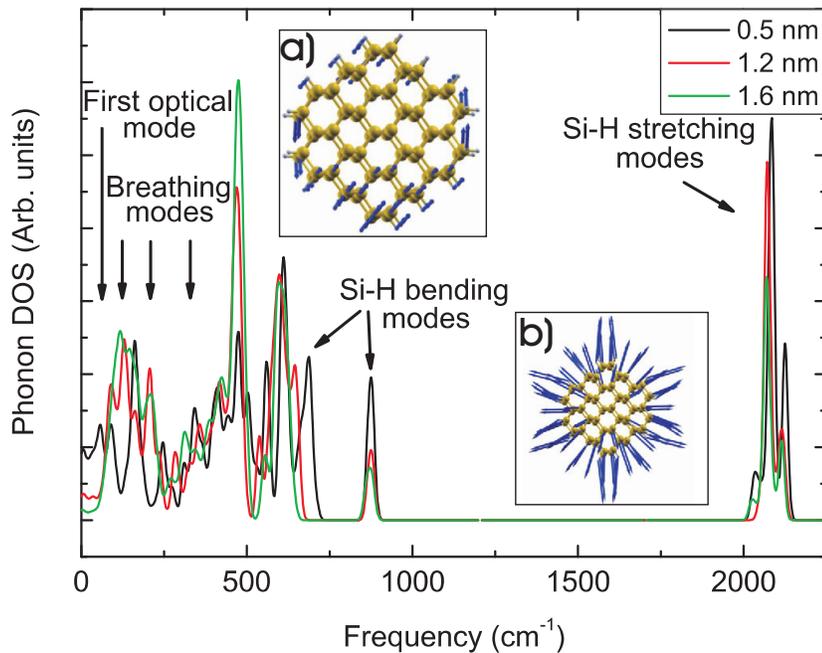


Figure 5.5: Phonon density of states (DOS) for wires with different diameter. The arrows indicate special eigenmodes. For exact frequencies of the breathing mode (shown with three arrows for the different wire diameters), see Table I. The eigenmodes of the first optical mode (a) and the breathing mode (b) for the 1.6 nm nanowire are shown in the insets.

d (nm)	ω_σ (cm ⁻¹)	ω_1 (cm ⁻¹)	ω_b (cm ⁻¹)	v_T (km/s)	v_L (km/s)
0.5	79.79	708.09	309.98	7.66	1.78
1.2	70.76	658.42	177.98	7.95	3.65
1.6	47.64	645.02	111.26	7.98	3.88
bulk				8.23	4.83

Table 5.1: The frequency of the lowest optical phonon mode (ω_σ), the frequency of the highest optical mode around 700 cm⁻¹ (ω_1), the frequency of the breathing mode (ω_b), the transversal acoustic sound velocity (v_T) and the longitudinal, i.e. torsion mode, sound velocity (v_L) for the different SiNW diameter. The calculated bulk acoustic velocities are given as reference.

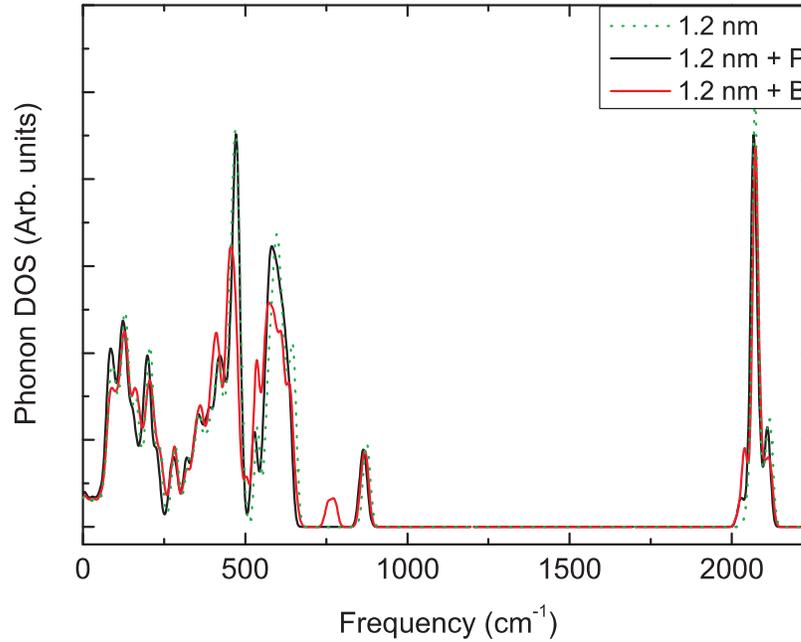


Figure 5.6: Phonon density of states (DOS) for the 1.2 nm nanowire, doped with one B or P atom. The dotted line shows the DOS of an undoped nanowire.

corresponds to the number of hydrogen atoms. Similarly the modes around 630 and 870 cm^{-1} are linked to the bending of the Si-H bonds [88].

When the phonon DOS of a 1.2 nm wire doped with a B or P atom are compared (see Fig. 5.6), an additional peak in the DOS around 760 cm^{-1} for the B-doped wire can be observed, which is a Si-H bending mode that is modified by the nearby Si-B bond because of the reconstruction. The corresponding mode in the P-doped spectrum has a lower frequency. This shift is caused by the difference in mass between the B and the P dopant (a B atom has a smaller mass, so the corresponding frequency will be higher).

5.2 Germanium nanowires

Also in germanium nanowires it is important to know if small highly doped unit cells, commonly used in *ab initio* calculations are stable. As in the previous subsection this can be shown using the phonon frequencies. These

5.2. GERMANIUM NANOWIRES

can then be used to study the dependency of the phonon frequencies and velocities on the wire diameter. Using the results obtained in the previous subsection, a comparison is made with silicon nanowires.

Previous theoretical research focused mainly on the electronic [67, 89] and optical properties [68, 69] of free standing GeNWs and the effect of quantum confinement on these properties and on the effects of B and/or P doping [90]. Long-wavelength continuum approaches [91, 82], applicable for micron-size wires, have been used to study polar and non-polar optical phonon modes. Here we will follow an atomistic approach based on full *ab initio* calculations.

To study the effect of the wire diameter on the phonon spectra we consider three different sized Ge nanowires, from the smallest possible hydrogen terminated wire in the [110] direction to more realistic sizes (from 0.5 nm to 1.6 nm). These wires are very comparable to the Si nanowires from the previous section. Cross sections of these three wires are shown in Fig. 5.7.

The phonon spectra, shown in Fig. 5.8, only contain positive real frequencies, thus all these structures are stable. The spectra are shown for frequencies smaller than 900 cm^{-1} . Higher frequencies are also present (around 2000 cm^{-1}), but these will be discussed later. The spectra near the Γ point

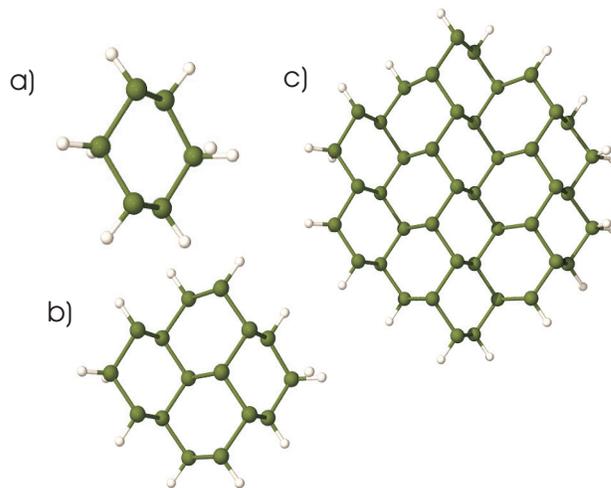


Figure 5.7: Cross section of the studied [110] GeNWs, having three different wire diameters: (a) 0.5 nm (containing 6 Ge and 8 H atoms), (b) 1.2 nm (containing 16 Ge and 12 H atoms), and (c) 1.6 nm (containing 42 Ge and 20 H atoms).

CHAPTER 5. PHONONS IN SEMICONDUCTOR NANOWIRES

contain four acoustic modes, containing the three standard acoustic modes, caused by translational symmetry and one mode, typical for one-dimensional systems and caused by the rotational symmetry. Of these modes, two are linear in q_z and two are quadratic in q_z . The two linear modes correspond to a translation in the wire direction and to the rotational mode. The two quadratic modes are softer modes and correspond to the two translational modes perpendicular to the wire, i.e. these are bending modes.

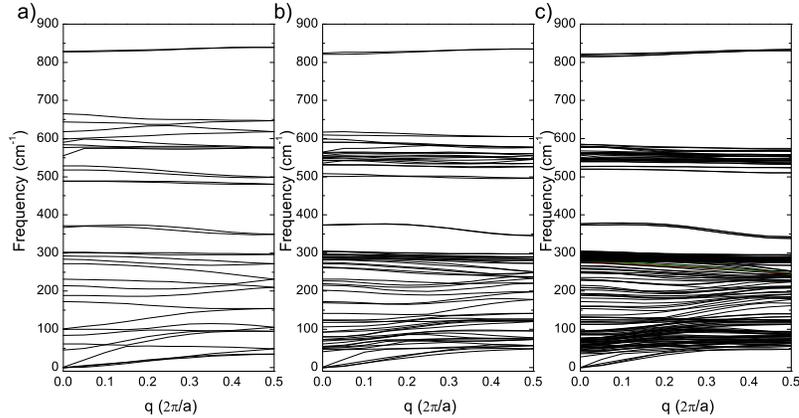


Figure 5.8: The calculated phonon spectra of a Ge nanowire with diameter: (a) 0.5 nm, (b) 1.2 nm, and (c) 1.6 nm. The spectra are only shown for frequencies smaller than 1000 cm^{-1} .

To be able to compare these spectra for the different wires (see Fig. 5.9), we calculated the phonon density of states (DOS), by using a smeared (using a Gaussian smearing function with a full-width at half-maximum of 10 cm^{-1}) histogram counting method. The DOS is normalized with respect to the number of phonon bands, so that a direct comparison is possible between the different sizes.

The peaks around 2000 cm^{-1} can be attributed to Ge-H stretching modes, while those around 820 cm^{-1} and 600 cm^{-1} are related to the Ge-H bending modes. These frequencies are within the expected experimental range, for example as measured in hydrogenated amorphous Ge [92, 88]. The frequency of the optical modes, e.g. those around 600 cm^{-1} , increases when the wire diameter decreases. This corresponds to a hardening of the optical phonon modes, which is similar to the result found for silicon nanowires [93]. The same behavior can be seen from the phonon frequencies listed in Table 5.2. In

5.2. GERMANIUM NANOWIRES

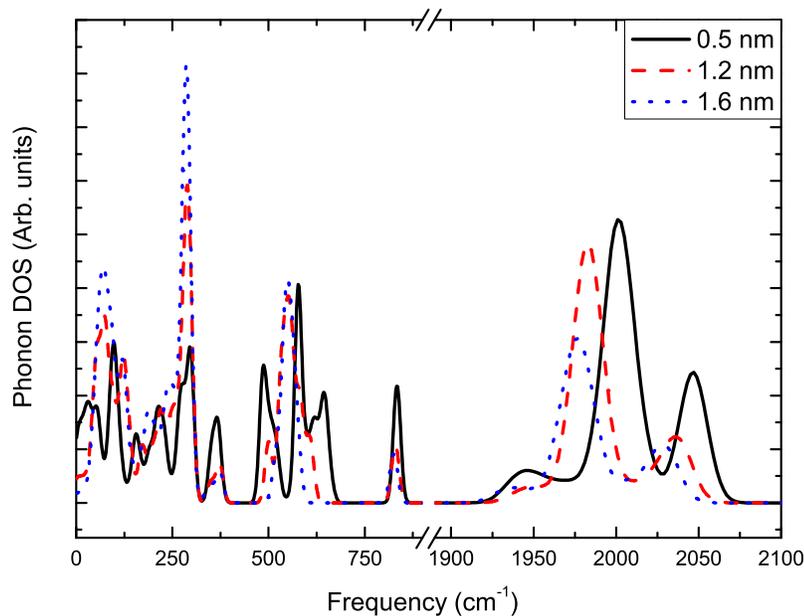


Figure 5.9: The phonon density of states of Ge nanowires with different diameters.

d (nm)	ω_σ (cm $^{-1}$)	ω_1 (cm $^{-1}$)	ω_b (cm $^{-1}$)	v_T (km/s)	v_L (km/s)
0.5	45.40	664.98	189.02	4.84	1.16
1.2	41.79	616.51	106.77	4.97	1.99
1.6	27.78	584.95	65.47	4.90	2.33

Table 5.2: The frequency of the lowest optical phonon mode (ω_σ), the frequency of the highest optical mode around 700 cm $^{-1}$ (ω_1), the frequency of the breathing mode (ω_b), the transversal acoustic sound velocity (v_T) and the longitudinal (i.e. torsion mode) sound velocity (v_L) for the different GeNW diameters.

that table, the frequencies of the first optical mode (ω_σ), the breathing mode (ω_b) and the highest optical mode near 600 cm $^{-1}$ (ω_1) are listed. Notice that a similar shift for the breathing frequency was found in carbon nanotubes [87]. The same table also lists the acoustic phonon velocities, computed using the slope of the $\omega - q$ curves, and which have a weak dependence on the wire

CHAPTER 5. PHONONS IN SEMICONDUCTOR NANOWIRES

diameter. For the longitudinal velocity we consider the acoustic mode in the wire direction and for the transversal mode the rotational mode. The longitudinal acoustic phonon velocities decrease when the wire diameter decreases, while this is not the case for the transversal velocities. This is also visible in the DOS in Fig. 5.9 for low frequencies, as the DOS is larger for smaller diameter wires. This softening of the acoustic modes is what one expects for Q1D structures. Compared to Si nanowires, the velocities in Ge are smaller, e.g. for the 0.5 nm nanowire, the transversal velocity in Si is 7.66 km/s compared to 4.84 km/s for Ge.

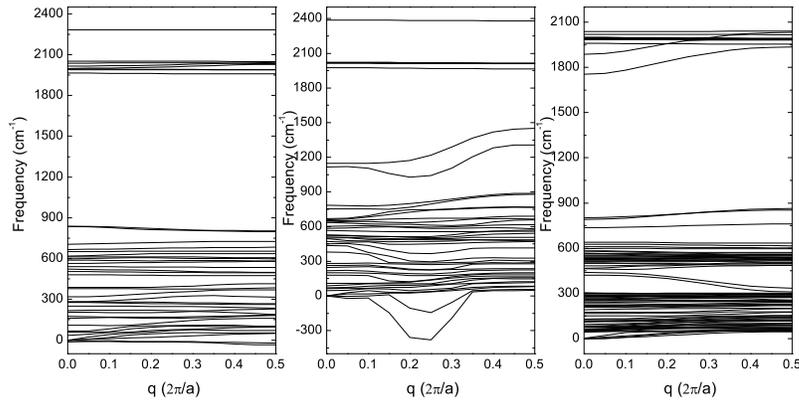


Figure 5.10: The obtained phonon spectra for doped wires. The dopant is placed on it's preferential position. From left to right: 0.5 nm wire doped with P, doped with B, and the 1.2 nm wire doped with P (the spectrum of the wire doped with B is not shown, as it looks similar). Imaginary frequencies (shown as negative frequencies) are present in case of the 0.5 nm nanowire, indicating a structural instability.

The phonon spectra can also be used to obtain an idea of the stability of structures under perturbations as imaginary frequencies indicate structural instabilities. All considered wires were stable when undoped. If we introduce a single substitutional B or P dopant in a small unit cell, thus corresponding to a high concentration of dopants, this changes. When the smallest possible wire is doped (distance between two dopants is 3.96\AA), imaginary frequencies occur. This instability is caused by the strong interactions between the dopants and the local distortions around it. The larger 1.2 nm wire remains stable when dopants are placed on the preferred positions (see 3), and the same is true for larger wires. From this we can conclude that the Ge-Ge zinc

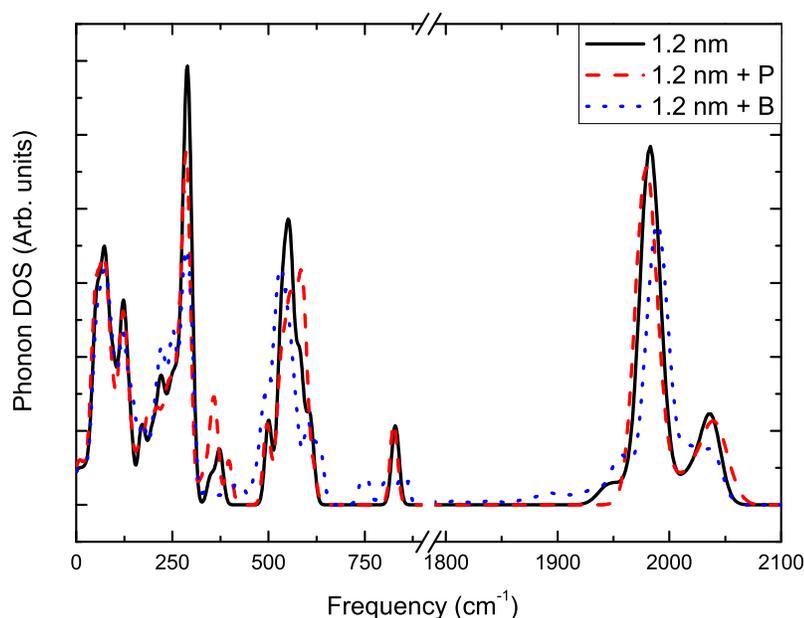


Figure 5.11: The phonon density of states of the 1.2 nm nanowire, in the undoped, B-doped and P doped case.

blende structure, which can be seen starting from the 1.2 nm nanowire, is rigid enough to allow for a local distortion. Enlarging the unit cell of the smallest wire in the wire direction allows long range structural relaxations, so that the phonon spectra becomes real. The corresponding spectra are shown in Fig. 5.10, where the imaginary frequencies are shown as negative frequencies.

Fig. 5.11 shows the DOS of the 1.2 nm wire, either doped with B, P or undoped. The differences between the undoped and the P doped spectra are much smaller than between the undoped and the B doped spectra, which could be expected by the size difference between B and P atoms.

5.3 Core-shell nanowires

Other interesting one-dimensional systems are the so-called coaxial heterostructures or core-shell nanowires. Coaxial Si/Ge heterostructures of such wires can readily be made [30, 94, 95, 29, 96]. Due to the sharp interface between a Ge core and a Si shell and the valence band offset, the interface

serves as a confinement potential, so that a one-dimensional hole gas can be created in undoped structures [95]. Additionally the Si shell can be used to create transparent contacts. Due to these properties, these heterostructures can be used as high switching speed transistors [94] and FETs [29]. P-i-n type heterostructures have already shown applications as solar cells [11].

Changing the sizes of the core and shell regions allows one to engineer the bandgaps, as was shown experimentally [96] and theoretically [97, 98]. The atomic structure of these wires can experimentally be studied by using transmission electron microscopy (TEM), but knowledge of the phonon spectra can lead to a characterization based on Raman spectroscopy. A first theoretical step in the understanding of the lattice dynamics of Si/Ge core-shell structures has been made using the empirical valence force field model containing two- and three-body interactions [99] to study the phonon spectra and the specific heat in these wires. Here, we go a step further and use a full *ab initio* approach to study *both* the electronic and the dynamical properties of Si/Ge core-shell nanowires and compare these results with those of full Si or Ge nanowires. We restrict ourselves to structures that are oriented in the [110] direction and are hydrogen passivated.

In this section we study two different sized nanowires with a diameter of 1.2 and 1.6 nm, both of them hydrogen passivated. For each of these diameters either a wire with a core composed of Si or Ge is constructed. For the 1.6 nm nanowire two different shell sizes are considered. A cross section of the resulting relaxed structures in case of a Ge core is shown in Fig. 5.12. In the following we label these wires as core-shell a, b and c. The reverse structures (i.e. where Ge is replaced with Si and vice versa) are also studied.

In order to understand the modification of the electronic band structure due to the presence of a core, we considered as an example the case of a Ge core and show in Fig. 5.13 the energy band structures. The wires with Si in the core have similar bandstructures. Notice that all core-shell structures have a direct bandgap (values are plotted in Fig. 5.16(c)), which is similar for Si and Ge nanowires and confirms earlier results [98]. Starting from a pure nanowire (either a Si or Ge one) and adding a core of the second material leads to a smaller bandgap and this downward trend is reversed when the quantity of the other material becomes larger, i.e. when x (concentration Ge) passes the 0.5 point. As is well-known the experimental bandgaps will be larger, as the LDA underestimates its size, but all relative trends are predicted correctly. In Fig. 5.16(a) and (b) respectively the effective electron

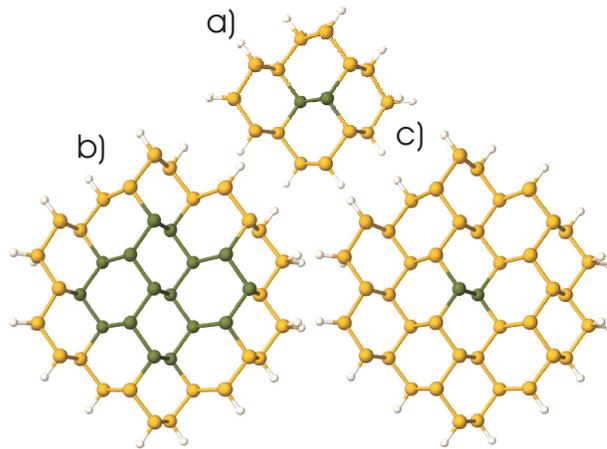


Figure 5.12: Relaxed unit cell of three different nanowires oriented along the $[110]$ direction, all with a Ge core: (a) a 1.2 nm $[110]$ wire with a core of 2 Ge atoms (green color), a shell of 14 Si atoms (yellow color) and a passivating layer of 12 H atoms (white color), (b) a 1.6 nm nanowire with 16 Ge core atoms, 26 Si shell atoms and 20 passivating H atoms, and (c) 1.6 nm wire with 2 Ge core atoms, 40 Si shell atoms and 20 H atoms.

and hole band mass are plotted as function of the concentration and structure type. The electron effective mass is quasi independent of the wire type and wire diameter. The hole effective mass strongly depends on the type of structure and of the wire size.

Next we calculate the dynamical properties and present the phonon energies as function of the wavevector in Fig. 5.14 for the three nanowires shown in Fig. 5.12. Notice that the low frequency (acoustic) modes around Γ exhibit the typical 1D features: 4 acoustic modes of which 2 are linear in q_z and 2 are quadratic in q_z [93]. All calculated phonon spectra have only real frequencies, and thus all structures can be considered stable (as the occurrence of imaginary frequencies would indicate a structural instability). To directly compare these spectra, it is useful to calculate the phonon density of states (DOS). Fig. 5.15 shows the DOS plot of core-shell b (see Fig. 5.12(b)). It is instructive to compare this DOS with the DOS of pure nanowires. In Fig. 5.15(a) the DOS of a pure 1.6 nm SiNW [93] is also added (red dashed line), while in Fig. 5.15(b) the DOS of a pure 1.6 nm GeNW [100] (black dashed line) and also of the core-shell structure c (red dotted line) is added.

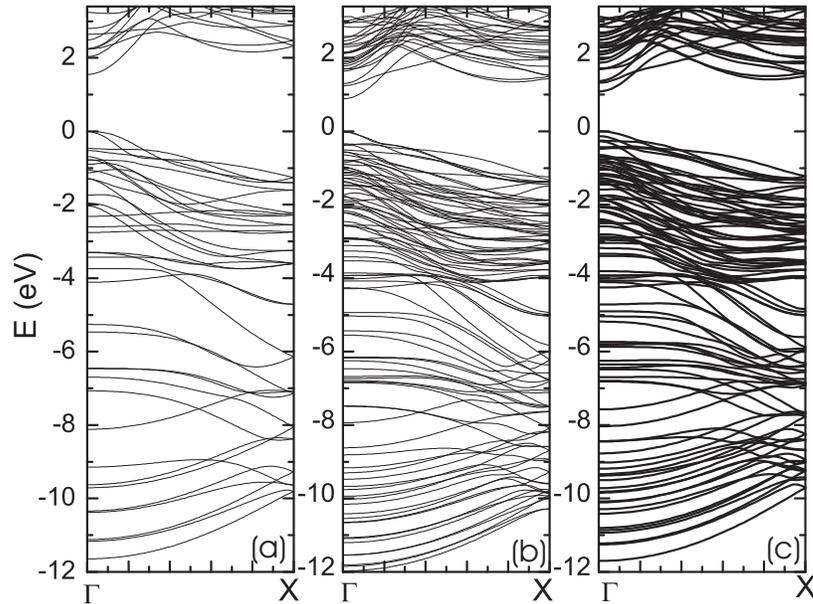


Figure 5.13: Bandstructure of Ge core nanowires corresponding to the structures depicted in Fig. 5.12

When comparing the DOS of the core-shell structure b with the one from either the Si or Ge nanowire, one can immediately identify what types of atoms (Si or Ge) are present in the shell, as the Si-H and Ge-H stretching modes (with frequencies around 2000 cm^{-1}) have different frequencies. The same observation can be made for the bending modes (frequencies around 870 cm^{-1} and 600 cm^{-1}), which in this case is determined by the Si shell. Both the DOS for the pure Si and Ge nanowire have a characteristic peak, 290 cm^{-1} for pure Ge (labeled I in Fig. 5.15) and one around 480 cm^{-1} for pure Si (labeled II in Fig. 5.15). The effect of the core atoms is visible in a lowering of this Ge peak (peak I) and the increase of the Si contribution around 480 cm^{-1} (peak II). In core-shell structure c the number of Ge atoms in the core is small as compared to those in the core-shell structure b (just 2 Ge atoms), which shows up as a decrease of the Ge peak (peak I) in the DOS (Fig. 5.15(b)) and an increase in the Si peak (peak II).

Next we look in more detail to some specific modes with their associated energies. In order to do this the acoustic phonon velocities (calculated for the transversal (rotational) (v_T) and longitudinal mode (v_L)), the first optical frequency (ω_σ), the bending mode around 600 cm^{-1} (ω_1) and the breathing

5.3. CORE-SHELL NANOWIRES

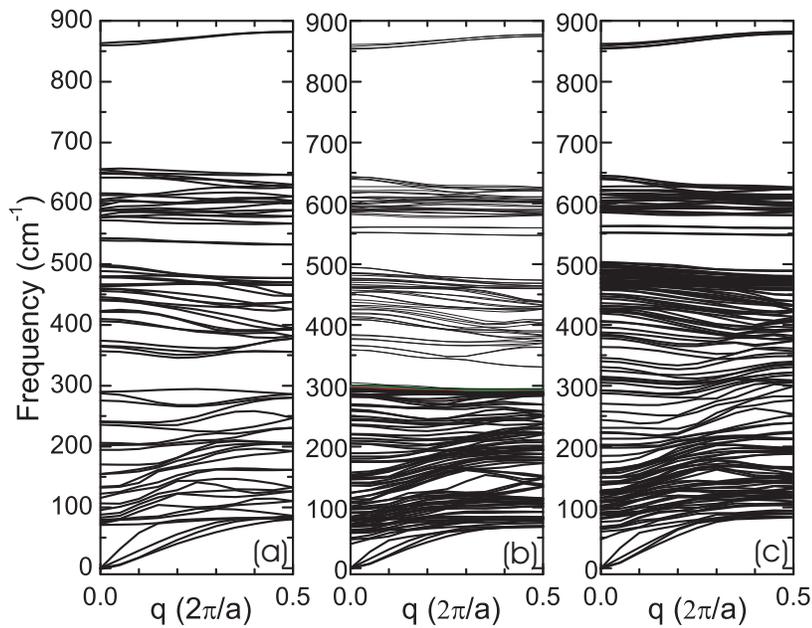


Figure 5.14: The phonon spectra of the Ge core nanowires corresponding to the structures depicted in Fig. 5.12. Frequencies higher than 900 cm^{-1} are not shown.

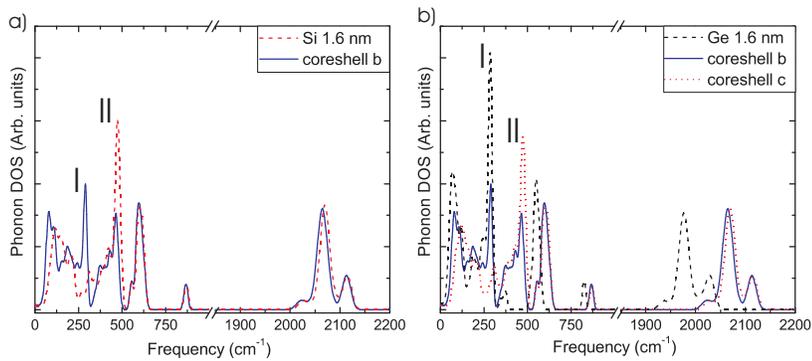


Figure 5.15: The phonon DOS of Ge core nanowires: (a) comparison between the coreshell structure b and a pure Si nanowire and (b) comparison between the coreshell structure b, a pure Ge nanowire, and the coreshell structure c. The peak labeled I is a typical Ge peak, while the peak labeled II is a typical Si peak.

mode (ω_b) are plotted in Fig. 5.16 as function of wire type (different symbols) and concentration of Ge (x-axis, ranging from 0, equal to no Ge at all, to 1, meaning only Ge atoms). Different dependencies can be observed: some properties (e.g. the transversal velocity v_T) only depend on the Ge concentration and not on the specific structure (i.e. the velocity does not depend on the type of atom in the core), while other frequencies (e.g. the bending mode ω_1) only depend on the type of atoms near the edges, which can be explained by the fact that these frequencies correspond to a Si-H or Ge-H bending mode. For the other properties the distinction between the two is not that clear. For the breathing mode ω_b it is rather the concentration that is important, while for the first optical frequency ω_σ and the longitudinal velocity v_L the type of wire is important which can be seen by the different behavior of the Si core versus the Ge core.

The hardening of optical frequencies and the softening of the acoustic frequencies (especially in case of the longitudinal velocity) with decreasing wire diameter, which was observed for both Si and Ge nanowires [93, 100] (see previous sections) is also present in the core-shell structures, as can be seen by comparing the obtained values for the 1.2 nm with the 1.6 nm nanowire.

This softening of the acoustic velocities with decreasing wire diameter in the different studied wires is in agreement with *ab initio* calculations [101, 102] of the Young's moduli, which increase with increasing wire diameter. Experimentally [103] it was also observed that the bending strength increases when the wire diameter increases. The specific phonon spectrum of each wire type can be used in experiments to distinguish between different wires using Raman spectroscopy, as was for example used by Kawashima *et al.* [72] to measure the Ge composition in SiGe nanowires and the effects of oxidation on this composition.

5.4 Conclusions

In conclusion we studied the phonon spectrum of silicon, germanium and Si/Ge core-shell nanowires, oriented in the [110] direction and passivated with hydrogens, by performing full *ab initio* DFPT calculations. We used the calculated phonon spectra to determine the stability of the nanowires used in chapter 3, as the occurrence of imaginary frequencies indicates structural instabilities. Our calculations indicate that, for the previously used unit cells of the nanowires (with diameters of 1.2 and 1.6 nm), all phonon

5.4. CONCLUSIONS

frequencies are real, also when doped with B or P. This is not the case for a smaller wire (diameter 0.5 nm), which has imaginary eigenfrequencies when doped, which is a clear indication that the wire is unstable (while stable when undoped). In that case, the wire can be stabilized by increasing the unit cell in the direction of the wire growth allowing for structural deformations. If these thin wires are already structurally stable, the same will be true for thicker wires. The identified phonon modes are also present in thicker wires. The effect of the wire diameter on the phonon frequencies and velocities is twofold: the longitudinal acoustic modes are softened when the wire diameter is decreased, while the optical modes are hardened. The calculated phonon DOS of these structures can be used to distinguish between the different core-shell types, e.g. it can be used to determine the core and shell materials and their relative concentrations. Optical frequencies and acoustic velocities can have two different types of dependencies: either on concentration or on structure.

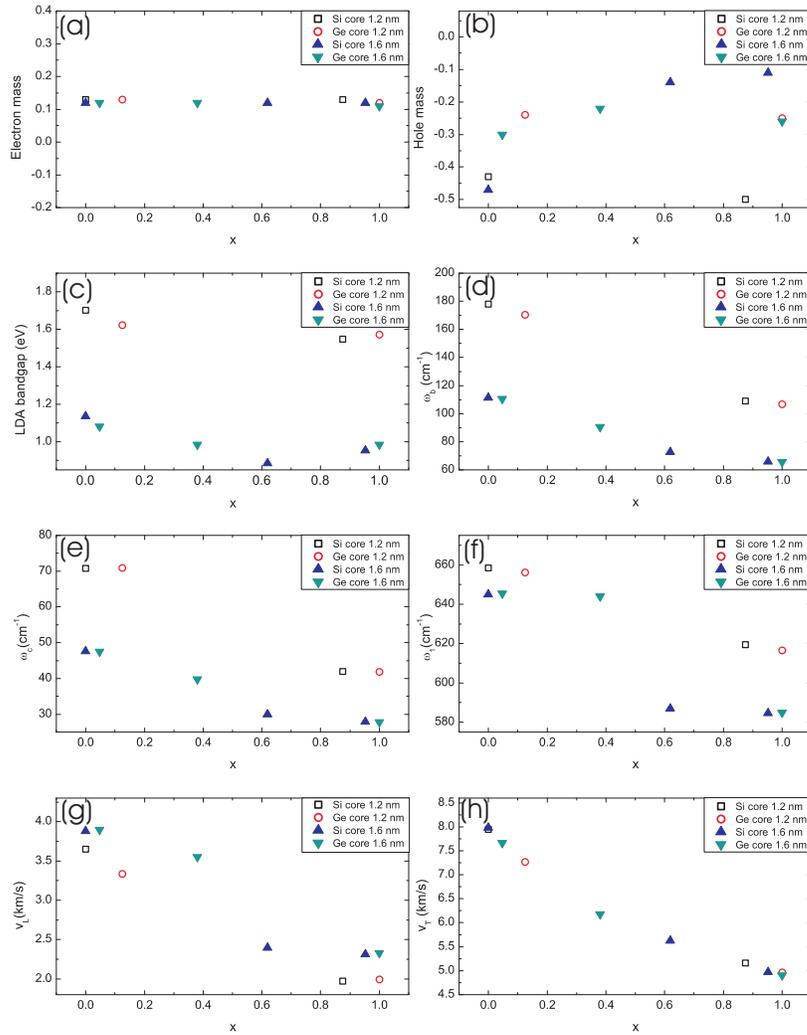


Figure 5.16: Effective masses, bandgaps, selected phonon frequencies and acoustic phonon velocities for the different core-shell nanowires. Each type of core-shell structure has a different symbol (depending on wire size and type of core atoms). The x axis depicts the concentration of Ge atoms, where 0 indicates the absence of Ge atoms (a pure Si nanowire) and 1 a wire only consisting of Ge. The different subplots show the following: (a) electron effective band mass, (b) hole effective band mass (c) the LDA bandgap, (d) the frequency of the breathing mode ω_b , (e) the frequency of the first optical mode ω_σ , (f) the frequency of the optical mode around 600 cm⁻¹ ω_1 , (g) the transversal sound velocity v_T , and (h) the longitudinal sound velocity v_L .

6

GW approximation: theory

In this chapter we give an introduction to the so-called GW approximation. The goal of this approximation is to calculate a more correct band structure. In the first section we explain why the band structures obtained using standard density-functional theory are not good enough, thus explaining the need of an improved calculation. This calculation is done in the GW approximation, which is an approximation derived from many-body perturbation theory, which makes use of the Green's functions. Therefore a small explanation of these functions and their properties is given. Next, we derive the Hedin equations [104], which are a set of coupled equations that can be solved to obtain the properties of interest (more specifically the self energy, which is necessary to calculate the two particle Green's function, when the one particle Green's function is known). In the last section we briefly discuss a practical G_0W_0 calculation.

This chapter is, in addition to the explicitly mentioned references, based on Refs. [105, 106, 107, 108, 109, 104].

6.1 The bandgap problem

The eigenvalues ϵ_i from DFT introduced in chapter 2 are completely artificial objects. These are the eigenvalues of a non-interacting system chosen to yield the same density as the many-body interacting system. Only the density has a physical meaning.

In fact, the only physically meaningful KS eigenvalue is the highest occupied one. If we use the notation $\epsilon_N(M)$ for the N'th eigenvalue of the M particle system, the negative of the ionization energy of the N-body system is given by $\epsilon_N(N) = -I$. Similarly the negative of the electron affinity is

given by $\epsilon_{N+1}(N+1) = -A$. When using an approximate xc functional the highest eigenvalues are not the best approximations to I and A . It is better to calculate them as total energy differences: $I = E(N-1) - E(N)$ and $A = E(N) - E(N+1)$, where $E(N)$ is the ground-state energy of the N electron system.

The bandstructures obtained by plotting the KS eigenvalues resemble in most cases only qualitatively the real band structure. If one uses these, one is in fact doing a mean-field approach to the many body theory. The bandgap Δ_{KS} obtained in this approach is severely underestimated because of the neglected derivative discontinuity Δ_{xc} by the standard local and semi-local xc functionals. Δ_{xc} is given by

$$\Delta_{xc} = \left. \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \right|_{N+\delta} - \left. \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \right|_{N-\delta} = v_{xc}^+ - v_{xc}^-, \quad (6.1)$$

where δ is an infinitesimal shift in the electron number. This is a system-dependent, but \mathbf{r} independent shift of the xc potential when going from the electron-rich to the electron-poor side of N . Similarly the bandgap Δ_{KS} is defined as

$$\Delta_{KS} = \left. \frac{\delta T_S[n]}{\delta n(\mathbf{r})} \right|_{N+\delta} - \left. \frac{\delta T_S[n]}{\delta n(\mathbf{r})} \right|_{N-\delta} = \epsilon_{N+1} - \epsilon_N. \quad (6.2)$$

The true bandgap is the discontinuity in the total ground state energy functional

$$\Delta = \left. \frac{\delta E[n]}{\delta n(\mathbf{r})} \right|_{N+\delta} - \left. \frac{\delta E[n]}{\delta n(\mathbf{r})} \right|_{N-\delta} = \Delta_{KS} + \Delta_{xc}, \quad (6.3)$$

where the last equality follows from the fact that only T_S and E_{xc} are not continuous functions. Standard exchange-correlation functionals (such as LDA and GGA) predict $\Delta_{xc} = 0$, so that this will lead to an underestimation of the bandgap. This is the so-called ‘‘bandgap problem’’. These energy levels are depicted in Fig. 6.1. An exact formulation of the gap can also be given as function of the ground-state energies by

$$\begin{aligned} E_g &= E(N+1) + E(N-1) - 2E(N) \\ &= [E(N+1) - E(N)] - [E(N) - E(N+1)] \\ &= -A(N) + I(N+1). \end{aligned} \quad (6.4)$$

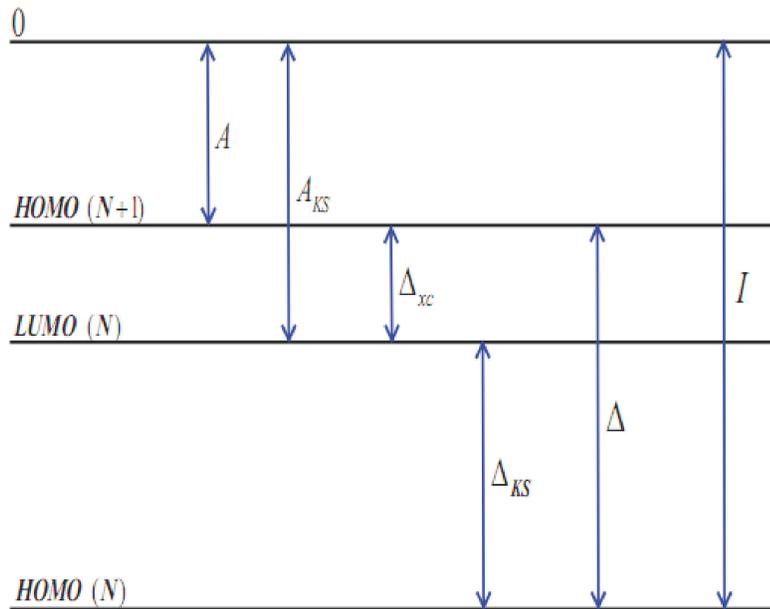


Figure 6.1: Some important Kohn-Sham eigenvalues and the different gaps (the Kohn-Sham gap Δ_{KS} and the many-body gap Δ) with respect to the vacuum level 0. The different electron ionization energies (I) and affinities (A and the Kohn-Sham A_{KS}) are shown.

In Fig. 6.2 the bandgaps calculated by LDA and experimentally determined are given. The underestimation is clearly visible. In the case of Ge the bandgap is even negative (semi-metal) when calculated within the LDA. As can be seen the application of the GW approximation leads to an improvement of the bandgaps. This approximation will be explained in the next sections.

6.2 Green's functions

When doing a many-body calculation, one needs to keep in mind what the goal of such calculation is: one is usually not interested in the many-body wavefunction itself, but in derived quantities. One will notice that the contained information in the Green's functions is smaller than the one contained

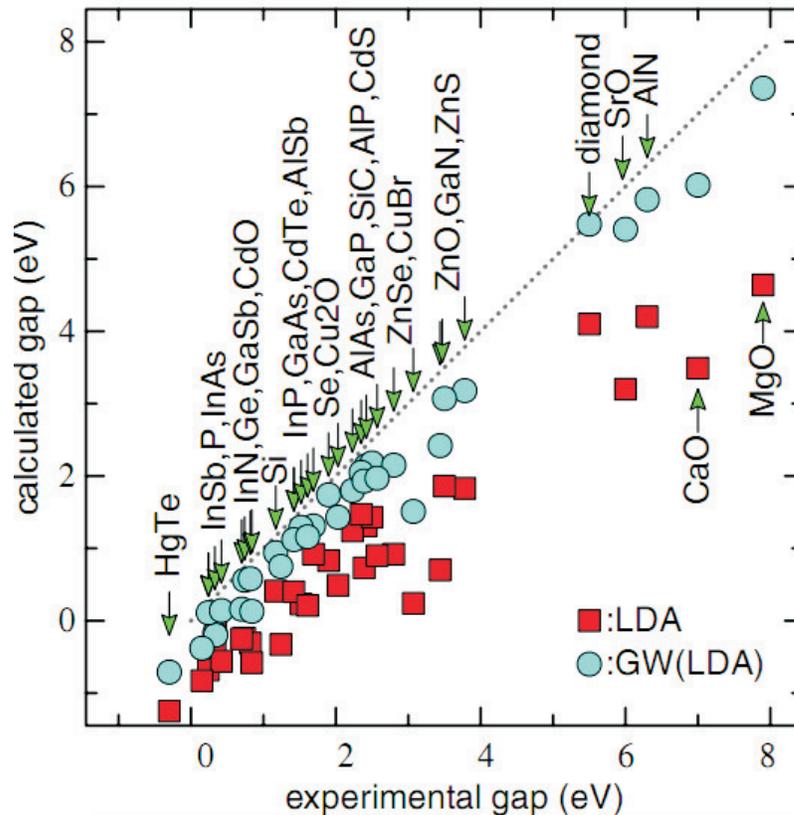


Figure 6.2: The experimental, LDA calculated and GW corrected values for different semiconductors. From Ref. [110].

in the many-body wavefunction, but it will still enable us to calculate the quantities of interest.

The one particle Green's function is given by

$$iG(1, 2) = \langle N | T[\hat{\psi}(1)\hat{\psi}^\dagger(2)] | N \rangle, \quad (6.5)$$

where $\hat{\psi}^\dagger$ and $\hat{\psi}$ are respectively the electron creation and annihilation operator and $|N\rangle$ is the normalized N particle wave function. The numbers (e.g. 1) are short hand notation for the position, time and spin variables $(\mathbf{r}_1, t_1, \sigma_1)$. T denotes the Wick time-ordering operator, defined as

$$T[\hat{\psi}(1)\hat{\psi}^\dagger(2)] = \begin{cases} \hat{\psi}(1)\hat{\psi}^\dagger(2) & t_1 > t_2 \\ -\hat{\psi}^\dagger(2)\hat{\psi}(1) & t_1 < t_2. \end{cases} \quad (6.6)$$

6.2. GREEN'S FUNCTIONS

This Green's function has a simple physical interpretation: when t_1 is larger than t_2 , it describes the probability to find an electron at time t_1 in \mathbf{r}_1 with spin σ_1 , when an electron was added at time t_2 in \mathbf{r}_2 with spin σ_2 to the ground-state system. When t_1 is smaller than t_2 , it describes the probability to find a hole in \mathbf{r}_2 with spin σ_2 at time t_2 , when an electron with spin σ_1 was removed from \mathbf{r}_1 at time t_1 .

Knowledge of the Green's function allows one to calculate the expectation value of any single-particle operator \hat{A} , given by

$$\begin{aligned}
 \hat{A} &= \int d^3r_1 \hat{\psi}^\dagger(\mathbf{r}_1) A(\mathbf{r}_1) \hat{\psi}(\mathbf{r}_1) \\
 \langle N | \hat{A}(\mathbf{r}_1) | N \rangle &= \int d^3r_1 \hat{A}(\mathbf{r}_1) \langle N | \hat{\psi}^\dagger(\mathbf{r}_1) \hat{\psi}(\mathbf{r}_1) | N \rangle \\
 &= \lim_{\mathbf{r}_2 \rightarrow \mathbf{r}_1} \int d^3r_1 \hat{A}(\mathbf{r}_1) \langle N | \hat{\psi}^\dagger(\mathbf{r}_1) \hat{\psi}(\mathbf{r}_2) | N \rangle \\
 &= - \lim_{\mathbf{r}_2 \rightarrow \mathbf{r}_1} \lim_{t_2 \rightarrow t_1^+} \int d^3r_1 \delta(t_2 - t_1^+) \hat{A}(\mathbf{r}_1) \langle N | T[\hat{\psi}(\mathbf{r}_1) \hat{\psi}^\dagger(\mathbf{r}_1)] | N \rangle \\
 &= -i \lim_{\mathbf{r}_2 \rightarrow \mathbf{r}_1} \lim_{t_2 \rightarrow t_1^+} \int d1 A(1) G(11^+), \tag{6.7}
 \end{aligned}$$

where t_1^+ is equal to $t_1 + \delta$, with δ a positive infinitesimal real number. Applying this to the particle density at point \mathbf{r}_1 gives $n(1) = -iG(11^+)$. The ground state energy can also be given as function of the one particle Green's functions by using the Galitskii-Migdal [111] formula

$$E = -\frac{i}{2} \sum_{\sigma_1 \sigma_2} \delta_{\sigma_1 \sigma_2} \int d^3r_1 \lim_{\mathbf{r}_2 \rightarrow \mathbf{r}_1} \lim_{t_2 \rightarrow t_1^+} \left[i \frac{\partial}{\partial t_1} + h_0(\mathbf{r}_1) \right] G(1, 2), \tag{6.8}$$

where h_o is the one particle Hamiltonian (for a proof see e.g. Ref. [105]).

Using the one-particle Green's function one can also obtain the excitation energies. This can be seen by introducing the closure relation in equation 6.5, where we denote the time difference as $\tau = t_1 - t_2$:

$$\begin{aligned}
 iG(\mathbf{r}_1, \mathbf{r}_2, \tau) &= \theta(\tau) \sum_i \langle N | \hat{\psi}(1) | N + 1, i \rangle \langle N + 1, i | \hat{\psi}^\dagger(2) | N \rangle \\
 &\quad - \theta(-\tau) \sum_i \langle N + 1, i | \hat{\psi}^\dagger(2) | N \rangle \langle N | \hat{\psi}(1) | N + 1, i \rangle. \tag{6.9}
 \end{aligned}$$

CHAPTER 6. GW APPROXIMATION: THEORY

The only terms of the closure relation that give non zero contributions are the $N + 1$ and $N - 1$ particle state, respectively in the case of τ positive or negative. The Green's function only depends on τ in the absence of any time-dependent external potential. The time dependence becomes clearer when transformed to the Schrödinger picture

$$\hat{\Psi}(1) = e^{i\hat{H}t_1}\hat{\Psi}(\mathbf{r}_1)e^{-i\hat{H}t_1}. \quad (6.10)$$

This gives

$$\begin{aligned} iG(\mathbf{r}_1, \mathbf{r}_2, \tau) &= \theta(\tau) \sum_i \langle N | \hat{\psi}(\mathbf{r}_1) | N + 1, i \rangle \langle N + 1, i | \hat{\psi}^\dagger(\mathbf{r}_2) | N \rangle e^{i(E_N - E_{N+1,i})\tau} \\ &\quad - \theta(-\tau) \sum_i \langle N | \hat{\psi}^\dagger(\mathbf{r}_2) | N - 1, i \rangle \langle N - 1, i | \hat{\psi}(\mathbf{r}_1) | N \rangle e^{i(E_{N-1,i} - E_N)\tau}. \end{aligned} \quad (6.11)$$

In this equation the excitation energies ϵ_i defined as

$$\begin{aligned} \epsilon_i^{N-1} &= E_N - E_{N-1,i} \\ \epsilon_i^{N+1} &= E_{N+1,i} - E_N \end{aligned} \quad (6.12)$$

can be introduced, giving

$$\begin{aligned} iG(\mathbf{r}_1, \mathbf{r}_2, \tau) &= \theta(\tau) \sum_i \langle N | \hat{\psi}(\mathbf{r}_1) | N + 1, i \rangle \langle N + 1, i | \hat{\psi}^\dagger(\mathbf{r}_2) | N \rangle e^{-i\epsilon_i^{N+1}\tau} \\ &\quad - \theta(-\tau) \sum_i \langle N | \hat{\psi}^\dagger(\mathbf{r}_2) | N - 1, i \rangle \langle N - 1, i | \hat{\psi}(\mathbf{r}_1) | N \rangle e^{-i\epsilon_i^{N-1}\tau}. \end{aligned} \quad (6.13)$$

This expression can be simplified by defining

$$\begin{aligned} f_i^{N-1}(\mathbf{r}) &= \langle N - 1, i | \hat{\psi}(\mathbf{r}_1) | N \rangle \\ f_i^{N+1}(\mathbf{r}) &= \langle N | \hat{\psi}(\mathbf{r}_1) | N + 1, i \rangle, \end{aligned} \quad (6.14)$$

so that equation 6.13 becomes

$$\begin{aligned} iG(\mathbf{r}_1, \mathbf{r}_2, \tau) &= \theta(\tau) \sum_i f_i^{N+1}(\mathbf{r}_1) f_i^{N+1*}(\mathbf{r}_2) e^{-i\epsilon_i^{N+1}\tau} \\ &\quad - \theta(-\tau) \sum_i f_i^{N-1}(\mathbf{r}_1) f_i^{N-1*}(\mathbf{r}_2) e^{-i\epsilon_i^{N-1}\tau}. \end{aligned} \quad (6.15)$$

6.3. DYSON EQUATIONS AND SELFENERGIES

We already see from this equation that the Green's function contains the complete excitation spectra. It is more convenient to work in the frequency domain, therefore we take the Fourier transform of 6.15, by making use of the Fourier transform of the Heaviside step function $\theta(\tau)$:

$$\theta(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\tau \theta(\tau) e^{i\omega\tau - \eta|\tau|} = \frac{i}{2\pi(\omega + i\eta)}, \quad (6.16)$$

resulting in

$$G(\mathbf{r}_1, \mathbf{r}_2, \omega) = \sum_i \frac{f_i^{N+1}(\mathbf{r}_1) f_i^{N+1*}(\mathbf{r}_2)}{\omega - \epsilon_i^{N+1} + i\eta} + \sum_i \frac{f_i^{N-1}(\mathbf{r}_1) f_i^{N-1*}(\mathbf{r}_2)}{\omega - \epsilon_i^{N+1} - i\eta}. \quad (6.17)$$

This result is the so-called Lehmann representation of the Green's function, which can also be rewritten by dropping the indices $N \pm 1$ and where it is assumed that $\pm\eta$ corresponds to the $N \pm 1$ system

$$G(\mathbf{r}_1, \mathbf{r}_2, \omega) = \sum_i \frac{f_i(\mathbf{r}_1) f_i^*(\mathbf{r}_2)}{\omega - \epsilon_i \pm i\eta}. \quad (6.18)$$

One can observe that the Green's functions have poles at the many-particle excitation energies. These values can be measured experimentally by photo-emission or inverse photo-emission. For semiconductors this will thus yield the size of the exact bandgap.

6.3 Dyson equations and selfenergies

The many-body Hamiltonian can be rewritten in second quantization, giving

$$\hat{H} = \int d^3r \hat{\Psi}^\dagger(\mathbf{r}) \hat{h}(\mathbf{r}) \hat{\Psi}(\mathbf{r}) + \frac{1}{2} \int \int d^3r d^3r' \hat{\Psi}^\dagger(\mathbf{r}) \hat{\Psi}^\dagger(\mathbf{r}') v(\mathbf{r}, \mathbf{r}') \hat{\Psi}(\mathbf{r}) \hat{\Psi}(\mathbf{r}'), \quad (6.19)$$

where $v(\mathbf{r}, \mathbf{r}')$ is the Coulomb interaction. The equation of motion in the Heisenberg picture of the annihilation operator is given by

$$i \frac{\partial}{\partial t} \hat{\Psi}(1) = [\hat{\Psi}(1), \hat{H}] = \hat{h}(1) \hat{\Psi}(1) + \int d2 v(1, 2) \hat{\Psi}^\dagger(2) \hat{\Psi}(2) \hat{\Psi}(1), \quad (6.20)$$

CHAPTER 6. GW APPROXIMATION: THEORY

so that the equation of motion of the Green's function 6.5 becomes

$$i\frac{\partial}{\partial t}G(1,2) = \delta(1,2) + \hat{h}(1)G(1,2) - i \int d3v(1^+,3) \langle N|T[\hat{\Psi}(1)\hat{\Psi}^\dagger(2)\hat{\Psi}^\dagger(3)\hat{\Psi}(3)]|N\rangle. \quad (6.21)$$

This is not a closed equation, because it contains the two-particle Green's function

$$G_2(1234) = -\langle N|T[\hat{\Psi}(1)\hat{\Psi}(2)\hat{\Psi}^\dagger(3)\hat{\Psi}^\dagger(4)]|N\rangle. \quad (6.22)$$

By using equation 6.22, the time evolution of the Green's function 6.21 can be rewritten as

$$i\frac{\partial}{\partial t}G(1,2) = \delta(1,2) + \hat{h}(1)G(1,2) - i \int d3v(1^+,3)G_2(1,3,2,3^+). \quad (6.23)$$

So in order to calculate the one-particle Green's function, one should know the two-particle Green's function, for the two-particle the three-particle,...

In order to stop this iteration scheme (which is as complex as the original problem), one uses many-body perturbation theory: in order to know the one-particle properties, one only needs to know the one-particle Green's function. In order to do that, a suitable expression of the two-particle Green's function in terms of the one-particle one should be found. This is done by defining the self-energy operator Σ as

$$\int d3\Sigma(1,3)G(32) = -i \int d3v(1^+,3)G_2(1,3,2,3^+), \quad (6.24)$$

so that equation 6.23 reads

$$\left[i\frac{\partial}{\partial t_1} - h(\mathbf{r}_1) \right] G(1,2) - \int d3\Sigma(1,3)G(3,2) = \delta(1,2). \quad (6.25)$$

This equation is the so-called Dyson equation.

This equation can be rewritten in the frequency domain, which gives some additional insight in the meaning of the Green's function. After a Fourier transform it becomes

$$[\omega - h(\mathbf{r})] G(\mathbf{r}, \mathbf{r}', \omega) - \int d^3r'' \Sigma(\mathbf{r}, \mathbf{r}'', \omega) G(\mathbf{r}'', \mathbf{r}', \omega) = \delta(\mathbf{r} - \mathbf{r}'). \quad (6.26)$$

6.3. DYSON EQUATIONS AND SELFENERGIES

Insertion of the Green's function in the Lehmann representation (see equation 6.18) yields

$$\sum_i \frac{f_i^*(\mathbf{r}')}{\omega - \epsilon_i \pm i\eta} \left([\omega - h(\mathbf{r})] f_i(\mathbf{r}) - \int d^3r'' \Sigma(\mathbf{r}, \mathbf{r}'', \omega) f_i(\mathbf{r}) \right) = \delta(\mathbf{r} - \mathbf{r}'). \quad (6.27)$$

In the following it is assumed that the system is not degenerated. Multiplication with $(\omega - \epsilon_j)$ and taking the limit $\omega \rightarrow \epsilon_j$ gives for the left-hand side

$$\begin{aligned} \lim_{\omega \rightarrow \epsilon_j} (\omega - \epsilon_j) \sum_i \frac{f_i^*(\mathbf{r}')}{\omega - \epsilon_i \pm i\eta} \left([\omega - h(\mathbf{r})] f_i(\mathbf{r}) - \int d^3r'' \Sigma(\mathbf{r}, \mathbf{r}'', \omega) f_i(\mathbf{r}) \right) \\ = f_j^*(\mathbf{r}') \left([\epsilon_j - h(\mathbf{r})] f_j(\mathbf{r}) - \int d^3r'' \Sigma(\mathbf{r}, \mathbf{r}'', \epsilon_j) f_j(\mathbf{r}) \right), \end{aligned} \quad (6.28)$$

while the right-hand side becomes

$$\lim_{\omega \rightarrow \epsilon_j} \delta(\mathbf{r} - \mathbf{r}') = 0. \quad (6.29)$$

$f_j^*(\mathbf{r}')$ will not vanish for all \mathbf{r}' values, so that the quasiparticle equation

$$h(\mathbf{r}) f_j(\mathbf{r}) + \int d^3r'' \Sigma(\mathbf{r}, \mathbf{r}'', \epsilon_j) f_j(\mathbf{r}) = \epsilon_j f_j(\mathbf{r}) \quad (6.30)$$

is obtained. This equation is also valid for the degenerate case: in this case an arbitrary perturbation $\hat{\phi}$ is introduced in \hat{h} , so that the symmetry is broken and the degeneracy is lifted. The above proof can then be used. Equation 6.30 is then obtained by taking the limit of vanishing external perturbation $\hat{\phi}$.

This equation shows that the self-energy can be seen as a non-local, frequency dependent potential in a Schrödinger-like equation of motion of the Lehmann amplitudes and energies. It should be mentioned that, contrary to the Kohn-Sham equations, this is not a mean-field formulation. The functions f_i and their energies ϵ_i are not single particle quantities, but these are defined as properties of the many-particle system (see equations 6.17 and 6.12).

CHAPTER 6. GW APPROXIMATION: THEORY

In the case of non-interacting fermions, the equation of motion for the non-interacting Green's function, denoted as G_0 , is

$$\left[i \frac{\partial}{\partial t_1} - h(\mathbf{r}_1) \right] G_0(1, 2) = \delta(1, 2). \quad (6.31)$$

This gives the usual interpretation of a Green's function: it is the formal inverse of the operator $(\omega - h)$ in frequency space. Using this interpretation of G_0^{-1} , the Dyson equation can be rewritten as

$$[G_0^{-1}(1, 3) - \Sigma(1, 3)] G(3, 2) = \delta(1, 2). \quad (6.32)$$

Symbolically, this can be written as

$$G = G_0 + G_0 \Sigma G. \quad (6.33)$$

Thus the self-energy can be seen as the connection between the non-interacting system (with Green's function G_0) and the interacting one (G). The self-energy Σ includes all effects of the interacting particles.

Usually one puts the Hartree potential into the one-particle Hamiltonian $h_0(\mathbf{r}_1) = h(\mathbf{r}_1) + v_H$, so that the self-energy denotes the remainder $\Sigma - v_H$. This new definition of the self energy is used from now on.

6.4 Hedin equations

To be able to obtain the one-particle Green's functions, the self-energy needs to be known. The Hedin's equations will give a set of equations that can be used to obtain this self-energy. To derive this set of equations the derivative technique of Schwinger is used. In this technique a small time dependent external potential $U(\mathbf{r}_1, \mathbf{r}_2, t)$ is introduced, that will be set to zero at the end of the derivation.

The one-particle Green's function in the presence of this potential becomes

$$iG(1, 2) = \frac{\langle N | T[\hat{S} \hat{\Psi}(1) \hat{\Psi}^\dagger(2)] | N \rangle}{\langle N | T[\hat{S}] | N \rangle}, \quad (6.34)$$

where the creation and annihilation operator (respectively $\hat{\Psi}^\dagger$ and $\hat{\Psi}$) are as previously defined. Because of the presence of the time-dependent potential,

they are now given in the interaction picture, so they still contain the time-dependence of the equilibrium Hamiltonian, but the time dependence due to the additional potential is contained in the time evolution operator \hat{S} . The definition of this operator is

$$T[\hat{S}] = T \left[\exp \left(-i \int_{-\infty}^{\infty} dt \int d^3r_1 d^3r_2 \hat{\Psi}^\dagger(\mathbf{r}_1, t^+) U(\mathbf{r}_1, \mathbf{r}_2, t) \hat{\Psi}(\mathbf{r}_2, t) \right) \right]. \quad (6.35)$$

The complete dependence on U is in this operator only. If $U = 0$, this gives again the previously used equilibrium Green's function 6.5. The first-order variation of $\delta G(1, 2)$ with respect to $\delta U(\mathbf{r}_3, \mathbf{r}_4, t_3)$ is thus given by

$$i\delta G(1, 2) = \frac{\langle N | T[\delta \hat{S} \hat{\Psi}(1) \hat{\Psi}^\dagger] | N \rangle}{\langle N | T[\hat{S}] | N \rangle} - iG(1, 2) \frac{\langle N | T[\delta \hat{S}] | N \rangle}{\langle N | T[\hat{S}] | N \rangle}. \quad (6.36)$$

Using definition 6.35 the variation $\delta \hat{S}$ can be obtained by taking the derivation of an exponential, taking into account that the product of operators $\hat{\Psi}^\dagger \hat{\Psi}$ commutes with other two-field operator products in a time-ordered product T

$$T[\delta \hat{S}] = -iT \left[\hat{S} \int_{-\infty}^{\infty} dt \int d^3r_3 d^3r_4 \hat{\Psi}^\dagger(\mathbf{r}_3, t^+) \delta U(\mathbf{r}_3, \mathbf{r}_4, t) \hat{\Psi}(\mathbf{r}_4, t) \right]. \quad (6.37)$$

Combining 6.36 and 6.37 and inserting the definition of one-particle and two-particle Green's functions gives

$$\delta G(1, 2) = - \int_{-\infty}^{\infty} dt \int d^3r_3 d^3r_4 \delta U(\mathbf{r}_3, \mathbf{r}_4, t) [G_2(1, \mathbf{r}_4 t, 2, \mathbf{r}_3 T^+) - G(1, 2) G(\mathbf{r}_4 t, \mathbf{r}_3 t^+)]. \quad (6.38)$$

From this equation we can obtain the final result of the Schwinger derivative technique, which is also valid for $t_3 \neq t^+$

$$\frac{\delta G(1, 2)}{\delta U(3, 4)} = -G_2(1, 4, 2, 3) + G(1, 2) G(4, 3). \quad (6.39)$$

This equation gives the connection between the two-particles Green's function and the one-particle Green's function. For a local potential $U(3)\delta(3, 4)$ this gives

$$\frac{\delta G(1, 2)}{\delta U(3)} = -G_2(1, 3, 2, 3^+) + G(1, 2) G(3, 3^+), \quad (6.40)$$

which can be used to rewrite the equation of motion 6.23

$$\begin{aligned} \left[i \frac{\partial}{\partial t_1} - \hat{h}(\mathbf{r}_1) \right] G(1, 2) + i \int d3v(1, 3)G(3, 3^+)G(1, 2) \\ - i \int d3v(1^+, 3) \frac{\delta G(1, 2)}{\delta U(3)} = \delta(1, 2). \end{aligned} \quad (6.41)$$

Introducing the identity $\int d5G^{-1}(4, 5)G(5, 2) = \delta(1, 2)$ finally gives

$$\begin{aligned} \left[i \frac{\partial}{\partial t_1} - \hat{h}(\mathbf{r}_1) + i \int d3v(1, 3)G(3, 3^+) \right] G(1, 2) \\ - i \left[\int d345v(1^+, 3) \frac{\delta G(1, 4)}{\delta U(3)} G^{-1}(4, 5) \right] G(5, 2) = \delta(1, 2). \end{aligned} \quad (6.42)$$

The third term in this expression is the Hartree potential, as $-iG(3, 3^+)$ is the electronic density. Comparing this expression with 6.32, gives an equation for the self-energy

$$\Sigma(1, 2) = i \int d345v(1^+, 3) \frac{\delta G(1, 4)}{\delta U(3)} G^{-1}(4, 5), \quad (6.43)$$

or by using the identity

$$\frac{\delta}{\delta U}(G^{-1}G) = G^{-1} \frac{\delta G}{\delta U} + \frac{\delta G^{-1}}{\delta U} G = 0, \quad (6.44)$$

this is

$$\Sigma(1, 2) = -i \int d345v(1^+, 3)G(1, 4) \frac{\delta G^{-1}(4, 5)}{\delta U(3)}. \quad (6.45)$$

To be able to evaluate this expression, a perturbative approach will be used, based on the interacting Green's function G and the dynamical screened Coulomb interaction. Earlier attempts to expand it using the non-interacting Green's function G_0 and the bare Coulomb potential v failed. Intuitively the screened Coulomb interaction W can be seen as being smaller than the bare one, as interactions between electrons are reduced by the screening of the other electrons. The exchange and correlation effects form a Coulomb hole around each electron. This combination of the Coulomb hole and the electron can be seen as a quasi-particle. It are these quasi-particles that interact

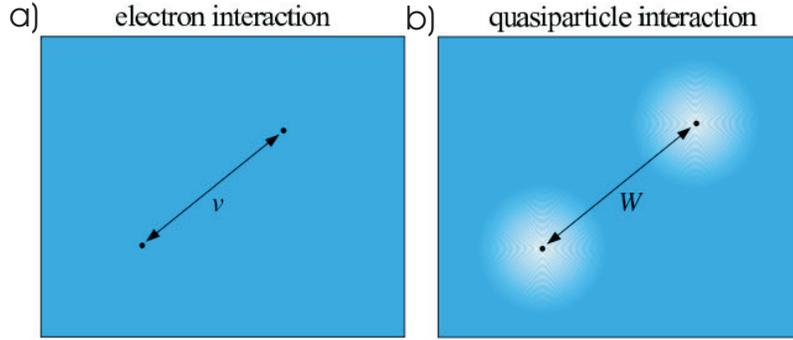


Figure 6.3: Schematic picture that shows the difference between (a) the bare Coulomb interaction v and (b) the screened interaction W . From Ref. [106].

with each other. This is schematically depicted in Fig. 6.3, which shows in (a) the Coulomb interaction between electrons and in (b) the quasi-particle interaction, interacting via the weaker screened interaction W .

This derivation follows Hedin. First a local classical potential V , which is the sum of the perturbation U and the Hartree potential is defined

$$V(1) = U(1) - i \int d2v(1, 2)G(2, 2^+). \quad (6.46)$$

This will be used to replace all references to U with references to V . Doing this by applying the chain-rule of derivatives in equation 6.45 of the self-energy, gives

$$\Sigma(1, 2) = -i \int d345v(1^+, 3) \frac{\delta G^{-1}(1, 4)}{\delta V(5)} \frac{\delta V(5)}{\delta U(3)} G(4, 2). \quad (6.47)$$

This can be rewritten by defining the time-ordered inverse dielectric function ϵ^{-1}

$$\epsilon^{-1}(1, 2) = \frac{\delta V(1)}{\delta U(2)}, \quad (6.48)$$

the irreducible vertex function $\tilde{\Gamma}$

$$\tilde{\Gamma}(1, 2; 3) = -\frac{\delta G^{-1}(1, 2)}{\delta V(3)}, \quad (6.49)$$

CHAPTER 6. GW APPROXIMATION: THEORY

and the dynamical screened Coulomb interaction W

$$W(1, 2) = \int d3v(1, 3)\epsilon^{-1}(3, 2). \quad (6.50)$$

“Irreducible” means that the differentiation is done with respect to V and not to U . The self-energy then becomes

$$\Sigma(1, 2) = i \int d34G(1, 4)W(3, 1^+)\tilde{\Gamma}(4, 2; 3). \quad (6.51)$$

Now the irreducible vertex function $\tilde{\Gamma}$ and the inverse dielectric function ϵ^{-1} will be worked out. For this the Dyson equation 6.42 $G^{-1} = G_0^{-1} - V - \Sigma$ is used in equation 6.49

$$\begin{aligned} \tilde{\Gamma}(1, 2; 3) &= \delta(1, 2)\delta(1, 3) + \frac{\delta\Sigma(1, 2)}{\delta V(3)} \\ &= \delta(1, 2)\delta(1, 3) + \int d45 \frac{\delta\Sigma(1, 2)}{\delta G(4, 5)} \frac{\delta G(4, 5)}{\delta V(3)} \\ &= \delta(1, 2)\delta(1, 3) - \int d4567 \frac{\delta\Sigma(1, 2)}{\delta G(4, 5)} G(4, 6) \frac{\delta G^{-1}(6, 7)}{\delta V(3)} G(7, 5) \\ &= \delta(1, 2)\delta(1, 3) + \int d4567 \frac{\delta\Sigma(1, 2)}{\delta G(4, 5)} G(4, 6) G(7, 5) \tilde{\Gamma}(6, 7; 3). \end{aligned} \quad (6.52)$$

In the second line the chain rule was used. The third line is obtained after using the identity 6.44 and in the last line the definition of the irreducible vertex function 6.49 was used.

For ϵ^{-1} , the definition of V is used

$$\begin{aligned} \epsilon^{-1}(1, 2) &= \frac{\delta(U(1) - i \int d3v(1, 3)G(3, 3^+))}{\delta U(2)} \\ &= \delta(1, 2) + \int d3v(1, 3)\chi(3, 2), \end{aligned} \quad (6.53)$$

where

$$\chi(1, 2) = -i \frac{\delta G(1, 1^+)}{\delta U(2)} \quad (6.54)$$

is the reducible (as the bare potential U is used) polarizability of the system. Similarly the irreducible polarizability $\tilde{\chi}$ can be defined

$$\tilde{\chi}(1, 2) = -i \frac{\delta G(1, 1^+)}{\delta V(2)}. \quad (6.55)$$

The connection between these two polarizabilities can be found by applying the chain-rule

$$\begin{aligned} \chi(1, 2) &= -i \int d3 \frac{\delta G(1, 1^+)}{\delta V(3)} \frac{\delta V(3)}{\delta U(2)} \\ &= \int d3 \tilde{\chi}(1, 3) \epsilon^{-1} \\ &= \tilde{\chi}(1, 2) + \int d34 \tilde{\chi}(1, 3) v(3, 4) \chi(4, 2). \end{aligned} \quad (6.56)$$

In the second line the definitions 6.55 and 6.48 are used. The last line is obtained by inserting equation 6.53.

The last remaining quantity $\tilde{\chi}$ can be obtained from 6.55, by applying the identity 6.44 and inserting the irreducible vertex function 6.49

$$\begin{aligned} \tilde{\chi}(1, 2) &= i \int d34 G(1, 3) \frac{\delta G^{-1}(3, 4)}{\delta V(2)} G(4, 1) \\ &= -i \int d34 G(1, 3) G(4, 1) \tilde{\Gamma}(3, 4; 2). \end{aligned} \quad (6.57)$$

The inverse dielectric function 6.53 can now be rewritten by inserting equation 6.56 repeatedly

$$\begin{aligned} \epsilon^{-1}(1, 2) &= \delta(1, 2) + \int d34 v(1, 3) \chi(3, 4) \\ &\quad \times \left[\delta(4, 2) + \int d56 v(4, 5) \chi(5, 6) \delta(6, 2) + \dots \right] \end{aligned} \quad (6.58)$$

$$= \delta(1, 2) + \int d34 v(1, 3) \chi(3, 4) \epsilon^{-1}(4, 2), \quad (6.59)$$

which can be used to get rid of the ϵ^{-1} dependency of W (equation 6.50)

$$\begin{aligned} W(1, 2) &= \int d3 \delta(1, 3) v(3, 2) + \int d345 v(1, 3) \chi(3, 4) \epsilon^{-1}(4, 3) v(3, 2) \\ &= v(1, 2) + \int d34 v(1, 3) \chi(3, 4) W(4, 2). \end{aligned} \quad (6.60)$$

CHAPTER 6. GW APPROXIMATION: THEORY

As all quantities are now properly defined, the limit of U going to zero can be safely taken.

To summarize the five Hedin equations are given below.

$$\begin{aligned}
 G(1, 2) &= G_0(1, 2) + \int d^3 4 G_0(1, 3) \Sigma(3, 4) G(4, 2) \\
 \tilde{\Gamma}(1, 2; 3) &= \delta(1, 2) \delta(1, 3) + \int d^4 5 6 7 \frac{\delta \Sigma(1, 2)}{\delta G(4, 5)} G(4, 6) G(7, 5) \tilde{\Gamma}(6, 7; 3) \\
 \tilde{\chi}(1, 2) &= -i \int d^3 4 G(2, 3) G(4, 2) \tilde{\Gamma}(3, 4; 1) \\
 W(1, 2) &= v(1, 2) + \int d^3 4 v(1, 3) \tilde{\chi}(3, 4) W(4, 2) \\
 \Sigma(1, 2) &= i \int d^3 4 G(1, 4) W(3, 1^+) \tilde{\Gamma}(4, 2; 3)
 \end{aligned}$$

This set of equations can be used to, starting from a certain approximation of the self-energy Σ , obtain higher order approximations by iterating through the equations. This is schematically depicted in Fig. 6.4, which is also called the Hedin “pentagon”. Unfortunately they are not numerical equations, as they contain a functional derivative. Therefore they can not be solved self-consistently by a computer code, but have to be iterated analytically to obtain an approximation. The simplest approximation is to start with

$$\tilde{\Gamma}(1, 2; 3) = \delta(1, 2) \delta(1, 3) \quad (6.61)$$

as the vertex equation, which yields the GW approximation:

$$\tilde{\Gamma}(1, 2; 3) = \delta(1, 2) \delta(1, 3) \quad (6.62)$$

$$\tilde{\chi}(1, 2) = -i G(1, 2) G(2, 1) \quad (6.63)$$

$$W(1, 2) = v(1, 2) + \int d^3 4 v(1, 3) \tilde{\chi}(3, 4) W(4, 2) \quad (6.64)$$

$$\Sigma(1, 2) = i G(1, 2) W(2, 1^+) \quad (6.65)$$

$$G(1, 2) = G_0(1, 2) + \int d^3 4 G_0(1, 3) \Sigma(3, 4) G(4, 2). \quad (6.66)$$

The naming of this approximation refers to equation 6.65 of the self-energy, as it is given by iGW .

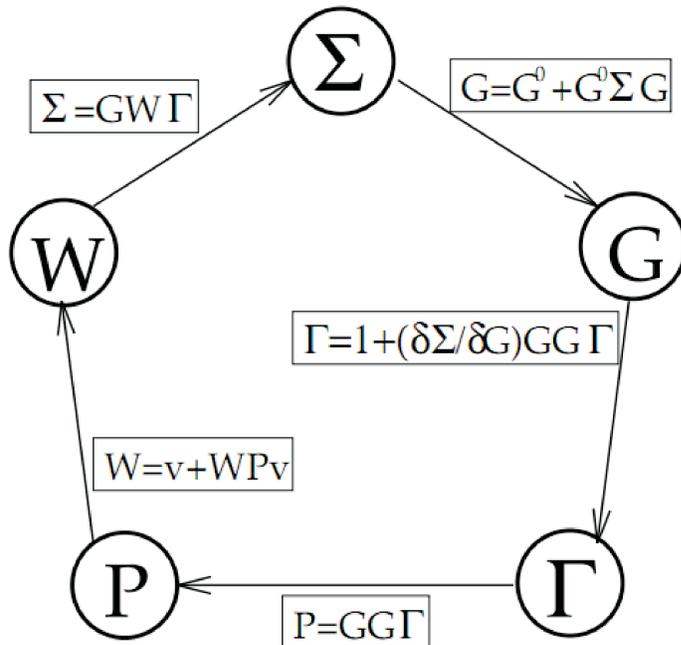


Figure 6.4: A schematic representation of the iterative determination of the self-energy using the Hedin equations, from Ref. [109].

If only one iteration is performed, starting from $\Sigma = 0$, so that $G = G_{KS} \equiv G_0$, one obtains the G_0W_0 approximation or “single-shot” GW, with as main equations:

$$\tilde{\chi}_0(1, 2) = -iG_0(1, 2)G_0(2, 1) \quad (6.67)$$

$$W_0(1, 2) = v(1, 2) + \int d^3x v(1, 3)\tilde{\chi}_0(3, 4)W_0(4, 2) \quad (6.68)$$

$$\Sigma(1, 2) = iG_0(1, 2)W_0(2, 1^+) \quad (6.69)$$

Equation 6.68 of the irreducible polarizability is equal to the one of independent particles (denoted χ_0). This approximation is only justified if the sum of higher order terms of W is small compared to the first order result. This approximation can be seen as a generalization of the Hartree-Fock approximation, where the bare Coulomb interaction is replaced by the screened one. From the full formula for the polarization, containing the vertex function,

it follows that the GW approximation is equivalent by only considering the contribution of the electron-hole pair excitations to the polarizability. All other interactions, such as the interaction between electron and hole and between electron-hole pairs are neglected.

6.5 Practical G_0W_0 calculations

In this section a brief overview of a practical calculation within a planewave framework is given, which is used in the next chapter.

In the frequency domain the self energy within the GW approximation becomes the convolution

$$\Sigma^{GW}(\mathbf{r}, \mathbf{r}'; \omega) = \frac{i}{2\pi} \int d\omega' e^{i\omega'\eta} G_0(\mathbf{r}, \mathbf{r}'; \omega + \omega') W(\mathbf{r}, \mathbf{r}'; \omega'), \quad (6.70)$$

where the screened interaction is given by

$$W(\mathbf{r}, \mathbf{r}'; \omega) = \int d^3r'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}''; \omega) v(\mathbf{r}'' - \mathbf{r}'). \quad (6.71)$$

This indicates that the interaction between quasi-particles is much smaller than the “bare” Coulomb interaction between the electrons, due to the repulsion of the electrons in the neighborhood of an electron, i.e. an exchange-correlation hole is created which screens the interaction. The dielectric function in the frequency domain is (Fourier transformed inverse of equation 6.53)

$$\epsilon(\mathbf{r}, \mathbf{r}'; \omega) = \delta(\mathbf{r} - \mathbf{r}') - \int d^3r'' v(\mathbf{r}'' - \mathbf{r}') \tilde{\chi}(\mathbf{r}'', \mathbf{r}'; \omega), \quad (6.72)$$

where the irreducible polarizability (in the RPA) can be shown [109] to be equal to

$$\begin{aligned} \tilde{\chi}_0(\mathbf{r}, \mathbf{r}'; \omega) = & \sum_{n\mathbf{k}, n'\mathbf{k}'} \phi_{n\mathbf{k}}(\mathbf{r}) \phi_{n'\mathbf{k}'}^*(\mathbf{r}) \phi_{n\mathbf{k}}^*(\mathbf{r}') \phi_{n'\mathbf{k}'}(\mathbf{r}') \times \\ & \left[\frac{f_{n\mathbf{k}}(1 - f_{n'\mathbf{k}'})}{\epsilon_{n\mathbf{k}}^0 - \epsilon_{n'\mathbf{k}'}^0 + \omega + i\eta} - \frac{(1 - f_{n\mathbf{k}})f_{n'\mathbf{k}'}}{\epsilon_{n'\mathbf{k}'}^0 - \epsilon_{n\mathbf{k}}^0 + \omega - i\eta} \right], \end{aligned} \quad (6.73)$$

where $f_{n\mathbf{k}}$ are the Fermi factors.

Usually one is more interested in the quasi-particle corrections to the bandstructures. To obtain the eigenvalues ϵ_i it is more efficient to use the

6.5. PRACTICAL G_0W_0 CALCULATIONS

quasi-particle equation 6.30, instead of solving the Dyson equation 6.33 and searching for the poles of the Green's function. In systems where the quasi-particle wavefunctions are known, one finds that the Kohn-Sham wavefunctions ϕ^{KS} are very similar to the quasi-particle ones. If one uses a standard DFT-like local approximation to the self energy

$$\Sigma(\mathbf{r}, \mathbf{r}'; \omega) = \delta(\mathbf{r} - \mathbf{r}')V_{xc}(\mathbf{r}), \quad (6.74)$$

the quasi-particle equation 6.30, becomes a Kohn-Sham equation

$$(\hat{h} + V_{xc})\psi_i^{KS} = \epsilon_i^{KS}\psi_i^{KS}. \quad (6.75)$$

Both observations indicate that the self-energy correction

$$\Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i) - V_{xc}\delta(\mathbf{r} - \mathbf{r}') \quad (6.76)$$

is small. Therefore one can use first order perturbation theory to obtain better approximations to the eigenenergies

$$\epsilon_i \approx \epsilon_i^{KS} + \langle \psi_i^{KS} | \Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i) - V_{xc} | \psi_i^{KS} \rangle. \quad (6.77)$$

This equation indicates that one does not need to calculate the complete Σ matrix, as one element is sufficient to calculate one corrected energy. This is however a nonlinear equation, as Σ depends on the quasi-particle energy. This can be simplified by using a linear expansion of Σ around the Kohn-Sham eigenenergy ϵ_i^{KS}

$$\Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i) \approx \Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i^{KS}) + (\epsilon_i - \epsilon_i^{KS}) \left. \frac{\partial \Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i^{KS})}{\partial \epsilon_i} \right|_{\epsilon_i^{KS}}. \quad (6.78)$$

Inserting this in equation 6.77 gives

$$\epsilon_i \approx \epsilon_i^{KS} + Z_i \langle \psi_i^{KS} | \Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i^{KS}) - V_{xc} | \psi_i^{KS} \rangle, \quad (6.79)$$

where the quasi-particle renormalization factor (also called the quasi-particle weight) Z_i is given by

$$Z_i = \left(1 - \langle \psi_i^{KS} | \left. \frac{\partial \Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i^{KS})}{\partial \epsilon_i} \right|_{\epsilon_i^{KS}} | \psi_i^{KS} \rangle \right)^{-1}. \quad (6.80)$$

CHAPTER 6. GW APPROXIMATION: THEORY

In practice all calculations are carried out in momentum space, so the formulations need to be transformed by applying Fourier transforms. The Fourier transform for the two variable function G^0 is defined as

$$G^0(\mathbf{r}, \mathbf{r}', \omega) \rightarrow G_{\mathbf{G}\mathbf{G}'}^0(\mathbf{q}, \mathbf{q}', \omega) = \frac{1}{\Omega} \int d^3r d^3r' e^{-i(\mathbf{G}+\mathbf{q})\cdot\mathbf{r}} G^0(\mathbf{r}, \mathbf{r}', \omega) e^{-i(\mathbf{G}'+\mathbf{q}')\cdot\mathbf{r}'}, \quad (6.81)$$

where \mathbf{G} denotes a reciprocal lattice vector and \mathbf{q} wave vectors in the IBZ. Applying this to the real space expression for the polarizability χ^0 in equation 6.73 gives

$$\begin{aligned} \chi_{\mathbf{G}\mathbf{G}'}^0(\mathbf{q}, \omega) &= \frac{2}{\Omega} \sum_{nn'\mathbf{k}} \langle n\mathbf{k} | e^{-i(\mathbf{G}+\mathbf{q})\cdot\mathbf{r}} | n'\mathbf{k} + \mathbf{q} \rangle \langle n'\mathbf{k} + \mathbf{q} | e^{-i(\mathbf{G}'+\mathbf{q}')\cdot\mathbf{r}'} | n\mathbf{k} \rangle \\ &\quad \times \left[\frac{f_{n\mathbf{k}}(1 - f_{n'\mathbf{k}'})}{\epsilon_{n\mathbf{k}}^0 - \epsilon_{n'\mathbf{k}'}^0 + \omega + i\eta} - \frac{(1 - f_{n\mathbf{k}})f_{n'\mathbf{k}'}}{\epsilon_{n'\mathbf{k}'}^0 - \epsilon_{n\mathbf{k}}^0 + \omega - i\eta} \right]. \end{aligned} \quad (6.82)$$

Similarly one obtains, using equations 6.53 and 6.50,

$$\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = \delta_{\mathbf{G}\mathbf{G}'} - \frac{4\pi}{|\mathbf{G} + \mathbf{q}|^2} \chi_{\mathbf{G}\mathbf{G}'}^0(\mathbf{q}, \omega) \quad (6.83)$$

$$W_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = \frac{4\pi}{|\mathbf{G} + \mathbf{q}|^2} \epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}, \omega). \quad (6.84)$$

In order to obtain the self energy at a given frequency, the convolution in equation 6.70 needs to be calculated. Therefore one needs to calculate ϵ^{-1} for different \mathbf{q} -points and for all frequencies. This is computational very time consuming, especially because the function $\chi^0(\omega)$ has many poles close to the real axis. In order to avoid this burden, a plasmon-pole model will be used in order to model the inverse dielectric function. The idea behind this approach is that this function is dominated by the peaks at which $\epsilon = 0$, i.e. the plasmon peaks. An additional benefit of using such a plasmon-pole model is that it allows to perform the convolution analytically.

The model used in this thesis is the Godby-Needs plasmon-pole model, which models ϵ^{-1} as follows

$$\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}, \omega) = \delta_{\mathbf{G}\mathbf{G}'} + \frac{\Omega_{\mathbf{G}\mathbf{G}'}^2(\mathbf{q})}{\omega^2 - (\tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) - i\eta)^2}. \quad (6.85)$$

The two parameters $\Omega_{\mathbf{G}\mathbf{G}'}^2$ and $\tilde{\omega}_{\mathbf{G}\mathbf{G}'}$ are determined in ABINIT by performing the RPA calculation at two frequencies: at $\omega = 0$ and at $\omega \sim i\omega_p$, where

6.5. PRACTICAL G_0W_0 CALCULATIONS

ω_p is the classical plasmon frequency. These frequencies are convenient as these are far away from the poles of χ^0 , so that the fit is more stable. Alternative choices for determining the fitting parameters exist [112] as well as alternate plasmon-pole models (e.g. Ref. [113]).

If W is split in the bare Coulomb interaction v and the remainder $W - v$, the self-energy can be split in an exchange (Σ^x) and a correlation (Σ^c) part. Symbolically this is

$$\Sigma = iG_0^{KS}W = iG_0^{KS}v + iG_0^{KS}(W - v) = \Sigma^x + \Sigma^c. \quad (6.86)$$

The exchange part of the self-energy reads

$$\Sigma^x(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int d\omega' G_0^{KS}(\mathbf{r}, \mathbf{r}', \omega + \omega') v(\mathbf{r}, \mathbf{r}') e^{i\omega'\eta}, \quad (6.87)$$

which can be evaluated analytically. This leads to the Hartree-Fock exchange term

$$\langle \psi_{n\mathbf{k}} | \Sigma^x | \psi_{n\mathbf{k}} \rangle = - \sum_{m\mathbf{q}}^{occ.} \int d^3r d^3r' \frac{\psi_{n\mathbf{k}}^*(\mathbf{r}) \psi_{m\mathbf{q}}(\mathbf{r}) \psi_{m\mathbf{q}}^*(\mathbf{r}') \psi_{n\mathbf{k}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (6.88)$$

Similarly the correlation part of the self-energy is given by

$$\Sigma^c(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int d\omega' G_0^{KS}(\mathbf{r}, \mathbf{r}', \omega + \omega') [W(\mathbf{r}, \mathbf{r}', \omega) - v(\mathbf{r}, \mathbf{r}')] e^{i\omega'\eta}. \quad (6.89)$$

In order to give an expression for the matrix elements of these self-energy parts, it is convenient to define the matrix elements $M_{\mathbf{G}}^{vc}(\mathbf{k}, \mathbf{q})$ between the occupied valence ($n = v$) and the unoccupied conduction ($n = c$) states

$$\begin{aligned} M_{\mathbf{G}}^{vc}(\mathbf{k}, \mathbf{q}) &= \int d^3r \psi_{v\mathbf{k}-\mathbf{q}}^*(\mathbf{r}) e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} \psi_{c\mathbf{k}} \\ &= \sum_{\mathbf{G}'} c_{v\mathbf{k}-\mathbf{q}}^*(\mathbf{G}' - \mathbf{G}) c_{c\mathbf{k}}(\mathbf{G}'). \end{aligned} \quad (6.90)$$

Using this notation the matrix elements of the exchange part can be written in momentum space as

$$\langle \psi_{n\mathbf{k}} | \Sigma^x | \psi_{m\mathbf{k}} \rangle = - \frac{4\pi}{\Omega} \sum_l^{occ.} \sum_{\mathbf{q}} \sum_{\mathbf{G}} \frac{M_{\mathbf{G}}^{ln}(\mathbf{k}, \mathbf{q}) [M_{\mathbf{G}}^{ln}(\mathbf{k}, \mathbf{q})]^*}{|\mathbf{G} + \mathbf{q}|^2}. \quad (6.91)$$

The correlation part, using the Godby-Needs approximation is given by

$$\begin{aligned} \langle \psi_{n\mathbf{k}} | \Sigma^c | \psi_{m\mathbf{k}} \rangle &= -\frac{2\pi}{\Omega} \sum_{l\mathbf{q}} \sum_{\mathbf{G}\mathbf{G}'} \frac{M_{\mathbf{G}}^{ln}(\mathbf{k}, \mathbf{q}) [M_{\mathbf{G}}^{ln}(\mathbf{k}, \mathbf{q})]^*}{|\mathbf{G} + \mathbf{q}| |\mathbf{G}' + \mathbf{q}|} \\ &\times \frac{\Omega_{\mathbf{G}\mathbf{G}'}^2(\mathbf{q})}{\omega_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) (\omega - \epsilon_{l\mathbf{k}-\mathbf{q}} + \text{sgn}(\mu - \epsilon_{l\mathbf{k}-\mathbf{q}}) \omega_{\mathbf{G}\mathbf{G}'}(\mathbf{q}))}. \end{aligned} \quad (6.92)$$

The *sgn* function in combination with the chemical potential μ is 1 when considering a valence band and -1 for a conduction band. With these self-energy matrix elements calculated, one can calculate the GW corrections to the eigenvalues by using equation 6.79.

7

Full quasi-particle corrected band structures of Si and Ge nanowires

7.1 Introduction

As explained in the previous chapter, one encounters the bandgap problem when doing a standard DFT calculation to calculate the bandstructure of a material. A solution to correct the underestimation of the bandgap is to use the GW approximation. Such an approach is used in this chapter, which is the result of a collaboration with the group of X. Gonze of the Université Catholique de Louvain (UCL).

As the one dimensional nature of the nanowires gives rise to some convergence issues, first the convergence in case of the 0.5 nm Ge nanowire (consisting of 6 Ge and 8 passivating H atoms) is discussed. Several convergence problems arise, which can be prevented by utilizing several recent techniques. These are discussed, as well as applied to the test system.

In the second part the obtained GW corrections are used to obtain the complete band spectrum. This can be achieved by performing an interpolation using Wannier functions. The obtained band spectra are compared with those from DFT using the local-density approximation (LDA) and the usability of a scissor shift operator is discussed. Also the electron and hole masses are calculated and compared, both in LDA and in the GW corrected approach. As the quasi-particle correction requires a starting point, we start with our converged results using the LDA as exchange-correlation functional.

To keep this chapter conveniently readable, most graphs of the convergence studies are displayed in the appendices. Only those graphs of direct importance are repeated in the text. To obtain the values of the convergence

parameters, not only these values are taken into account, but also the rate of their change, i.e. the slope in the convergence graphs.

As the GW calculation uses the results obtained within LDA as a starting point, first the LDA convergence is studied. Such a convergence study was also performed in the previous chapters, but here we explicitly show it, so that we can compare it with the GW convergence studies.

7.2 LDA convergence

For the LDA calculation three main parameters need a convergence study. The separation between the nanowire and its mirror images (caused by applying periodic boundary conditions), which is represented by the distance in the x- and y-direction (the wire is considered to be oriented along the z-direction), is the first parameter. As the wire possesses a spherical symmetry, this distance is taken equal in the x and y direction. The number of basis functions (here plane waves), determined by a cutoff energy (*ecut*), and the number of \mathbf{k} -points used to perform the integrations over the Brillouin zone are the two other parameters. For the \mathbf{k} -point grid a shifted Monkhorst-Pack grid [52] is chosen. For the integration in x- and y-direction a single \mathbf{k} -point is sufficient, which is a consequence of the inverse proportionality of the Brillouin zone with respect to the real space. Therefore the number of \mathbf{k} -points in the z-direction (in ABINIT given by the parameter *ngkpt(3)*) is used as the convergence parameter. Because time reversal symmetry can be used and the fact that we only need to take into account the irreducible Brillouin zone, this number has to be divided by two to obtain the actual used number of \mathbf{k} -points.

The main quantity that has to converge is the energy difference between the highest valence band and the lowest conduction band, e.g. the bandgap. Thus, for the 0.5 nm nanowire the bandgap is the distance between the 16th and the 17th band. For completeness it is also checked if the unit cell dimension in the z-direction (*acell(3)*) and the total energy are converged.

For each set of parameters a full relaxation is performed, so that all forces are smaller than $5 \cdot 10^{-5}$ Hartree/Bohr and all components of the stress tensor are smaller than $5 \cdot 10^{-7}$ Hartree/Bohr³.

From Fig. 7.1, which shows the LDA bandgap as a function of the three parameters, one can conclude that a unit cell size of 23 Bohr in x- and y-direction, a plane wave energy cutoff of 18 Hartree, and a sampling of 2

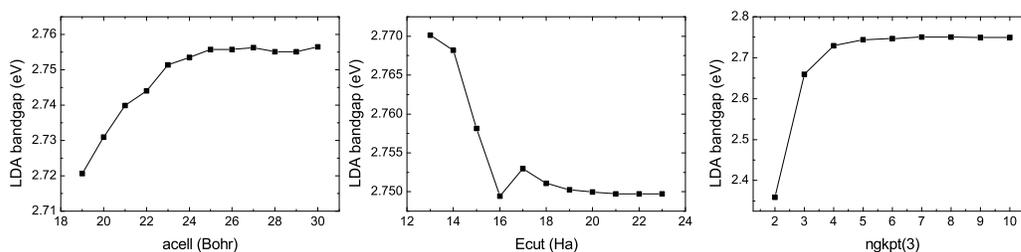


Figure 7.1: Convergence of the bandgap, as function of the unit cell size in directions perpendicular to the wire (*acell*), the plane wave energy cutoff (*ecut*), and the number of \mathbf{k} -points in the z -direction (*ngkpt(3)*).

symmetry inequivalent \mathbf{k} -points is sufficient to have converged results within a 0.01 eV energy range. The other convergence indicators (total energy and unit cell size in the z -direction) are also converged with these values (see appendix A-1). Therefore these values are used as a starting point for the quasi-particle corrections. The converged value of the LDA bandgap is equal to 2.76 eV.

7.3 GW convergence

Using these LDA results, the GW corrections can be calculated. To obtain convergence, more parameters should be taken into account. A model GW calculation consists of three steps: first the calculation of the LDA wavefunctions and eigenvalues (the generation of a KSS structure file), second the calculation of the screened dielectric matrix (calculation of the ϵ^{-1} , also called the screening calculation) and finally the calculation of the self energy and the corrections of the bandenergies.

7.3.1 Convergence parameters

KSS generation

For the first step, the generation of the KSS file, three parameters are of main importance: the number of calculated bands (it is necessary to include a sufficiently large number of empty bands), characterized in ABINIT by the parameter *nband*, the size of the unit cell (in x - and y -directions), and finally the number of used \mathbf{k} -points.

The initial value of the number of bands can be taken large, as in the subsequent calculations a smaller number can always be requested. Therefore only one KSS file should be generated, containing a large number of bands. The typical rule of thumb is to include 10-20 times the number of occupied bands. In this study this resulted in 160-320 bands.

For the unit cell dimensions or the number of \mathbf{k} -points on the other hand, a separate KSS file is needed for each new combination of values. Also, since the wire has a direct bandgap located at the gamma point, the gamma point (Γ) needs to be included in the \mathbf{k} -point mesh. Therefore an unshifted mesh is used, which is slightly less efficient in sampling the Brillouin zone.

Screening calculation

For the calculation of the dielectric matrix, several quantities contain summations, as is shown in the previous chapter. These summations are cutoff at certain values, identified by several parameters. Here I briefly discuss these parameters, which are present in ABINIT.

The first of these parameters is *ecuteps*, which determines the cutoff energy of the \mathbf{G} vectors used to represent the independent-particle susceptibility $\chi^{(0)}$ (equation 6.82), the dielectric matrix ϵ (equation 6.83), and its inverse (used in equation 6.83). Since this parameter determines the dimensions of ϵ and W , it has a large influence on the required memory and disk space (to store the dielectric matrix), so a careful convergence study is needed.

The next cutoff parameter is *ecutwfn*, which determines the cutoff energy of the plane wave set used to represent the wavefunctions in the formula that generates the independent-particle susceptibility χ^0 (equation 6.82)). This parameter mainly influences the required memory, while the required cpu time is not much affected.

As said previously, the number of bands (summation in the formula of χ^0 (equation 6.82)), the unit cell size, and the number of \mathbf{k} -points are also important and influence the disk space, memory requirements and cpu time.

For the calculation of the dielectric matrix one also needs to calculate the independent particle polarizability for all frequencies. By using a plasmon-pole model, this cumbersome calculation can be avoided. Here we used the plasmon-pole model of Godby and Needs. In this approach the screening must be calculated at zero frequency and for another imaginary frequency, usually taken of the order of the plasmon frequency. Here a value of 16.7 eV is used. The final results should not depend on this value.

Sigma calculation

For the sigma calculation, convergence parameters similar to the ones of the screening calculation are relevant: *ecutwfn* (cutoff on the \mathbf{G} -vectors used in the summation of the matrix elements $M_{\mathbf{G}}^{ln}$ equation 6.90), number of bands (e.g. in equations 6.91 and 6.92), number of \mathbf{k} -points, and the unit cell dimensions.

A new cutoff parameter is introduced, called *ecutsigx*, which determines the cutoff energy of \mathbf{G} -vectors used to generate the exchange part of the self-energy operator Σ^x (equation 6.91).

7.3.2 Convergence study

For these convergence studies, one parameter is systematically varied, while keeping the others fixed at reasonable values.

Cutoff parameters

As the calculation of the screening is the most time consuming part, first the convergence study of *ecuteps* is done, as this largely determines the time, memory, and disk space needed for the screening calculation. In Fig. 7.2(a), the GW corrected bandgap is shown as function of *ecuteps*. It can be observed that a value of 3.5 Ha already leads to a converged value. From the complete graphs in the appendix (Fig. A-2), one also sees that this cutoff influences only the correlation part of the self energy.

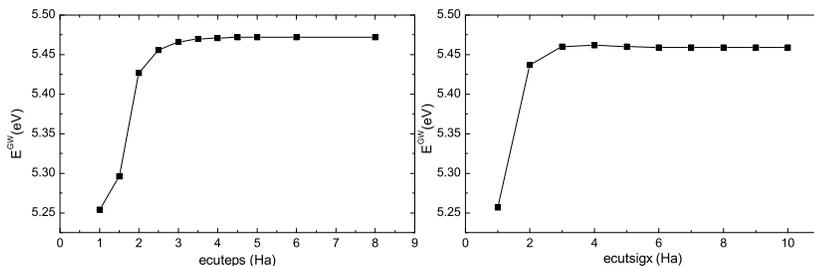


Figure 7.2: Convergence of the GW gap as function of (a) *ecuteps* and (b) *ecutsigx*

For the the exchange part of the sigma calculation, the main cutoff parameter is *ecutsigx*. In Fig. 7.2(b) this parameter is varied and the corrected

CHAPTER 7. FULL QUASI-PARTICLE CORRECTED BAND STRUCTURES OF SI AND GE NANOWIRES

bandgap is calculated for each value. A value of 4 Ha can be considered as converged (the slope is also almost zero). This parameter influences both the exchange and correlation part of the self energy (as can be seen from Fig. A-3).

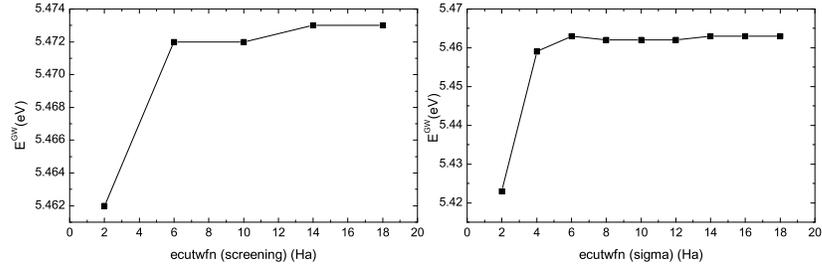


Figure 7.3: Convergence of the GW gap as function of (a) $ecutwfn$ set during the screening calculations and (b) $ecutwfn$ set during the sigma calculations

In both the screening and sigma calculation the parameter $ecutwfn$ plays a role. From Fig 7.3 (and Figs. A-4 and A-5 in the appendices), it can be seen that the value for the screening part can be lower (6 Ha is already sufficient) than that for the sigma part (8 Ha). But since this parameter only has a small influence on the required cpu time, it only affects the memory requirements, the maximum value is used if the available memory (less than 16 Gb) permits it.

KSS parameters

In this part the convergence parameters are discussed that require a separate calculation of the KSS file for each parameter (e.g. the unit cell size, the number of bands, and the \mathbf{k} -point grid used). Some of these convergences lead to numerical problems, which are discussed, as well as possible solutions for these problems.

The first parameter we studied is the dependency on the number of \mathbf{k} -points. As can be seen from Fig. 7.4(a) (black curve), the value of the GW corrected bandgap increases with increasing number of \mathbf{k} -points and there is no sign of convergence (the curve shows a linear slope). A similar behavior is seen when the unit cell is increased in the lateral direction (not shown). This lack of convergence can be explained by the presence of a long range Coulomb interaction. This causes interactions between the wire images, and due to its

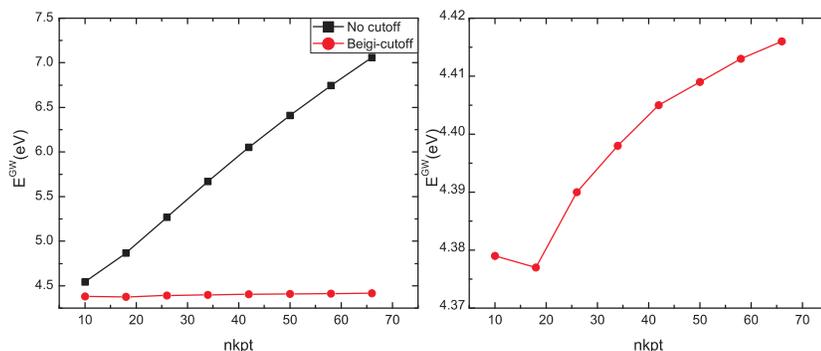


Figure 7.4: Convergence of the bandgap with respect to the number of \mathbf{k} -points. (a) With and without Beigi cutoff, (b) zoom of the convergence with Beigi cutoff.

long range, it requires huge unit cells (which cannot be calculated with the present computer resources) to reach the point that this Coulomb interaction between the image cells can be neglected. The reason that this Coulomb interaction is so dominantly present in GW calculations (in contrast with the LDA calculations, where a small unit cell and \mathbf{k} -point grid is sufficient to converge the band gap) is that the quasi-particle approach introduces charges, because the Green functions are linked to electron addition and removal, and this is not screened completely by the passivating hydrogen layer. In particular the problem occurs for the $\mathbf{q} \rightarrow 0$ limit.

A possible way around this problem is to introduce a truncation of the Coulomb interaction after a certain radius. This truncation is done most efficiently in Fourier space. Two available techniques (one by Ismael-Beigi [114] and the other by Rozzi *et al.* [115]) are present in the ABINIT program. Here the Beigi approach is used. In this approach the Coulomb term v in the equations for W (equation 6.84) and ϵ (equation 6.83) is replaced by a truncated interaction v_c , thus

$$\begin{aligned} W_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) &= \epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q})v_c(\mathbf{q} + \mathbf{G}') \\ \epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) &= \delta_{\mathbf{G}\mathbf{G}'} - v_c(\mathbf{q} + \mathbf{G}')\chi_{\mathbf{G}\mathbf{G}'}^0(\mathbf{q}). \end{aligned} \quad (7.1)$$

The truncated Coulomb interaction for a wire has the form

$$v_c(\mathbf{r}) = \frac{\theta(x, y)}{|\mathbf{r}|}, \quad (7.2)$$

CHAPTER 7. FULL QUASI-PARTICLE CORRECTED BAND STRUCTURES OF SI AND GE NANOWIRES

where $\theta(x, y)$ is zero unless x and y are inside the Wigner-Seitz unit cell. The Fourier transform of this quantity is

$$v_c(\mathbf{k}) = \int dx dy \theta(x, y) 2K_0(|k_z|\rho) \cos(k_x x + k_y y), \quad (7.3)$$

where $\rho = (x^2 + y^2)^{1/2}$ and $K_0(z)$ is the modified Bessel function. This integral is of finite extent, so when $k_z \rightarrow 0$, the divergence originates from K_0 , so that the divergent term reads

$$-2 \ln(|k_z|) \int dx dy \theta(x, y) \cos(k_x x + k_y y). \quad (7.4)$$

When $(k_x^2 + k_y^2)^{1/2} \neq 0$, this divergence vanishes, as the projection of \mathbf{k} in the xy plane is a reciprocal lattice vector. Therefore the cutoff radius depends only on the used unit cell, so no additional convergence parameters are needed.

The effect of the Beigi cutoff technique on the convergence can be clearly seen in Fig. 7.4(a) (red curve). A detailed view of the convergence graphs in the case of the Beigi approach is shown in Fig. 7.4(b). As can be seen, convergence is much better.

For the convergence with respect to the number of bands a technique (the extrapolar treatment [116]) exists that allows to decrease the number of used unoccupied bands (10 times smaller values are possible) during a GW calculation. This decrease of the amount of bands can be important, as this decreases the amount of memory that is needed to perform the calculation. To be exact, a GW calculation should take into account a number of states of the same order of the dimensionality of the Hilbert space, so equal to the number of basis functions. In a plane wave approach this number of basis functions is huge. To improve this convergence, the eigenenergies of states not explicitly treated are replaced by a common energy, determined from the highest computed state through a single parameter. These states are taken into account by utilizing the closure relation

$$\sum_{i > N_b} |i\rangle \langle i| = \mathbf{1} - \sum_{i \leq N_b} |i\rangle \langle i|. \quad (7.5)$$

The introduced parameter, called *gwencomp*, requires an additional convergence study. Note that it is also possible to obtain a calculated value by

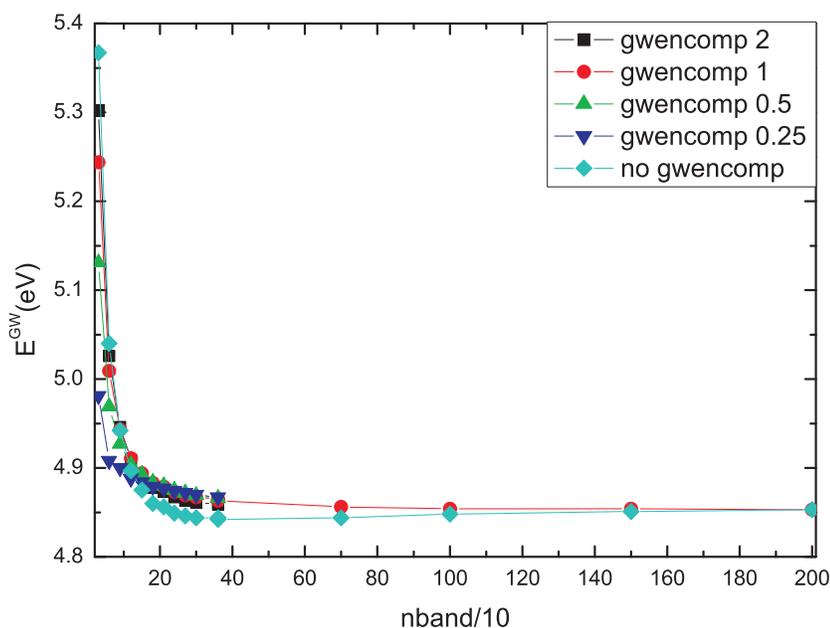


Figure 7.5: Convergence of the bandgap with respect to the number of bands in the calculation. The convergence is shown for different values of the parameter *gwencomp*.

checking the fulfilment of a sum rule (see Ref. [116] for details). This convergence is shown in Fig. 7.5. In this figure, the same number of bands and the same value of *gwencomp* is taken for the screening and the subsequent sigma calculation. When the extrapolar technique is used, we find that the bandgap is converged when using 180 bands. Such a convergence is not obtained when the extrapolar technique is not used. For small number of bands, it seems that the calculation using the extrapolar technique is converging towards another value for the bandgap, compared to the calculations without using the technique. However, when the number of bands is increased further, both approaches converge towards the same value. It is important to notice that the usage of the extrapolar technique leads to a much faster convergence (180 bands are sufficient versus 2000 bands). This influences both memory requirements and cpu time substantially. In this case the best value to take for *gwencomp* is 1 eV, as this leads to the fastest convergence, with the smallest slope in the convergence graphs. From Fig. A-6 it follows that the convergence of the bandgap is faster than the convergence of the individ-

CHAPTER 7. FULL QUASI-PARTICLE CORRECTED BAND STRUCTURES OF SI AND GE NANOWIRES

ual parts (band energies of the highest valence E_{GW}^{16} and lowest conduction band E_{GW}^{17}). The usage of this technique is not limited to one-dimensional (or more general low dimensional systems), but it can also be used in bulk systems.

Finally, the convergence with respect to the inter wire distance and thus of the lateral size of the unit cell, characterized by the parameter $acell$, is performed. Here we used the previously discussed extrapolar technique. This convergence for the value of the bandgap, in the presence (red dots) and absence (black squares) of the Beigi cutoff, is shown in Fig. 7.6. Convergence in the case where we also applied the Beigi cutoff is already obtained for a lateral unit cell size of 26. Please note the units on the y-axis, as the differences are of the order 0.001 eV, which is smaller than the desired accuracy.

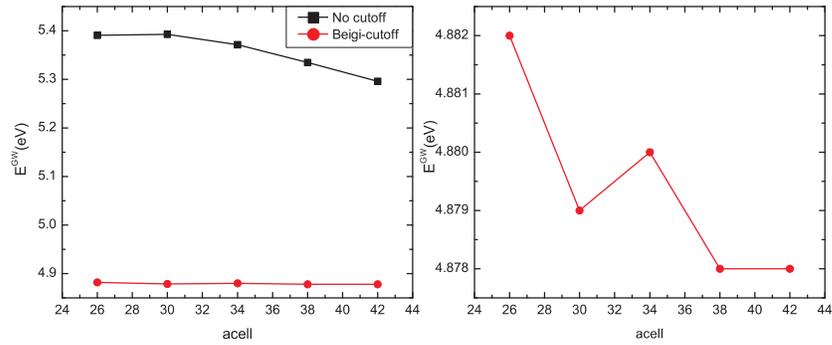


Figure 7.6: Convergence of the bandgap with respect to the size of the unit cell in the directions perpendicular to the wire (parameter $acell$). (a) In the absence and presence of the Beigi cutoff, and (b) a zoom of the convergence when the Beigi cutoff is applied.

To conclude this part, we calculated the GW corrected band gap of 0.5 and 1.2 nm Si and Ge nanowires, and of the 1.6 nm Si nanowire. The obtained LDA values, the GW results and the size of the GW correction are shown in Table 7.1. These corrections are larger for the 0.5 nm nanowire than for the 1.2 and 1.6 nm nanowires. These values are comparable to values reported in literature, which are also added to Table 7.1. Experimental results for these wire sizes are not available, as the experimental wires have a larger diameter. For a comparison with larger diameter nanowires (and [112] oriented nanowires), see Fig. 8.2 in chapter 8.

7.4. COMPLETE BAND STRUCTURE

Wire diameter	LDA gap	GW gap	GW correction	Literature
Ge:				
0.5 nm	2.76	4.87	2.11	4.5 ¹
1.2 nm	1.57	2.98	1.41	
Si:				
0.5 nm	3.20	5.55	2.35	5 ²
1.2 nm	1.70	3.32	1.62	3.12 ³ -3.2 ² -3.4 ⁴
1.6 nm	1.14	2.31	1.18	2.2 ² -2.32 ³ -2.33 ⁴

Table 7.1: The LDA and GW band gap values for different wires, given in eV. The size of the GW correction is also given. The last column show reference values for the GW bandgap as present in the literature. ¹ denotes Ref. [68], ² Ref. [117], ³ Ref. [60], and ⁴ Ref. [118].

7.4 Complete band structure

When one wants to obtain the LDA band energies in an arbitrary \mathbf{k} -point, all one has to perform is a non-selfconsistent calculation in that \mathbf{k} -point. Calculation of the complete band structure can then be performed by doing these non-selfconsistent calculations in all \mathbf{k} -points along the desired \mathbf{k} -point path. For the GW corrected band energies it is not that simple, as all \mathbf{k} -points in which one wants to calculate the corrections should be in the initial \mathbf{k} -point grid used to do the screening calculations. Therefore large initial grids should be used, which increases the required computing resources. Additionally, if the \mathbf{k} -point is present in the grid, a time consuming sigma calculation has to be performed in this point for all bands (especially if one compares the time needed with that of a LDA non-selfconsistent calculation).

A solution is to use an efficient and accurate interpolation, so that based on a limited number of \mathbf{k} -points the complete bandstructure can be obtained. Interpolations based on Wannier functions are capable of doing this.

First a basic introduction to Wannier functions is given, followed by the practical application of them to obtain the GW corrected band structures. These are used to study the goodness of a scissor shift operation (commonly used to shift the conduction bands so that the band gap has a correct size) and the difference of the calculated band masses using LDA and using the GW corrected energy bands.

7.4.1 Wannier functions

Wannier functions are orthonormal localized functions that span the same space as the eigenstates of a band or a group of bands. They are defined in terms of the Bloch functions $\psi_{n\mathbf{k}}(\mathbf{r})$ as

$$|\mathbf{R}n\rangle = \frac{V}{(2\pi)^3} \int d\mathbf{k} e^{-i\mathbf{k}\mathbf{R}} |\psi_{n\mathbf{k}}\rangle, \quad (7.6)$$

so that

$$|\psi_{n\mathbf{k}}\rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}\mathbf{R}} |\mathbf{R}n\rangle, \quad (7.7)$$

where V denotes the volume of the real-space unit cell and \mathbf{R} a lattice vector. This set of Wannier functions forms an orthonormal set, as $\langle \mathbf{R}_1 n | \mathbf{R}_2 m \rangle = \delta_{\mathbf{R}_1 \mathbf{R}_2} \delta_{nm}$ and $\sum_{\mathbf{R}n} |\mathbf{R}n\rangle \langle \mathbf{R}n| = 1$, which can easily be shown making use of the orthonormality of the Bloch functions $\psi_{n\mathbf{k}}(\mathbf{r})$.

It is well known that Bloch wavefunctions are not unique. For a single isolated band (i.e. the band does not become degenerate with another band anywhere in the Brillouin zone) this leads to the freedom to choose the phases of the Bloch orbitals as a function of \mathbf{k}

$$|\psi_{n\mathbf{k}}\rangle \rightarrow e^{i\phi_n(\mathbf{k})} |\psi_{n\mathbf{k}}\rangle, \quad (7.8)$$

where ϕ_n is a real function of \mathbf{k} . In case of degenerate bands (which form a so-called “composite group”) this leads to the freedom to mix the bands at each \mathbf{k} -vector

$$|\psi_{n\mathbf{k}}\rangle \rightarrow \sum_m U_{mn}^{\mathbf{k}} |\psi_{m\mathbf{k}}\rangle, \quad (7.9)$$

where $U_{mn}^{\mathbf{k}}$ is a unitary matrix. The isolated band case can be seen as a special case of the composite group. This non-uniqueness allows to construct different Wannier functions. From this set of choices, one looks for those that are the most localized, which are called the maximally localized Wannier functions (MLWF). Marzari and Vanderbilt [119] defined the spread-functional Ω as

$$\Omega = \sum_n [\langle \mathbf{r}^2 \rangle_n - \langle \mathbf{r} \rangle_n^2], \quad (7.10)$$

7.4. COMPLETE BAND STRUCTURE

which should be minimized to obtain the set of MLWFs. For any operator A , $\langle A \rangle_n$ denotes the expectation value $\langle \mathbf{0}n | A | \mathbf{0}n \rangle$. Other localization criteria exist, but this one leads to an elegant formalism, as the spread-functional can be split in an invariant, diagonal and off-diagonal contribution:

$$\Omega = \underbrace{\sum_n \left[\langle r^2 \rangle_n - \sum_{\mathbf{R}m} |\langle \mathbf{R}m | \mathbf{r} | \mathbf{0}n \rangle|^2 \right]}_{\Omega_I} + \underbrace{\sum_n \sum_{\mathbf{R}m \neq \mathbf{0}n} |\langle \mathbf{R}m | \mathbf{r} | \mathbf{0}n \rangle|^2}_{\tilde{\Omega}} \quad (7.11)$$

The first term can be shown to be gauge invariant [119] and the last term can be split into a diagonal and an off-diagonal part

$$\tilde{\Omega} = \underbrace{\sum_{m \neq n} \sum_{\mathbf{R}} |\langle \mathbf{R}m | \mathbf{r} | \mathbf{0}n \rangle|^2}_{\Omega_{OD}} + \underbrace{\sum_n \sum_{\mathbf{R} \neq \mathbf{0}} |\langle \mathbf{R}n | \mathbf{r} | \mathbf{0}n \rangle|^2}_{\Omega_D}. \quad (7.12)$$

If we define the vectors \mathbf{b} as the vectors connecting each \mathbf{k} -point to its nearest neighbors and w_b the corresponding weights chosen so that

$$\sum_{\mathbf{b}} w_b b_\alpha b_\beta = \delta_{\alpha\beta} \quad (7.13)$$

is fulfilled and $M_{mn}^{\mathbf{k}\mathbf{b}}$ as

$$M_{mn}^{\mathbf{k}\mathbf{b}} = \langle u_{m\mathbf{k}} | u_{n, \mathbf{k}+\mathbf{b}} \rangle, \quad (7.14)$$

where $u_{n\mathbf{k}}$ is the cell periodic part of the Bloch functions, the different parts of the spread functional can be given by

$$\begin{aligned} \Omega_I &= \frac{1}{N} \sum_{\mathbf{k}\mathbf{b}} w_b \left(Tr \mathbf{1} - \sum_{mn} |M_{mn}^{\mathbf{k}\mathbf{b}}|^2 \right) \\ \Omega_{OD} &= \frac{1}{N} \sum_{\mathbf{k}\mathbf{b}} w_b \sum_{m \neq n} |M_{mn}^{\mathbf{k}\mathbf{b}}|^2 \\ \Omega_D &= \frac{1}{N} \sum_{\mathbf{k}\mathbf{b}} w_b \sum_n \left(-Im \ln M_{nn}^{\mathbf{k}\mathbf{b}} - \mathbf{b} \cdot \langle \mathbf{r} \rangle_n \right)^2. \end{aligned} \quad (7.16)$$

The trace $Tr \mathbf{1}$ equals the number of bands. These equations can then be minimized to find the Wannier functions with the smallest spread. Special care should be taken in case of entangled bands (see e.g. Ref. [120]).

7.4.2 GW corrected bands

We performed the minimization procedure explained in the previous subsection using the wannier90 code [121]. This code needs the overlap matrices M_{mn}^{kb} and an initial guess of the projection of the Bloch states $|\psi_{n\mathbf{k}}\rangle$ onto trial localized orbitals $|g_n\rangle$

$$A_{mn}^{(\mathbf{k})} = \langle \psi_{n\mathbf{k}} | g_n \rangle. \quad (7.17)$$

Usage of the minimization procedure yields the transformation matrix $U_{mn}^{\mathbf{k}}$. The Hamiltonian of the system can then be expressed in the basis of the MLWFs and can be subsequently used in a Slater-Koster interpolation [122] scheme to compute the Hamiltonian on a finer \mathbf{k} -point mesh.

Here we first search the optimal initial \mathbf{k} -point mesh needed to obtain good Wannier functions in the LDA case. For this mesh the GW corrections are calculated. These are then used to perform the interpolation of the band structures using the MLWFs. These are the same as in the LDA case, as the one-shot G_0W_0 approximation uses the LDA wavefunctions and only corrects the band energies.

In Fig. 7.7 the interpolated Wannier band structure for the LDA calculation of the 0.5 nm Ge nanowire is shown. These Wannier functions are obtained using a disentanglement procedure using an inner window from the bottom of the lowest valence band to a value of 5 eV and an outer window to 6 eV. In (a) the Wannier functions were obtained using an initial (unshifted) $1 \times 1 \times 5$ \mathbf{k} -point grid, corresponding to 3 symmetry inequivalent \mathbf{k} -points. The \mathbf{k} -point grid is unshifted, so that it contains the Γ point, needed to calculate the band gap in the GW calculations. The energy bands in these points are indicated by red dots. It can be clearly observed that the interpolation using the Wannier functions is not converged in this case. In (b) a finer \mathbf{k} -point grid was used, containing 6 inequivalent \mathbf{k} -points corresponding to an unshifted $1 \times 1 \times 10$ grid. The agreement with calculated non-selfconsistent band energies is excellent, as can be observed from (c), where the blue dots represent additionally obtained non-selfconsistently band energies.

If for the same \mathbf{k} -point mesh the GW corrected eigenvalues are used, the band structures can again be obtained by doing the Wannier based interpolation, which results in the band structure shown in Fig. 7.8(a) (black lines).

The obtained GW band structures can be used to study the validity of the commonly applied “scissor shift” operation. This operation aims to correct

7.4. COMPLETE BAND STRUCTURE

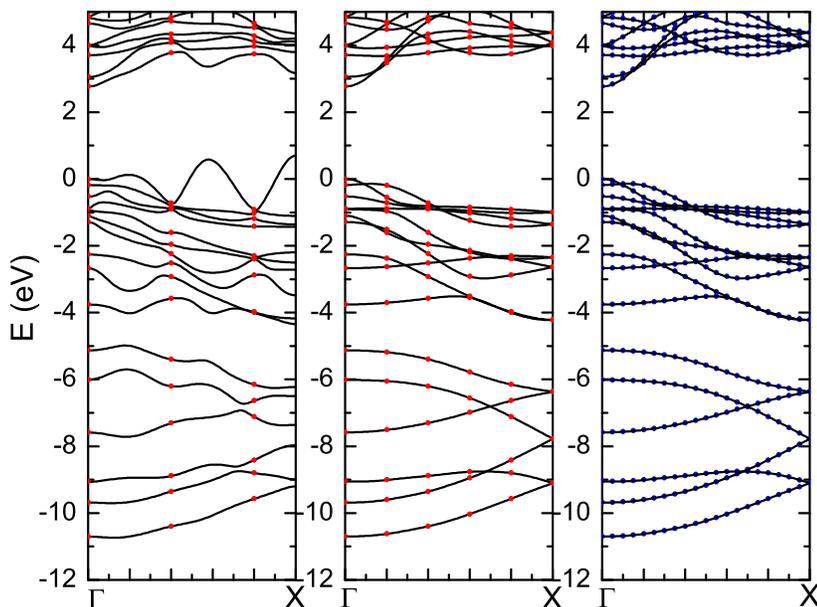


Figure 7.7: Wannier interpolated bandstructure (solid black lines) of the 0.5 nm Ge nanowire. The red dots are the calculated (using non-selfconsistent calculations) LDA values. (a) Interpolation obtained using an unshifted 1x1x5 grid. (b) Interpolation using an unshifted 1x1x10 grid. (c) Comparison of the interpolated band structure of (b) with additional calculated band energies (using non-selfconsistent iterations), shown as blue dots.

the LDA band structures by shifting all conduction bands so that the band gap gets the same value as the one calculated in a one-shot GW calculation. Thus this approach neglects all curvature differences between the LDA band structure and the GW corrected one. As we now have the full GW corrected band structure, we can test this approach. This is done in Fig. 7.8(a), where the red lines are the scissor shifted LDA energy bands. As can be seen from a zoom around the region close to the bandgap (which is shown in Fig. 7.8(b)), this approach is only valid for \mathbf{k} -points close to the Γ point and for bands close to the highest valence and lowest conduction band. One also observes that the gap between the highest valence and lowest conduction band in other \mathbf{k} -points, is still, even after applying the scissor shift, underestimated in the LDA case (the highest valence band is too high in energy, while the lowest conduction band is too low in energy). The GW correction is clearly not a

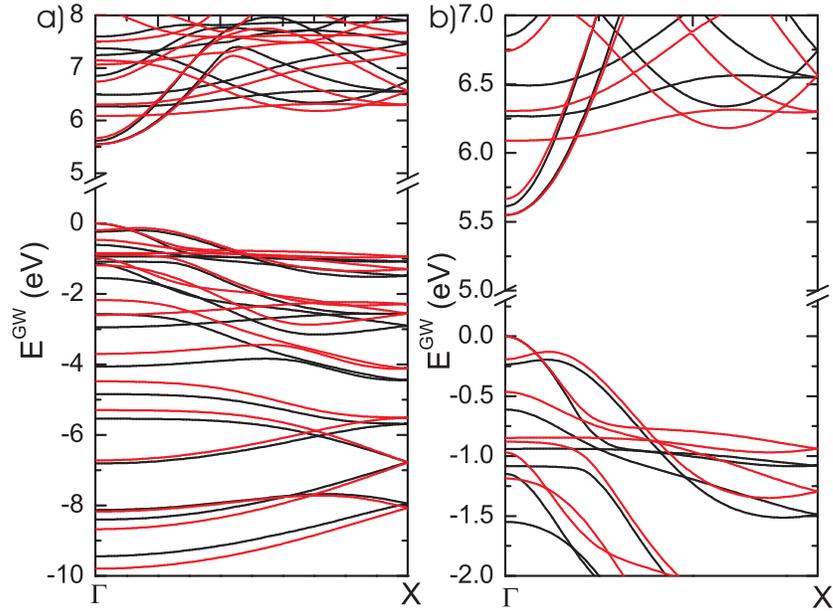


Figure 7.8: (a) Wannier interpolated GW corrected band structure (black) and the uncorrected scissor shifted LDA band structure (red) of a 0.5 nm Si nanowire. (b) A zoom of the region around the bandgap.

uniform correction, which is what one implies when performing a scissor shift operation. To check if the slopes near the Γ point, i.e. the hole and electron mass, are different in the GW corrected case as compared to the LDA case, these were calculated by fitting a quadratic curve for small \mathbf{k} -points. As could be expected from a visual inspection of Fig. 7.8 the band masses are equal in both cases.

7.5 Conclusions

In this chapter we have shown a systematic convergence study of the GW calculations for one-dimensional nanowires. It was shown that it is necessary to utilize a cutoff of the Coulomb interactions if one wants to achieve convergence. The application of the extrapolar technique allows to reduce the number of bands by a factor of 10. The rule of thumb to take an amount of empty bands equal to 10 times the number of occupied bands is also valid in case of nanowires. Using a Wannier interpolation we found good band

7.5. CONCLUSIONS

structures can be achieved. From these it was demonstrated that the scissor shift operation leads to substantial differences with the full GW results, when one compares the band energies in \mathbf{k} -points distant from the Γ point or for bands with energies that deviate from the highest valence or lowest conduction band. This can have measurable differences in the optical spectra. The validity close to Γ was also demonstrated by comparing the band masses for the GW corrected and uncorrected results, which did not show significant differences.

Appendix: Convergence graphs

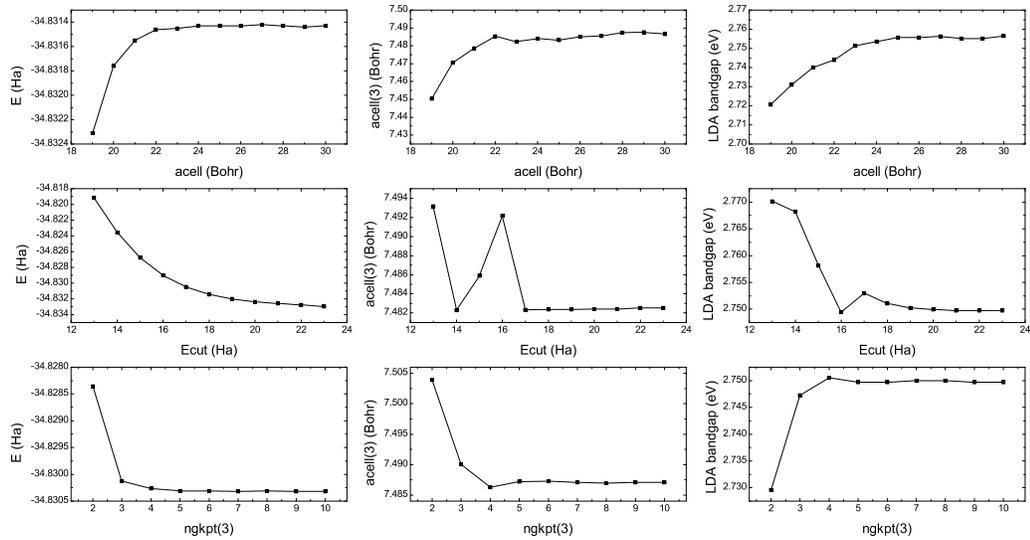


Figure A-1: Convergence of the LDA calculations as function of the transversal size of the unit cell ($acell$), the plane wave energy cutoff ($ecut$), and the number of k -points ($ngkpt(3)$). The convergence of the total energy, the unit cell size in the wire direction ($acell(3)$) and the LDA bandgap are shown.

Appendix: Convergence graphs

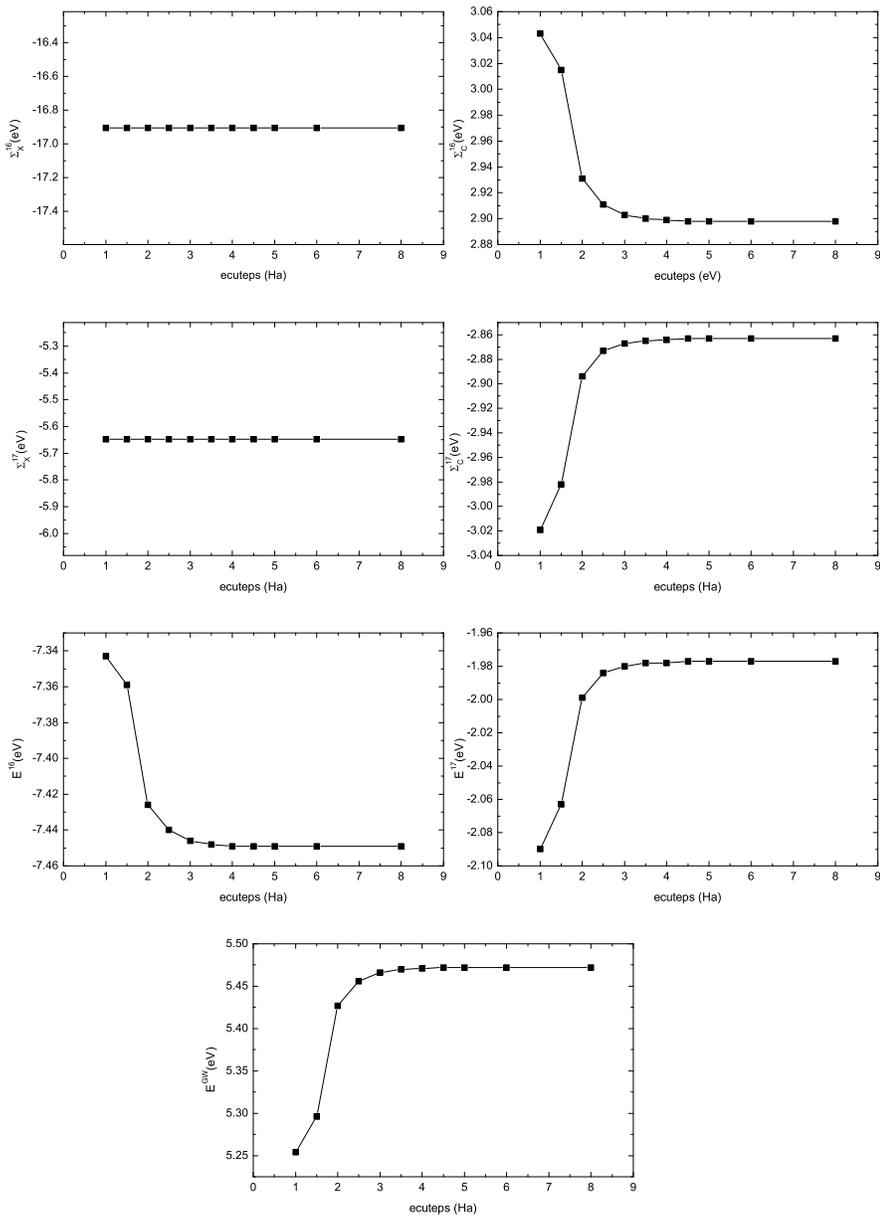


Figure A-2: Convergence of the GW calculations (shown are the correlation and exchange part of the self energy for the highest valence (16) and lowest conduction band (17), the energies of these bands and the GW bandgap) with respect to *cuteps*.

CHAPTER 7. FULL QUASI-PARTICLE CORRECTED BAND STRUCTURES OF SI AND GE NANOWIRES

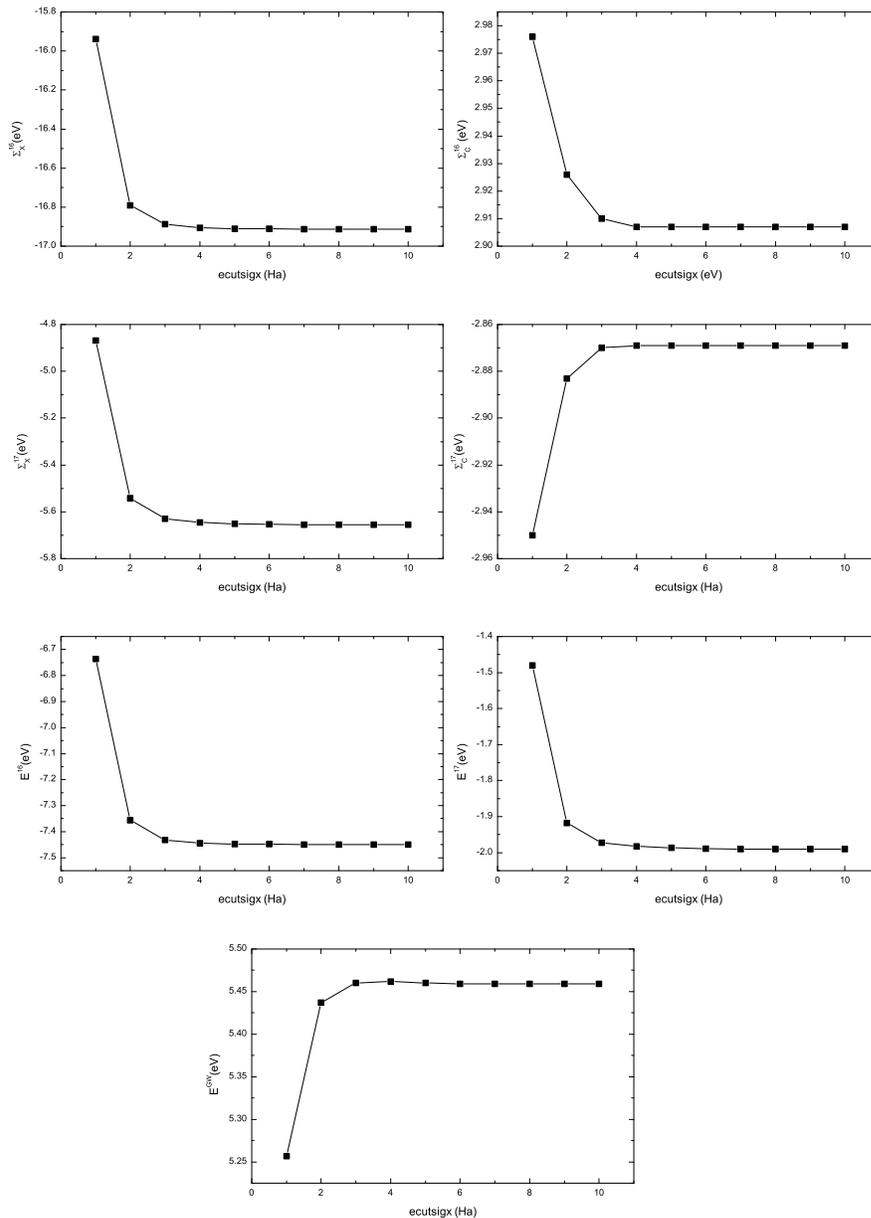


Figure A-3: Convergence of the GW calculations (shown are the correlation and exchange part of the self energy for the highest valence (16) and lowest conduction band (17), the energies of these bands and the GW bandgap) with respect to $ecutsigx$.

Appendix: Convergence graphs

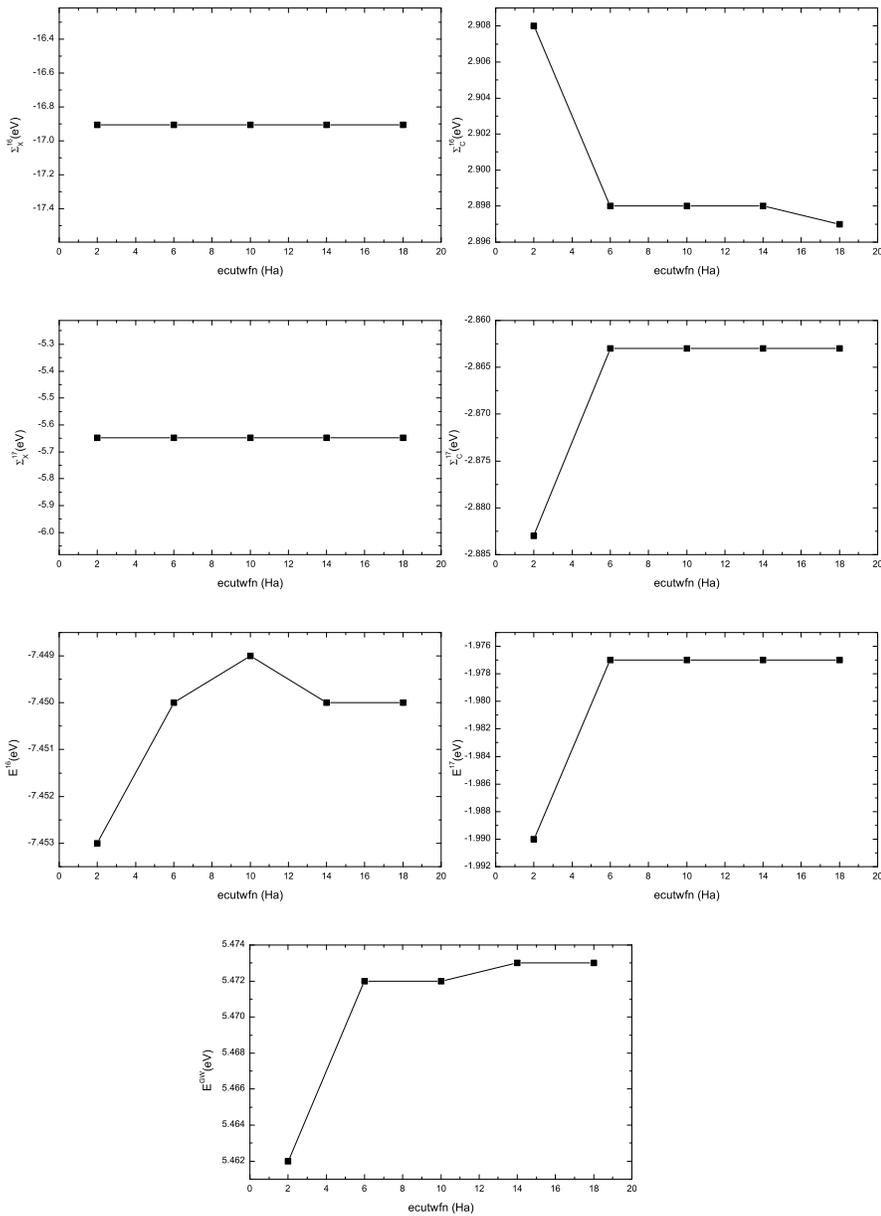


Figure A-4: Convergence of the GW calculations (shown are the correlation and exchange part of the self energy for the highest valence (16) and lowest conduction band (17), the energies of these bands and the GW bandgap) with respect to $ecutwfn$, set in the screening calculation.

CHAPTER 7. FULL QUASI-PARTICLE CORRECTED BAND STRUCTURES OF SI AND GE NANOWIRES

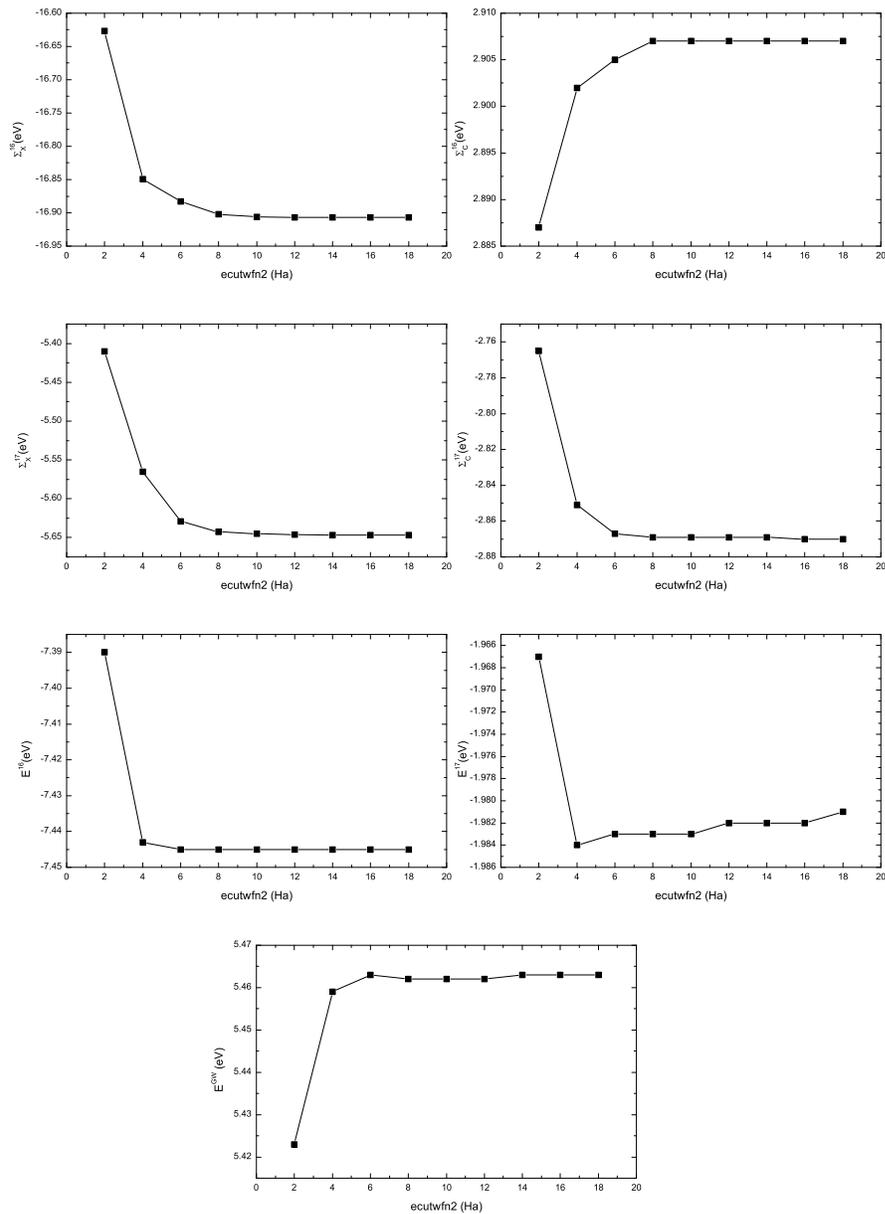


Figure A-5: Convergence of the GW calculations (shown are the correlation and exchange part of the self energy for the highest valence (16) and lowest conduction band (17), the energies of these bands and the GW bandgap) with respect to $cutwfn$, set in the sigma calculation.

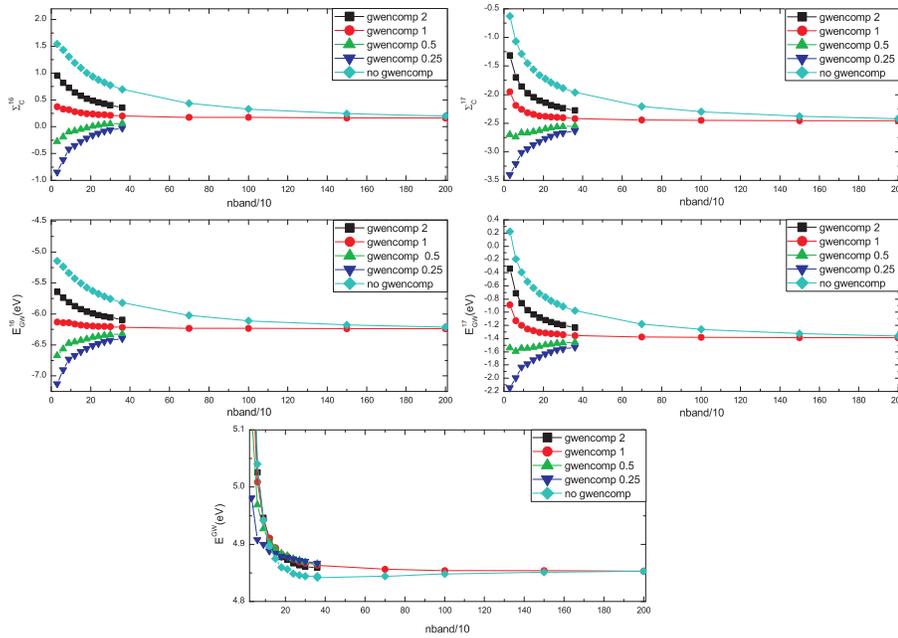


Figure A-6: Convergence of $nband$ (number of bands) using different values for the extrapolar correction (the parameter $gwencomp$). Shown are the correlation part of the self energy for the highest valence (16) and lowest conduction band (17), the energies of these bands and the GW bandgap. The exchange part of the self energy (Σ_X) is not shown, as this does not depend on the number of bands or the used $gwencomp$.

CHAPTER 7. FULL QUASI-PARTICLE CORRECTED BAND STRUCTURES OF SI AND GE NANOWIRES

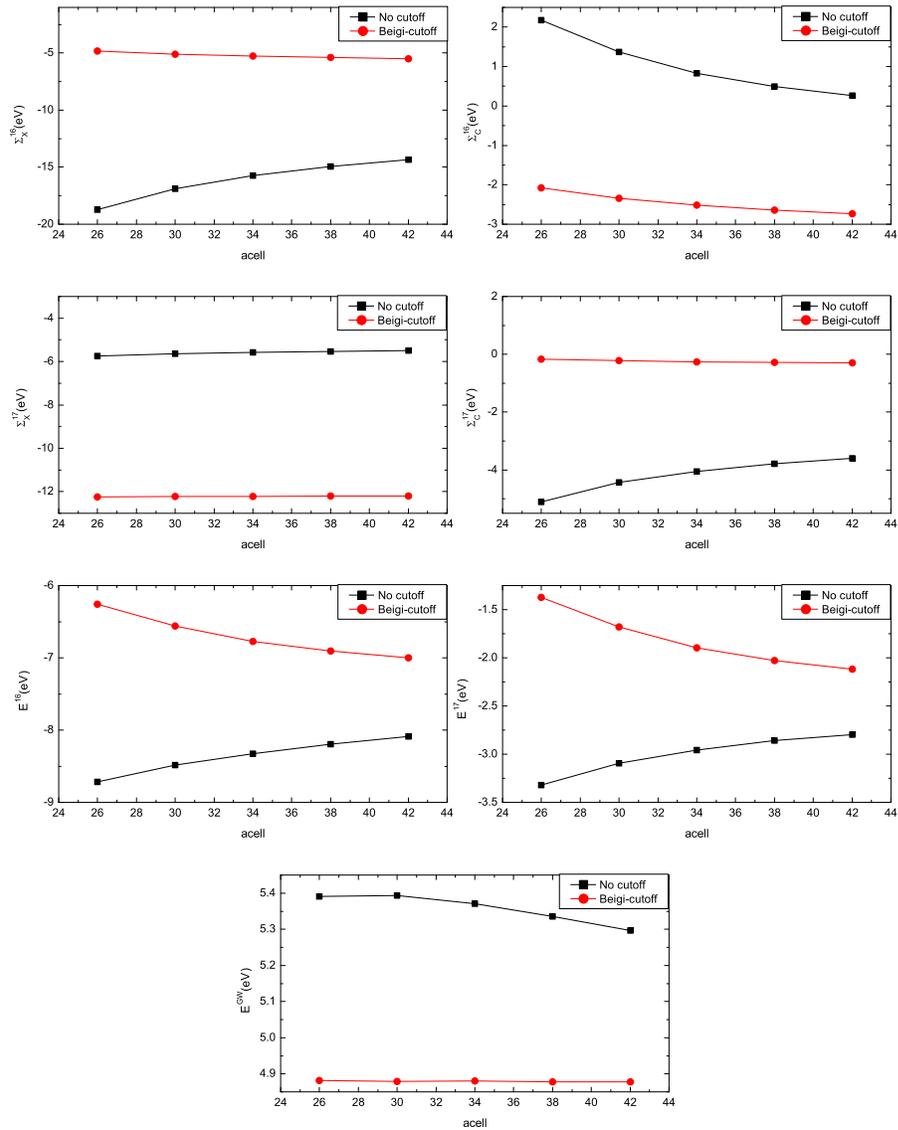


Figure A-7: Convergence of the GW calculations (shown are the correlation and exchange part of the self energy for the highest valence (16) and lowest conduction band (17), the energies of these bands and the GW bandgap) with respect to *acell* (size of unit cell in directions perpendicular to the wire) in the absence (black) and presence (red) of a Beigi cutoff.

8

Electronic structure of doped and undoped Si nanowires using a hybrid functional

8.1 Introduction

In the previous chapter it was shown that it is necessary to utilize the GW approximation in order to obtain good band gaps. A simple scissor shift is not enough to yield also good band structures. The application of the GW approximation, combined with the Wannier interpolation to obtain better band structures, is however limited due to its huge computational demands to small systems.

If one is able to study larger systems, new interesting research questions can be tackled. For example it is interesting to see how the band structure changes when the diameter is increased, especially since bulk silicon is an indirect band gap semiconductor, while wires grown in the [110] direction have direct band gaps due to folding of the Brillouin zone. Related to B or P dopants one can investigate to formation of an impurity band when the dopant concentration is reduced substantially. This requires a large unit cell in the growth direction.

An approach that can be used to overcome this problem is the usage of a so-called hybrid functional. These functionals basically combine Hartree-Fock and DFT and they provide improved electronic properties in comparison with standard DFT, but slightly less accurate structural properties. The main advantage compared to GW calculations is that the computational cost is smaller than full GW calculations. It is also possible to perform structural

relaxations using hybrid functionals.

A hybrid functional approach was already used to calculate the electronic structure of semiconducting and metallic carbon nanotubes, yielding good agreement with experimental data [123, 124]. More recently, Rurali *et al.* [125] also applied hybrid functionals to silicon nanowires to accurately determine electronic band gaps for varying diameter and orientation.

In this chapter, obtained in collaboration with E. Durgun and Ph. Ghosez from the Université de Liège (ULg), we study the properties of Si nanowires using an alternative hybrid functional recently proposed by Bilc *et al.* [126]. This so-called B1-WC functional combines the GGA of Wu and Cohen [127], which is known to be very accurate in predicting structural properties of solids, with 16 % exact exchange through the B1 scheme. In this chapter we investigate how accurate it is for Si nanowires, while it was originally designed for the study of oxides [126].

8.2 Computational details

The calculations presented here are performed in two steps. Starting from coordinates obtained with the ABINIT [63] or SIESTA [128] package (or based on these results), they are again optimized by using a plane wave basis set and projector augmented wave (PAW) [129, 130] method implemented in the VASP package [131, 132]. The exchange correlation potential has been approximated by local density approximation (LDA). A 8 Å vacuum layer between the wire images is included in order to prevent the interaction between wire images in the directions perpendicular to the wire. The Monkhorst-Pack scheme [52] was used to sample the Brillouin zone using a (1x1x15) \mathbf{k} -point mesh. For the plane wave basis set a cutoff energy of 500 eV was used. The convergence criterium for the energy is chosen as 10^{-5} eV between two ionic steps, and the maximum force allowed on each atom is $0.05\text{eV}/\text{Å}$.

In the second set of calculations a linear combination of atomic orbitals and norm conserving effective core potentials, as implemented in the CRYSTAL code [133], was used. In this step we used the hybrid B1-WC functional [134] functional. Other parameters were taken to be the same as those used during the VASP calculations. Because we used previously relaxed ABINIT or SIESTA calculations as a starting point for the hybrid calculations, only a few relaxation steps were needed and these steps did not have a significant effect on the bandstructures.

8.3 Results

One of the main goals of using the hybrid functional is to obtain better bandgaps without the computational burden of a GW calculation. So first we check if the bandgaps are indeed improved. To do that we plotted the calculated bandgaps for different diameter sizes. This is done in Fig. 8.1, that shows the GW bandgaps, the HFH hybrid bandgaps (obtained by Rurali *et al.* [125]), the B1-WC bandgaps and the LDA bandgaps. The curves are fits of the form

$$E_{gap} = E_{gap,bulk} + C \left(\frac{1}{d} \right)^\alpha, \quad (8.1)$$

where d is the diameter of the wire, as proposed by Delerue *et al.* [135]. For an effective mass particle-in-a-box approach the exponent α is equal to 2. Here we obtain a value of ~ 1 for the B1-WC hybrid, in agreement with the result for the HFH hybrid [125]. As expected the LDA bandgaps are underestimated. Compared to the HFH hybrid, the B1-WC hybrid has a slightly different slope. Scanning tunneling spectroscopy was also used to measure the bandgaps of Si nanowires [23]. These measurements are indicated on Fig. 8.1 with hollow triangles. The triangles pointing up are for [112] oriented wires, while the downwards pointing hollow triangle corresponds to a 3 nm diameter [110] oriented nanowire. From these experimental results it can be concluded that the hybrid functionals give good predictions of the bandgaps, at least for larger diameters.

Since the bandgaps are improved, let us see how the overall band structure obtained using the hybrid functionals compares to the full GW band structures. As the band gap of the hybrid functional is still a bit underestimated with respect to the GW bandgap, a small scissor shift is applied to align the conduction bands. For the 0.5 nm Si nanowire this is an additional shift of the conductance bands of 1.2 eV. In Fig. 8.2 this comparison is shown. The black lines are the hybrid results and the red lines the Wannier interpolated GW results (see previous chapter). Fig. 8.2(b) is a zoom of the region around the bandgap. If this is compared to the LDA result (see Fig. 7.8), it can be seen that the overall agreement with the GW calculations is slightly better (especially for the lowest conduction band) in case of the hybrid functional. It should be noted that the GW results were obtained using the LDA wavefunctions and that one traditionally only expects rigid shifts of the eigenvalues, while the hybrid wave functions can look different. Therefore it would be

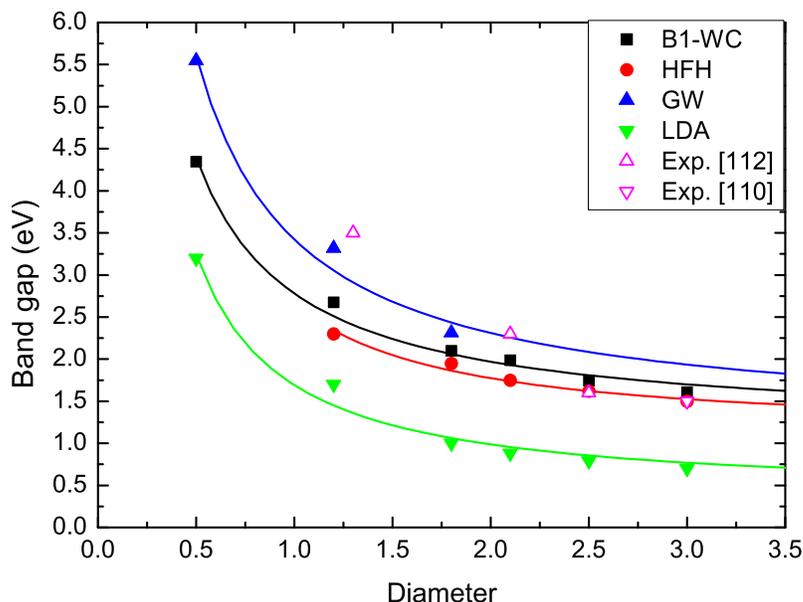


Figure 8.1: Bandgaps calculated using GW, HFH hybrid (see Ref. [125]), B1-WC hybrid and LDA. As reference the experimentally measured bandgaps for small diameter nanowires are added for [112] and [110] oriented nanowires (data from Ref. [23]).

interesting to use the wave functions and bandenergies obtained using a calculation with the hybrid functional as starting point for a GW calculation.

As the hybrids allow to study wires with much larger diameters, one can study the evolution of the band structure as function of the wire diameter. In Fig. 8.4 the band structures for wires with diameters ranging from 0.5 to 3 nm are shown. The corresponding relaxed structures are shown in Fig. 8.3. Note that when the wire diameter increases, the band gap decreases (as was showed previously) and the indirect band gap (bulk Si has an indirect band gap) starts to appear. Still all wires exhibit direct band gaps, as can be expected based on the folding of the bulk bands to the [110] direction. The conduction band minima in bulk silicon is located at about 85% along Γ to X . Therefore there exist 6 equivalent conduction band minima on $\pm x$, $\pm y$, and $\pm z$. When constructing a [110] wire 2 of these minima are projected onto the Γ point, while the other 4 are projected between the Γ point and the zone boundary. For the minima projected on Γ , based on the effective

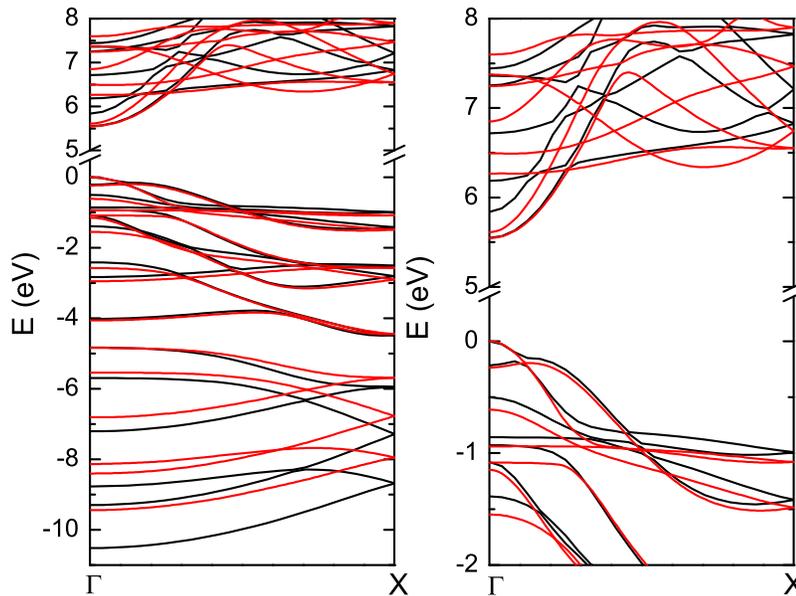


Figure 8.2: (a) Hybrid band structure (black lines) compared to GW band structure (red lines), and (b) a zoom in of the region around the bandgap.

mass approximation, both the longitudinal and transversal mass determine the confinement in the direction perpendicular to the growth direction. As the longitudinal mass is larger, it is the relevant mass for the confinement effect in the cross sectional plane. For the 4 other projections the effective mass lies in between the longitudinal and transversal mass. Therefore the conduction band at Γ has a smaller upward shift due to confinement [118]. When increasing the wire diameter it are these bands that start to become visible.

In chapter 3 we studied the formation energy of B and P atoms. While some of these results were obtained for highly doped wires, we showed that these formation energies converged and that these wires are actually stable (see chapter 5). What we did not consider was the limit of a single dopant and the associated forming of an impurity band. To obtain the exact impurity level with high accuracy it is necessary that the bandgap itself is also correct. Therefore we use the hybrid functional to calculate the impurity bands. We do this for B and P, by increasing the unit cell in the wire direction, so that the distance between dopants increases and thus the dopant concentration

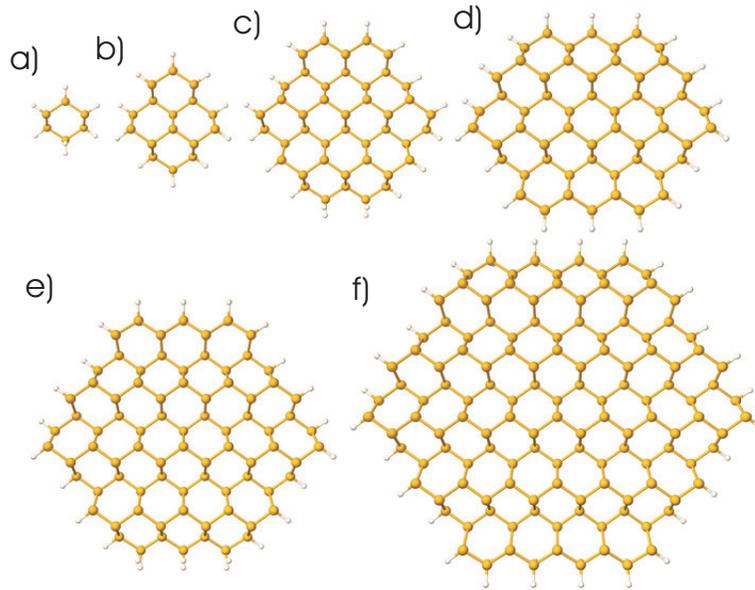


Figure 8.3: The different considered nanowires, ordered from (a) to (f) by increasing diameter.

decreases. In Fig. 8.5 the band structures of a P doped (placed on the preferential near edge position, see chapter 3) 1.2 nm nanowire are plotted for different number of unit cells in the z-direction. When the number of unit cells is increased the formation of an impurity band located near the bottom of the (undoped) conduction band, as P acts as a donor, can be observed. The width of this band decreases and it becomes more separated from the other bands. This is also seen in Fig. 8.7(a) that shows respectively the energy difference between the bottom of the (undoped) conduction band and the Fermi level, which is a measure of the location of the impurity band, and in (b) that shows the energy difference between the bottom of the impurity band and the Fermi level. This last energy difference thus corresponds to a measure of the width of the impurity band. As can be observed both values are converging. The width is converging to 0, which corresponds to a band without dispersion and the impurity band location is converging to 0.62 eV below the conduction band or equivalently 2.37 eV above the valence band. Similarly the band structure of a B doped nanowire is calculated, again for different numbers of unit cells in the z-direction. This is shown in Fig. 8.6. The formation of an acceptor level can be seen. We define the

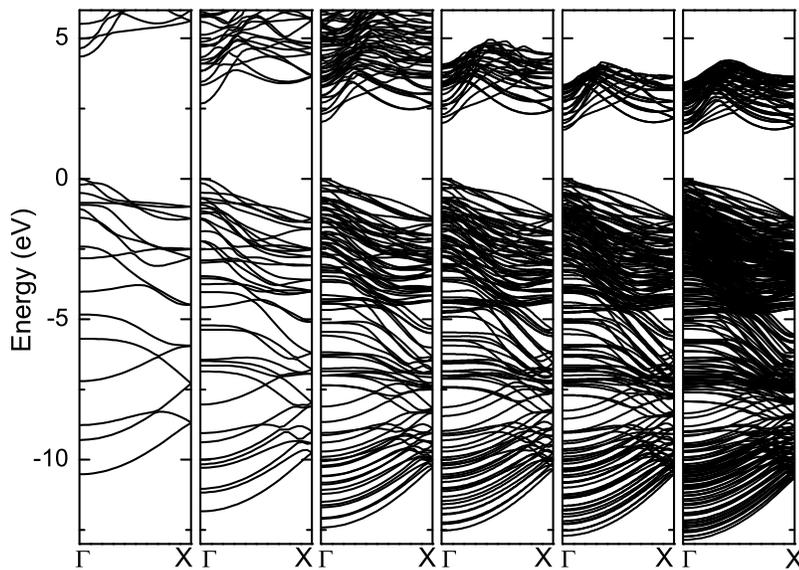


Figure 8.4: Band structures for different sized nanowires. From left to right the wire diameter is increased. The corresponding structures are shown in Fig. 8.3.

energy difference between the top of the valence band and the Fermi level as a measure of the location of the band and the difference between the top of the impurity band and the Fermi level as a measure of the width of the band. Both quantities are plotted in Fig. 8.7 (c) and (d) respectively. The width is converging towards a dispersion free band, while the location of the band is converging towards an energy level of 0.31 eV above the valence band (or 2.41 eV below the conduction band). Experimentally measuring the localization of this level is very challenging and therefore there are no experimental data available.

8.4 Conclusions

In this chapter we used hybrid functionals to be able to perform calculations of thick Si nanowires, while still obtaining an improvement over standard LDA calculations, especially with respect to the band gaps. This allowed us to study the evolution of band spectra of [110] directed nanowires with increasing diameter (till wires with a diameter of 3 nm). The band structures

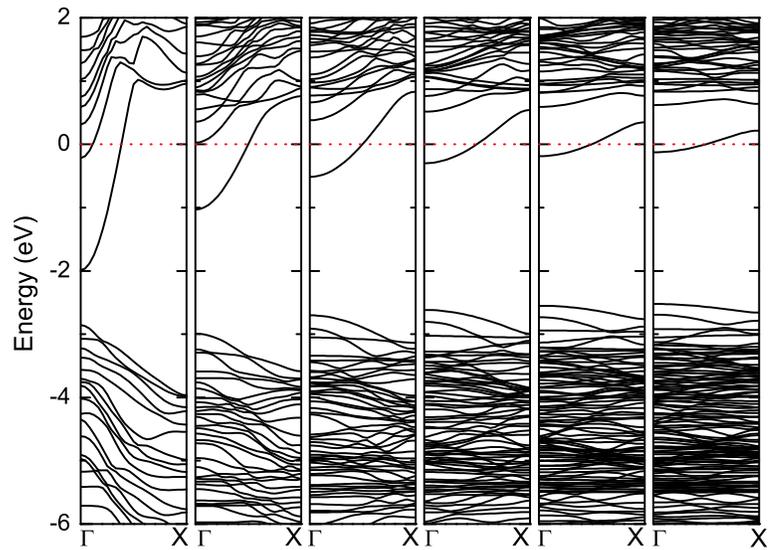


Figure 8.5: Band structures for a P doped 1.2 nm nanowire as function of the number of unit cells in the wire direction. From left to right the number of unit cells is increased from 1 to 6.

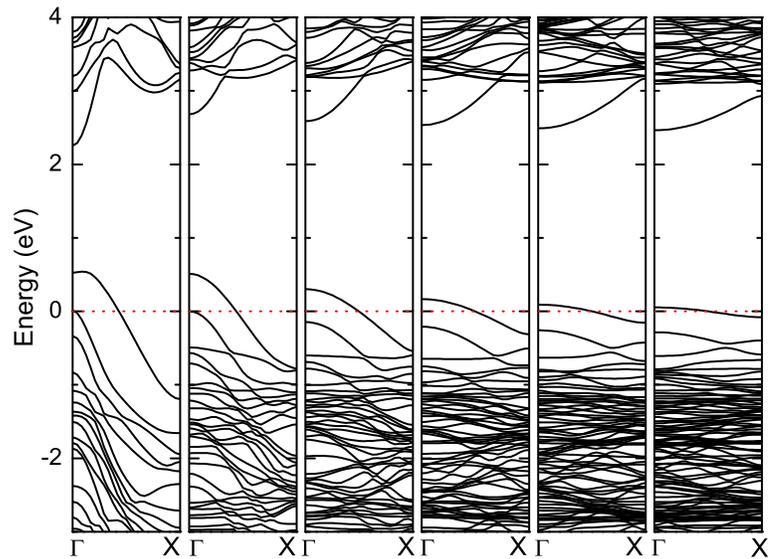


Figure 8.6: Band structures for a B doped 1.2 nm nanowire as function of the number of unit cells in the wire growth direction. From left to right the number of unit cells is increased from 1 to 6.

start to show more bulk like features, such as the indirect band gap (but the direct band gap is still smaller). Using large super cells in the wire direction, it was possible to study the formation of a donor (P) or acceptor (B) type single impurity band. The widths of these bands converge towards 0, so they become dispersionless. The P donor impurity band is located 0.62 eV below the conduction band and the B acceptor level is located at 0.31 eV above the valence band.

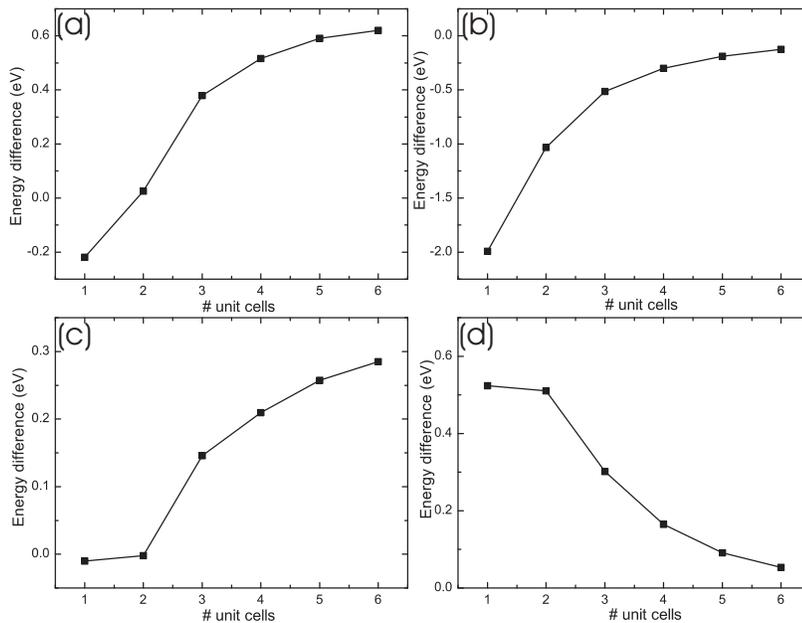


Figure 8.7: Convergence of several measures of the impurity bands as function of the number of unit cells in the wire direction. All energies are given with respect to the Fermi level. (a) The energy of the bottom of the conduction band in case of a P dopant, (b) the energy of the bottom of the P impurity band, (c) the energy of the top of the valence band in case of a B dopant, and (d) the energy of the top of the B band.

9

Conclusions and outlook

In this thesis I investigated the properties of silicon and germanium nanowires and silicon/germanium core-shell nanowires grown in the [110] direction by means of *ab initio* calculations.

In the **first chapter** an introduction is given to nanowire growth and several applications are discussed. The in this thesis utilized theoretical techniques are explained in other chapters. In **chapter 2** an introduction to density functional theory is given. This was followed in **chapter 4** by a brief explanation of phonon calculations using the density functional perturbation theory. The last theoretical chapter is **chapter 6**, in which the GW approximation is explained.

Inspired by the importance of doping for future applications and the unexplained experimental result that boron doped Si nanowires have higher conductivities than similarly phosphorus doped nanowires, I started with an investigation of the electronic properties of doped hydrogen passivated nanowires on an atomic level, which is shown in **chapter 3**. For this I used the formation and segregation energies to determine the most favorable dopant localization in a nanowire. This showed that the near edge position (a position close to the edge of the nanowire, but not bounded to one of the passivating hydrogen atoms) is the preferred position. As these nanowires have a large surface to volume ratio, the number of a dangling bond defects, which is the absence of one of the passivating hydrogen atoms, can be larger than the number of dopants. Therefore the effect of such a defect on the preferred localization of dopants was investigated. This defect increases the segregation energy substantially, especially in the case of P atoms where it becomes larger than 1 eV. For B atoms this segregation energy is smaller. These results offer an explanation for the observed conductance differences between B and P doped nanowires, as the additional electron of P dopants

can become trapped by a dangling bond, so that it can not contribute to the conductivity. This segregation towards the edges of the nanowire was later on shown experimentally.

Because we (and other authors) used small unit cells with a very high dopant concentration, the valid question can be raised if these nanowire structures are stable. Ground state calculations alone do not provide this information, therefore we perturbed the ionic positions using density functional perturbation theory. The calculated phonon spectra will contain imaginary frequencies when the structure is not stable. All previously used structures were stable, but an even thinner and doped nanowire (with a diameter of 0.5 nm, i.e. the thinnest possible nanowire in the [110] direction) is not stable. Increasing the unit cell in the growth direction stabilizes the structure, as long range relaxations become possible. The effect of the wire diameter on the phonon spectra is twofold: the (longitudinal) acoustic mode is hardened, while the optical modes are softened. As the phonon density of states allows to identify different contributions to the spectra, it was used in core-shell nanowires to distinguish between different core materials and shell sizes. The frequencies and velocities itself can have two different dependencies on the specific core-shell structure: the concentration of one of the materials (e.g. the Ge concentration) or the type of material in the core region. For these structures we also calculated the effective electron and hole masses, and the electron effective mass is quasi independent of the concentration or wire structure, while the hole mass depends strongly on the specific structure. These results can be used to experimentally characterize the wires (e.g. by Raman spectroscopy) and they are shown in chapter 5.

In **chapter 7** we performed a systematic convergence study of GW calculations on nanowires, where we tested the application of a Coulomb cutoff and an extrapolation technique. It was shown that these techniques allow to reduce the computational cost of doing a GW calculation. Using a Wannier interpolation we obtained full GW corrected bandstructures. These showed that the standard application of a scissor shift operation to shift the conduction bands of an LDA calculation so that the band gap corresponds to the quasi-particle band gap, yields only a good overall bandstructure for \mathbf{k} -points close to the Γ point and for bands close to the edges of the band gap. This difference in band structure can have an influence on the optical spectra.

While the techniques discussed in chapter 7 allow a reduction in the computational cost of a GW calculation, these calculations are still too demanding to treat larger systems. Therefore we used in **chapter 8** a hybrid

functional. This functional yields better band gaps (compared to both experiments and GW calculations) than a standard LDA calculation with comparable computational resources, and by comparing with a full GW calculation it was shown that the overall bandstructure is also slightly better than the corresponding LDA results. With this hybrid we investigated the effect of the wire diameter on the bandspectra, where the lowering of the bands associated with the indirect bandgap in bulk silicon could be observed. The formation of a B or P single impurity band, obtained by systematically reducing the concentration was observed and its localization with respect to the other bands was determined. Its width was also seen to converge towards a dispersionless band.

Currently the typical wire diameters treated in theoretical calculations and in experiments are different, as the theoretically treated wires are very thin (the largest being around 3 nm) compared to the ones grown experimentally. The continuing increase in computer processing power (i.e. Moore's law) will allow to treat larger diameter wires and/or use more precise theoretical techniques. On the other hand, experiments will be refined so that the typical wire diameters will become smaller. So in the future it will be possible to do calculations on real experimental realized nanowires and vice versa. However, most results currently obtained theoretically for thin nanowires can be used to understand experiments.

Future applications of nanowires will also become more and more important. For example, they will be used as solar cells. As was pointed out in the introduction p-i-n type of nanowires already showed promising performances. Other wire geometries are currently investigated for example at IMEC, where they also want to give a functional use to the oxide layer present on the grown wires. Theoretical calculations can help here to probe the effects of such an oxide layer.

10

Samenvatting

In deze thesis heb ik de eigenschappen van silicium en germanium nanodraden en van silicium/germanium kern-schil nanodraden bestudeerd met behulp van *ab initio* berekeningen.

Het **eerste hoofdstuk** bevat een inleiding over de groei van zulke nanodraden en uitleg van enkele toepassingen. De theoretische achtergrond van de gebruikte technieken wordt in andere hoofdstukken gegeven. Zo bevat **hoofdstuk 2** een inleiding tot de dichtheidsfunctionaaltheorie. **Hoofdstuk 4** bespreekt dan kort de berekening van fononen met behulp van de dichtheidsfunctionaal perturbatie theorie. Het laatste theoretische hoofdstuk is **hoofdstuk 6**, waarin de GW benadering wordt uitgelegd.

Aangezien het belangrijk is om de eigenschappen van gedopeerde draden te kennen voor toekomstige toepassingen en om het nog niet verklaarde experimenteel resultaat dat boron gedopeerde Si nanodraden een hogere conductiviteit hebben dan fosfor gedopeerde nanodraden (voor een zelfde concentratie aan doperende atomen) te verklaren, heb ik in **hoofdstuk 3** de elektronische eigenschappen van waterstof gepassiveerde nanodraden bestudeerd op een atomair niveau. Hiervoor heb ik gebruik gemaakt van de vormings- en segregatie energie om op die manier de preferentiële positie van de doperende atomen te vinden. Dit toonde aan dat posities aan de rand van de nanodraad te verkiezen zijn boven posities gelegen in het centrum van de draad. Aangezien nanodraden een grote oppervlakte/volume-verhouding hebben, kunnen het aantal “dangling-bond” (een ontbrekend passiverend waterstof-atoom) defecten groter zijn dan het aantal doperende atomen. Er werd dan ook onderzocht op welke manier dit de eigenschappen beïnvloedt. De aanwezigheid van een dergelijk defect vergroot de segregatie energie aanzienlijk, tot zelfs meer dan 1 eV voor P atomen, terwijl het voor B atomen veel kleiner is. Het grote verschil in segregatie energie zou kunnen verklaren

waarom de conductiviteit lager is voor P gedopeerde nanodraden, aangezien de aanwezigheid van het dangling-bonddefect ervoor zorgt dat de extra ladingsdragers inactief worden. Het energetisch voordelig zijn van randposities is ook experimenteel waargenomen.

Aangezien wij (en andere auteurs) gebruik gemaakt hebben van kleine eenheidscellen die een zeer hoge concentratie doperende atomen bevatten, kan men zich afvragen of deze structuren nog stabiel zijn. Berekeningen van de grondtoestand leveren deze informatie niet op. Daarom maakten we in **hoofdstuk 5** gebruik van dichtheidsfunctionaal perturbatie theorie om de posities van de atomen te verstoren en op die manier met behulp van de fonon-spectra de stabiliteit te bepalen. Imaginaire frequenties duiden immers op instabiliteit. Alle voorgaande structuren zijn stabiel, maar wanneer men een nog kleinere diameter draad neemt (van 0.5 nm), dan is deze niet meer stabiel na dopering. Deze kan wel gestabiliseerd worden door de eenheidscel te vergroten in de groeirichting, omdat er dan langere afstands relaxaties mogelijk worden. De diameter van de draad heeft twee verschillende effecten op de fononspectra: als eerste worden de akoestische modes zachter (de akoestische snelheid neemt af) wanneer de diameter van de draad afneemt en een tweede effect is dat de optische modes harder worden (dus de frequenties nemen toe). De fononspectra laten ook toe, in de vorm van toestandsdichtheden (DOS), om een onderscheid te maken tussen draden met verschillende diameters en materialen, zodat het dan ook gebruikt werd om in kern-schil draden de diameter en de gebruikte materialen (Si of Ge) in de schil te bepalen. Wanneer men de snelheden en frequenties als functie van de concentratie en het draadtype (het materiaal dat zich in de kern bevindt) bekijkt, dan kan men twee verschillende verbanden onderscheiden: ofwel hangt de snelheid of frequentie enkel af van de concentratie, ofwel is het materiaal dat zich in de kern (en het andere dus in de schil) bevindt het belangrijkste. Qua elektronische eigenschappen blijkt het toevoegen van een kleine kern van een ander materiaal ervoor te zorgen dat de bandkloof afneemt ten opzichte van de zuivere draad. De effectieve massa van de elektronen is quasi onafhankelijk van de concentratie of van het type draad, terwijl de effectieve massa van de holtes sterk afhangt van de structuur en concentratie. Deze resultaten kunnen gebruikt worden om experimenteel nanodraden te karakteriseren (bv. met Raman spectroscopie).

In **hoofdstuk 7** werd een systematische convergentiestudie van GW berekeningen op nanodraden uitgevoerd, waarbij twee nieuwe technieken getest werden, namelijk een Coulomb cutoff techniek en een extrapolaire

techniek. Beide technieken laten toe om de computationele kost van een GW berekening te verminderen. Vervolgens werd gebruik gemaakt van een Wannier interpolatie om een volledig GW gecorrigeerd bandspectrum te bekomen. Aan de hand van deze bandspectra werd er getoond dat het toepassen van een “scissor shift” (dit is het uniform verschuiven van de conductiebanden om zo een correcte bandkloof te bekomen) enkel goed is voor banden dicht bij de bandkloof gelegen en voor \mathbf{k} -punten dicht bij het Γ punt. Dit blijkt ook uit de berekende bandmassa’s, die zowel bij een niet GW gecorrigeerde berekening als een gecorrigeerde hetzelfde resultaat opleveren. Het verschil in dispersie van de banden dicht bij de bandkloof tussen de LDA en de GW berekeningen kan aanleiding geven tot verschillen in de optische spectra.

Alhoewel de technieken besproken in hoofdstuk 7 de computationele kost van een GW berekening aanzienlijk verminderen, zijn deze berekeningen toch te veeleisend om grotere systemen te beschrijven. Daarom hebben we in **hoofdstuk 8** gebruik gemaakt van hybride functionalen. Een dergelijke functionaal levert betere bandkloven op dan een zelfde LDA of GGA berekening en laat bovendien toe om een relaxatie uit te voeren. Een vergelijking met de volledige gecorrigeerde GW resultaten toonde bovendien aan dat de bekomen bandenstructuur ook goed is (iets beter dan de LDA bandenstructuur). Deze hybride werd dan gebruikt om de invloed van de diameter van de nanodraad op de bandenstructuur te onderzoeken. Bij toenemende diameter komen de banden, die in bulk Si voor de indirecte bandkloof zorgen, lager in energie te liggen. De vorming van een B of P onzuiverheidsband bij afnemende concentratie werd eveneens bestudeerd en de lokalisatie van deze band ten opzichte van de andere banden werd bepaald. Ook de breedte van deze band convergeerde naar een dispersieloze band, zoals verwacht kan worden van een enkele onzuiverheid.

Momenteel zijn de typische diameters van de nanodraden bestudeerd in theoretische berekeningen (tot ongeveer 3 nm) veel kleiner dan de experimenteel gerealiseerde nanodraden. De steeds toenemende computersnelheid (zoals voorspeld door de wet van Moore) zal toelaten om nanodraden met steeds grotere diameters te beschouwen en/of meer geavanceerde theoretische technieken toe te passen. Tegelijkertijd zullen de experimenten steeds meer verfijnd worden, zodat de diameters ook zullen verkleinen. In de toekomst zal het dus mogelijk zijn om berekeningen uit te voeren op experimentele nanodraden en omgekeerd experimenten te doen op theoretisch berekende draden. De meeste van de reeds theoretisch gevonden resultaten voor dunne

CHAPTER 10. SAMENVATTING

nanodraden kunnen echter nu ook al gebruikt worden om de huidige experimenten te begrijpen.

Toekomstige toepassingen van nanodraden zullen ook steeds belangrijker worden. Zo zullen nanodraden bijvoorbeeld gebruikt worden als zonnecellen. Zoals reeds in de inleiding besproken, hebben p-i-n nanodraden reeds een veelbelovende performantie. Andere vormen van nanodraden worden momenteel onderzocht, bijvoorbeeld door IMEC. Daar wil men ook de steeds voorkomende oxide-laag als functioneel deel van de zonnecel gebruiken. Theoretische berekeningen kunnen hier helpen om de eigenschappen van deze draden te onderzoeken.

Curriculum Vitae

Name: **Hartwin Peelaers**

Born: 29/09/1983, Edegem

Contact: Campus Groenenborger,
Groenenborgerlaan 171,
B-2020 Antwerp, Belgium

E-mail: hartwin.peelaers@ua.ac.be



Education

2005–2009 Department of Physics, University of Antwerp
PhD in Physics

2001–2005 Department of Physics, University of Antwerp
Master in Physics, highest honors

Teaching experience

Exercises: *Biostatistics*: courses in 2006-2007 and 2007-2008.

Numerical methods: courses in 2008-2009 and 2009-2010.

Computer skills

Operating systems: Windows, Linux

Office: Word, Excel, Access, PowerPoint

Programming languages: C/C++, Fortran, Python, java, bash
scripting, MPI

Ab initio packages: ABINIT, Siesta, Wannier90

Mathematical software: Mathematica, MATLAB, Maple, Comsol

Other: Linux cluster administration

Language Proficiency

Dutch (native), English (fluent), French (good)

List of publications

- H. Peelaers, B. Partoens, and F. M. Peeters: *Formation and segregation energies of B and P doped, and BP codoped silicon nanowires*, Nano Lett. **6**, 2781 (2006)
- H. Peelaers, B. Partoens, D.V. Tatyanyenko, and F. M. Peeters: *Dynamics of scattering on a classical 2D artificial atom*, Phys. Rev. E **75**, 036606 (2007)
- H. Peelaers, B. Partoens, and F. M. Peeters: *Properties of B and P doped Ge nanowires*, Appl. Phys. Lett. **90**, 263103 (2007). Also selected for the Virtual Journal of Nanoscale Science & Technology **16**, 2 (2007).
- H. Peelaers, B. Partoens, and F. M. Peeters: *First-principles study of doped Si and Ge nanowires*, Physica E **40**, 2169 (2008).
- H. Peelaers, B. Partoens, and F. M. Peeters: *Phonon Band Structure of Si Nanowires: A Stability Analysis*, Nano Lett. **9**, 107 (2009).
- H. Peelaers, B. Partoens, and F. M. Peeters: *Phonons in Ge nanowires*, Appl. Phys. Lett. **95**, 122110 (2009). Also selected for the Virtual Journal of Nanoscale Science & Technology **20**, 14 (2009).
- H. Peelaers, B. Partoens, and F. M. Peeters: *Electronic and dynamical properties of Si/Ge core-shell nanowires* (submitted)
- H. Peelaers, B. Partoens, M. Giantomassi, T. Rangel, E. Goossens, X. Gonze, G.-M. Rignanese, and F. M. Peeters: *Full quasi-particle corrected band structures of Si and Ge nanowires* (in preparation)
- H. Peelaers, E. Durgun, B. Partoens, Ph. Ghosez, and F. M. Peeters: *Electronic structure of doped and undoped Si nanowires using a hybrid functional* (in preparation)

Bibliography

- [1] <http://www.itrs.net/>.
- [2] Y. Huang, X. Duan, Q. Wei, and C. M. Lieber, *Science* **291**, 630 (2001), <http://www.sciencemag.org/cgi/reprint/291/5504/630.pdf>.
- [3] P. X. Gao and Z. L. Wang, *J. Phys. Chem. B* **106**, 12653 (2002).
- [4] Y. Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, *Nano Lett.* **3**, 149 (2003).
- [5] X. Duan, Y. Huang, Y. Cui, J. Wang, and C. M. Lieber, *Nature* **409**, 66 (2001).
- [6] Y. Cui and C. M. Lieber, *Science* **291**, 851 (2001).
- [7] M. H. Huang, S. Mao, H. Feick, H. Yan, Y. Wu, H. Kind, E. Weber, R. Russo, and P. Yang, *Science* **292**, 1897 (2001).
- [8] Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K.-H. Kim, and C. M. Lieber, *Science* **294**, 1313 (2001).
- [9] Y. Cui, Q. Wei, H. Park, and C. M. Lieber, *Science* **293**, 1289 (2001).
- [10] F. Patolsky, G. Zheng, O. Hayden, M. Lakadamyali, X. Zhuang, and C. M. Lieber, *Proc. Natl. Acad. Sci. USA* **101**, 14017 (2004).
- [11] B. Tian, X. Zheng, T. J. Kempa, Y. Fang, N. Yu, G. Yu, J. Huang, and C. M. Lieber, *Nature* **449**, 885 (2007).
- [12] B. Tian, T. J. Kempa, and C. M. Lieber, *Chem. Soc. Rev.* **38**, 16 (2009).
- [13] Z. L. Wang and J. Song, *Science* **312**, 242 (2006).
- [14] E. Stern, J. F. Klemic, D. A. Routenberg, P. N. Wyrembak, D. B. Turner-Evans, A. D. Hamilton, D. A. LaVan, T. M. Fahmy, and M. A. Reed, *Nature* **445**, 519 (2007).
- [15] V. Pott and A. M. Ionescu, *Microelectron. Eng.* **83**, 1718 (2003).

BIBLIOGRAPHY

- [16] J. Martinez, R. V. Martinez, and R. Garcia, *Nano Lett.* **8**, 3636 (2008).
- [17] V. Schmidt, J. V. Wittemann, S. Senz, and U. Gsele, *Adv. Mat.* **21**, 2681 (2009).
- [18] R. S. Wagner and W. C. Ellis, *Appl. Phys. Lett.* **4**, 89 (1964).
- [19] Y. Xia, P. Yang, Y. Sun, Y. Wu, B. Mayers, B. Gates, Y. Yin, F. Kim, and H. Yan, *Adv. Mat.* **15**, 353 (2003).
- [20] Y. Wu and P. Yang, *J. Am. Chem. Soc.* **123**, 3165 (2001).
- [21] C. M. Lieber and Z. L. Wang, *MRS Bull.* **32**, 99 (2007).
- [22] A. M. Morales and C. M. Lieber, *Science* **279**, 208 (1998).
- [23] D. D. D. Ma, C. S. Lee, F. C. K. Au, S. Y. Tong, and S. T. Lee, *Science* **299**, 1874 (2003).
- [24] Y. Wu, Y. Cui, L. Huynh, C. J. Barrelet, D. C. Bell, and C. M. Lieber, *Nano Lett.* **4**, 433 (2004).
- [25] H. Jagannathan, M. Deal, Y. Nishi, J. Woodruff, C. Chidsey, and P. C. McIntyre, *J. Appl. Phys.* **100**, 024318 (2006).
- [26] D. Wang and H. Dai, *Angew. Chem. Int. Ed.* **41**, 4783 (2002).
- [27] T. Bryllert, L.-E. Wernersson, T. Lowgren, and L. Samuelson, *Nanotechnology* **17**, S227 (2006).
- [28] C. Thelander, P. Agarwal, S. Brongersma, J. Eymery, L. Feiner, A. Forchel, M. Scheffler, W. Riess, B. Ohlsson, U. Gsele, and L. Samuelson, *Mater. Today* **9**, 28 (2006).
- [29] J. Xiang, W. Lu, Y. Hu, Y. Wu, H. Yan, and C. M. Lieber, *Nature* **441**, 489 (2006).
- [30] L. J. Lauhon, M. S. Gudixsen, D. Wang, and C. M. Lieber, *Nature* **420**, 57 (2002).
- [31] F. Patolsky, B. P. Timko, G. Zheng, and C. M. Lieber, *MRS Bull.* **32**, 142 (2007).

- [32] G. Zheng, F. Patolsky, Y. Cui, W. U. Wang, and C. M. Lieber, *Nat. Biotechnol.* **23**, 1294 (2005).
- [33] X. Bi, W. L. Wong, W. Ji, A. Agarwal, N. Balasubramanian, and K.-L. Yang, *Biosensors and Bioelectronics* **23**, 1442 (2008).
- [34] Y. Li, F. Qian, J. Xiang, and C. M. Lieber, *Mater. Today* **9**, 18 (2006).
- [35] Y. Cui, X. Duan, J. Hu, and C. M. Lieber, *J. Phys. Chem. B* **104**, 5213 (2000).
- [36] J.-Y. Yu, S.-W. Chung, and J. R. Heath, *J. Phys. Chem. B* **104**, 11864 (2000).
- [37] R. M. Martin, *Electronic Structure* (Cambridge University Press, Cambridge, 2004).
- [38] U. von Barth, *Phys. Scr.* **T109**, 9 (2004).
- [39] K. Capelle, *Braz. J. Phys.* **36**, 1318 (2006).
- [40] R. O. Jones, Introduction to density functional theory and exchange-correlation energy functionals, in *Computational Nanoscience: Do it yourself!*, edited by J. Grotendorst, S. Blügel, and D. Marx, , NIC Series Vol. 31, pp. 45–70, John von Neumann Institute for Computing, Jülich, 2006.
- [41] M. Born and R. Oppenheimer, *Annalen der Physik* **389**, 457 (1927).
- [42] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [43] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [44] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- [45] S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).
- [46] O. Gunnarsson, M. Jonson, and B. I. Lundqvist, *Phys. Rev. B* **20**, 3136 (1979).
- [47] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).

BIBLIOGRAPHY

- [48] A. D. Becke, *J.Chem.Phys* **98**, 5648 (1993).
- [49] C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- [50] H. Hellmann, *Einführung in die Quantenchemie* (Deuticke, Leipzig, 1937).
- [51] R. P. Feynman, *Phys. Rev.* **56**, 340 (1939).
- [52] H. J. Monkhorst and J. D. Pack, *Phys. Rev. B* **13**, 5188 (1976).
- [53] N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).
- [54] D. D. D. Ma, C. S. Lee, and S. T. Lee, *Appl. Phys. Lett.* **79**, 2468 (2001).
- [55] K.-K. Lew, L. Pan, T. E. Bogart, S. M. Dilts, E. C. Dickey, J. M. Redwing, Y. Wang, M. Cabassi, T. S. Mayer, and S. W. Novak, *Appl. Phys. Lett.* **85**, 3101 (2004).
- [56] Y. Cui, L. J. Lauhon, M. S. Gudiksen, J. Wang, and C. M. Lieber, *Appl. Phys. Lett.* **78**, 2214 (2001).
- [57] N. Wang, Y. H. Tang, Y. F. Zhang, C. S. Lee, I. Bello, and S. T. Lee, *Chem. Phys. Lett.* **299**, 237 (1999).
- [58] A. K. Singh, V. Kumar, R. Note, and Y. Kawazoe, *Nano Lett.* **5**, 2302 (2005).
- [59] T.-L. Chan, C. V. Ciobanu, F.-C. Chuang, N. Lu, C.-Z. Wang, and K.-M. Ho, *Nano Lett.* **6**, 277 (2006).
- [60] X. Zhao, C. M. Wei, L. Yang, and M. Y. Chou, *Phys. Rev. Lett.* **92**, 236805 (2004).
- [61] A. K. Singh, V. Kumar, R. Note, and Y. Kawazoe, *Nano Lett.* **6**, 920 (2006).
- [62] M. V. Fernandez-Serra, C. Adessi, and X. Blase, *Phys Rev Lett* **96**, 166805 (2006).

- [63] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, and D. C. Allan, *Comput. Mater. Sci.* **25**, 478 (2002).
- [64] G. Cantele, E. Degoli, E. Luppi, R. Magri, D. Ninno, G. Iadonisi, and S. Ossicini, *Phys. Rev. B* **72**, 113303 (2005).
- [65] S. Ossicini, E. Degoli, F. Iori, E. Luppi, R. Magri, G. Cantele, F. Trani, and D. Ninno, *Appl. Phys. Lett.* **87**, 173120 (2005).
- [66] A. Stesmans and V. V. Afanasev, *J. Phys.: Condens. Matter* **10**, L19 (1998).
- [67] R. Kagimura, R. W. Nunes, and H. Chacham, *Phys. Rev. Lett.* **95**, 115502 (2005).
- [68] M. Bruno, M. Palummo, A. Marini, R. D. Solo, V. Olevano, A. N. Kholod, and S. Ossicini, *Phys. Rev. B* **72**, 153310 (2005).
- [69] S. P. Beckman, J. Han, and J. R. Chelikowsky, *Phys. Rev. B.* **74**, 165314 (2006).
- [70] E. Tutuc, J. O. Chu, J. A. Ott, and S. Guha, *Appl. Phys. Lett.* **89**, 263101 (2006).
- [71] G. Imamura, T. Kawashima, M. Fujii, C. Nishimura, T. Saitoh, and S. Hayashi, *Nano Lett.* **8**, 2620 (2008).
- [72] T. Kawashima, G. Imamura, T. Saitoh, K. Komori, M. Fujii, and S. Hayashi, *J. Phys. Chem. C* **111**, 15160 (2007).
- [73] M. C. Putnam, M. A. Filler, B. M. Kayes, M. D. Kelzenberg, Y. Guan, N. S. Lewis, J. M. Eiler, and H. A. Atwater, *Nano Lett.* **8**, 3109 (2008).
- [74] J. E. Allen, E. R. Hemesath, D. E. Perea, J. L. Lensch-Falk, LiZ.Y., F. Yin, M. H. Gass, P. Wang, A. L. Bleloch, R. E. Palmer, and L. J. Lauhon, *Nature Nanotech* **3**, 168 (2008).
- [75] J. E. Allen, D. E. Perea, E. R. Hemesath, and L. J. Lauhon, *Adv. Mat.* **21**, 3067 (2009).

BIBLIOGRAPHY

- [76] P. Xie, Y. Hu, Y. Fang, J. Huang, and C. M. Lieber, Proc Natl Acad Sci USA **106**, 15254 (2009), <http://www.pnas.org/content/106/36/15254.full.pdf+html>.
- [77] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).
- [78] X. Gonze, Phys. Rev. A **52**, 1096 (1995).
- [79] S. Baroni, P. Giannozzi, and A. Testa, Phys. Rev. Lett. **58**, 1861 (1987).
- [80] R. M. Sternheimer, Phys. Rev. **96**, 951 (1954).
- [81] M. J. Verstraete and X. Gonze, Phys. Rev. B **74**, 153408 (2006).
- [82] M. Stroschio and M. Dutta, *Phonons in Nanostructures* (Cambridge University Press, Cambridge, 2001).
- [83] E. P. Pokatilov, D. L. Nika, and A. A. Balandin, Superlattices Microstruct. **38**, 168 (2005).
- [84] N. Mingo and L. Yang, Phys. Rev. B **68**, 245406 (2003).
- [85] S. P. Hepplestone and G. P. Srivastava, Nanotechnology **17**, 3288 (2006).
- [86] A. K. Buin, A. Verma, A. Svizhenko, and M. P. Anantram, Nano Lett. **8**, 760 (2008).
- [87] R. Saito and G. Dresselhaus, *Physical properties of carbon nanotubes* (Imperial College Press, London, 1998).
- [88] G. Lucovsky, R. J. Nemanich, and J. C. Knights, Phys. Rev. B **19**, 2064 (1979).
- [89] J. T. Arantes and A. Fazzio, Nanotechnology **18**, 295706 (6pp) (2007).
- [90] H. Peelaers, B. Partoens, and F. M. Peeters, Appl. Phys. Lett. **90**, 263103 (2007).
- [91] F. Comas, I. Camps, G. E. Marques, and N. Studart, J. Appl. Phys. , 033525 (2007).

- [92] G. A. N. Connell and J. R. Pawlik, *Phys. Rev. B* **13**, 787 (1976).
- [93] H. Peelaers, B. Partoens, and F. M. Peeters, *Nano Lett.* **9**, 107 (2009).
- [94] Y. Hu, J. Xiang, G. Liang, H. Yan, and C. M. Lieber, *Nano Lett.* **8**, 925 (2008).
- [95] W. Lu, J. Xiang, B. P. Timko, Y. Wu, and C. M. Lieber, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10046 (2005).
- [96] J.-E. Yang, C.-B. Jin, C.-J. Kim, and M.-H. Jo, *Nano Lett.* **6**, 2679 (2006).
- [97] R. N. Musin and X.-Q. Wang, *Phys. Rev. B* **71**, 155318 (2005).
- [98] R. N. Musin and X.-Q. Wang, *Phys. Rev. B* **74**, 165308 (2006).
- [99] Y. Zhang and Y. Xiao, *Eur. Phys. J. B* **63**, 425 (2008).
- [100] H. Peelaers, B. Partoens, and F. M. Peeters, *Appl. Phys. Lett.* **95**, 122110 (2009).
- [101] L. Ma, J. Wang, J. Zhao, and G. Wang, *Chem. Phys. Lett.* **452**, 183 (2008).
- [102] P. W. Leu, A. Svizhenko, and K. Cho, *Phys. Rev. B* **77**, 235305 (2008).
- [103] S. Hoffmann, I. Utke, B. Moser, J. Michler, S. H. Christiansen, V. Schmidt, S. Senz, P. Werner, U. Gosele, and C. Ballif, *Nano Lett.* **6**, 622 (2006).
- [104] L. Hedin, *Phys. Rev.* **139**, A796 (1965).
- [105] A. Fetter and J. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).
- [106] C. Friedrich and A. Schindlmayr, Many-body perturbation theory: The gw approximation, in *Computational Nanoscience: Do it yourself!*, edited by J. Grotendorst, S. Blügel, and D. Marx, , NIC Series Vol. 31, pp. 335–355, John von Neumann Institute for Computing, Jülich, 2006.

BIBLIOGRAPHY

- [107] F. Bruneval, *Exchange and Correlation in the Electronic Structure of Solids, from Silicon to Cuprous Oxide: GW Approximation and beyond*, PhD thesis, Ecole Polytechnique, Palaiseau Cedex, 2005.
- [108] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
- [109] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, Solid State Phys. **54**, 1 (1999).
- [110] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
- [111] V. M. Galitskii and A. B. Migdal, Zh. ksp. Teor. Fiz. **34**, 139 (1958).
- [112] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
- [113] G. E. Engel and B. Farid, Phys. Rev. B **47**, 15931 (1993).
- [114] S. Ismail-Beigi, Phys. Rev. B **73**, 233103 (2006).
- [115] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, Phys. Rev. B **73**, 205119 (2006).
- [116] F. Bruneval and X. Gonze, Phys. Rev. B **78**, 085125 (2008).
- [117] M. Bruno, M. Palummo, A. Marini, R. D. Sole, and S. Ossicini, Phys. Rev. Lett. **98**, 036807 (2007).
- [118] J.-A. Yan, L. Yang, and M. Y. Chou, Phys. Rev. B **76**, 115319 (2007).
- [119] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).
- [120] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).
- [121] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008).
- [122] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
- [123] P. V. Avramov, K. N. Kudin, and G. E. Scuseria, Chem. Phys. Lett. **370**, 597 (2003).

- [124] V. Barone, J. E. Peralta, J. Uddin, and G. E. Scuseria, *J.Chem.Phys* **124**, 024709 (2006).
- [125] R. Rurali, B. Aradi, T. Frauenheim, and A. Gali, *Phys. Rev. B* **76**, 113303 (2007).
- [126] D. I. Bilc, R. Orlando, R. Shaltaf, G.-M. Rignanese, J. I. niguez, and P. Ghosez, *Phys. Rev. B* **77**, 165107 (2008).
- [127] Z. Wu and R. E. Cohen, *Phys. Rev. B* **73**, 235116 (2006).
- [128] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, *J.Phys.:Condens.Matter* **14**, 2745 (2002).
- [129] G. Kresse and D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).
- [130] P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).
- [131] G. Kresse and J. Hafner, *Phys. Rev. B* **47**, 558 (1993).
- [132] G. Kresse and J. Furthmüller, *Phys. Rev. B* **54**, 11169 (1996).
- [133] R. Dovesi, R. Orlando, B. Civalleri, C. Roetti, V. R. Saunders, and C. M. Zicovich-Wilson, *Z. Kristallogr.* **220**, 571 (2005).
- [134] A. D. Becke, *J.Chem.Phys* **104**, 1040 (1996).
- [135] C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* **48**, 11024 (1993).