SISSA - SCUOLA
INTERNAZIONALE
SUPERIORE
DI STUDI AVANZATI
TRIESTE
Strada Costiera 11

ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

ENERGY–BAND DISCONTINUITIES AT LATTICE–MATCHED SEMICONDUCTOR HETEROJUNCTIONS

Thesis submitted
for the degree of "Doctor Philosophiae"

October 1989

Candidate:

M. Peressi

Supervisors:

A. Baldereschi
S. Baroni

TRIESTE

ENERGY–BAND DISCONTINUITIES AT LATTICE–MATCHED SEMICONDUCTOR HETEROJUNCTIONS

Thesis submitted
for the degree of "Doctor Philosophiae"

October 1989

Candidate:

M. Peressi

Supervisors:

A. Baldereschi
S. Baroni

Acknowledgements

I wish to express my gratitude to Prof. A. Baldereschi and Prof. S. Baroni for their support during the course of this work. I am also grateful to Prof. R. Resta for clarifying discussions, to Dr. A. Fasolino for a precious reading of the first part of the manuscript, to Dr. P. Giannozzi and S. De Gironcoli for help their during the computational work.

ENERGY–BAND DISCONTINUITIES
AT LATTICE–MATCHED SEMICONDUCTOR HETEROJUNCTIONS

INTRODUCTION

Semiconductor quantum structures have attracted growing attention during the last two decades from scientists devoted both to basic research and applied science. The field of semiconductor heterostructures covers a variety of aspects and different systems of increasing complexity, ranging from the simple *heterojunction*, i.e. the contact between two different semiconductors, to more complex multilayered structures. High–quality semiconductor heterostructures are nowadays routinely realizable thanks to the technological progress, and find application in new optoelectronic and microelectronic devices.

From the point of view of basic physics, this activity has opened many new problems. In the present work I focus the investigation on a particular one, i.e. the electronic properties at an isolated interface between two different semiconductors, and, in particular, to the valence and conduction band–edge offsets (VBO and CBO), usually collectively referred to as the *band offset*. This problem is important first of all from a fundamental point of view, since it is related to the basic question about the existence of an absolute reference energy scale to be used for the energy–band structures of all semiconductors. Furthermore, since the band discontinuities are one of the key parameters governing the transport properties at the semiconductor heterojunctions, it is very important for the technological consequences to know the VBO and CBO values and the factors which influence them; a large number of experimental data are now available but their spread reveals the difficulty of precise and accurate measurements. The main question concerning the band lineup is whether it is determined only by the bulk properties of the two constituents, or if it is also related to some interface–specific phenomena (e.g. interface orientation, substitutional disorder, atomic dislocations...) which critically affect it and could therefore be used to *tune* it. A theoretical approach to the problem is necessary not only to discriminate between contradictory experimental results, but also for a general understanding of the problem.

The present work is limited to heterojunctions between materials having the same lattice parameters, and the attention is focused in particular on the VBO.

The achievements of the present work can be summarized into two main points:

i) Accurate self–consistent (SCF) supercell calculations are performed in order to obtain both the values of VBO for some systems and the general qualitative features of isovalent and heterovalent heterojunctions. I will try to identify similarities and differences between these classes of heterojunctions, in order to achieve a general understanding of the interplay between bulk and interface phenomena. As for the isovalent interfaces, previous studies for GaAs/AlAs — where the two systems have the same anion— have suggested that the VBO does not depend on the interface orientation nor on its abruptness, and it is determined only by bulk properties. In order to check if this is the case also for a general isovalent heterojunction, I have studied also the interface InP/GaInAs,

where no common ions exist between the two materials: the investigation of this system confirms the properties suggested by the previous calculations, and also show a new aspect, i.e. the additivity of anionic and cationic effects. Conversely, calculations performed for heterovalent heterojunctions show that for this class of systems the VBO is critically affected by the interface details, but its value at polar interfaces can be obtained from the one calculated for non–polar interfaces once it is corrected by a classical electrostatic contribution easily evaluated from bulk macroscopic quantities.

ii) The linear response method —only recently applied to the study of the heterojunctions— is also based on SCF first–principles electronic calculations techniques, but relates the problem of the interface to that of the response of a crystal to the perturbation caused by a substitutional impurity. The method is here analyzed in all its details and fully developed: in particular the most suitable choice of the unperturbed reference system is discussed, together with the importance of the terms of order higher than linear both in the charge redistribution and in the VBO. The investigation of the method allows also to analyze the spatial extension of the response to a single atomic substitution. The linear–response method is proven to be a powerful tool not only to explain the general qualitative properties of lattice–matched heterojunctions, but also, if properly applied, to give quantitative results.

The thesis is organized as follows: Ch. 1 is a brief introduction to the field of the physics of semiconductor heterostructures: it contains a survey of the materials studied in this thesis, and describes the various types of heterojunctions, with brief comments about their structural characterization and growth techniques.

Ch. 2 is specifically devoted to the band–offset problem. After some preliminary definitions, I present the most widely used methods for energy–level measurements, and discuss the trends resulting from the experimental data.

A theoretical approach for a systematic study of the band–offset problem, aimed at understanding the nature of the physical phenomena occurring at the interface, is necessary. Ch. 3 highlights the general features and states a crucial problem common to any theoretical approach, even if hidden under several different formulations: i.e. the possible existence of an absolute energy scale to which the band structures of the various materials may be referred. The most significant "model" theories are then critically reviewed in this perspective. The results of some new calculations are also reported in this Chapter; they are attempt of reformulating one of the models in order to make it applicable to the more general case of lattice–matched heterojunctions. The Chapter ends with a general discussion about this kind of model theories which underlines their merits and limits, and stresses the necessity of a more accurate and systematic *ab–initio* theoretical approach, to which the remaining part of the work is devoted.

Ch. 4 explains the basic features of the self–consistent density–functional ab–initio pseudopotential method, together with some technical details for its practical implementation to the specific case of heterojunctions; the Chapter also contains a short review of some recent self–consistent calculations which are comparable to the new ones presented in this work.

The original results are mainly collected in the remaining Chapters 5 and 6. Ch. 5 reports the results of the self–consistent supercell calculations, with a description of the technical details in the first part. I study some typical heterojunctions, and more exactly one representative example for each of the following classes of heterostructures, in order of increasing complexity:

a) *isovalent common–anion*: I choose GaAs/AlAs from the III-V group, since there exists a large number of experimental data and theoretical predictions.

b) *isovalent no–common–ion*: I choose as a prototype among the III-V group $InP/Ga_{1-x}In_xAs$, which has a great technological importance, it is lattice–matched at the composition $x = 0.53$, and has been extensively studied from an experimental point of view but not theoretically;

c) *heterovalent*: I study Ge/GaAs which, among the group IV — group III-V heterojunctions, is the simplest and most studied.

Calculations are performed for the three main crystallographic orientations, i.e. (001), (110), and (111). For $InP/Ga_{0.47}In_{0.53}As$ I also study separately the anionic and cationic contribution to the VBO. This system also introduces a new problem, i.e. the treatment of the alloy: the calculations are first performed using the virtual–crystal approximation (VCA). These results are then compared with the ones obtained by considering the "true" Ga and In atoms separately in the gallium–indium arsenide region, necessarily in some ordered configuration. In particular the case $InP/(GaAs)_1(InAs)_1$ along the (001) direction is considered, since it is now possible to grow it epitaxially. For this system, I present only some preliminar study; here the presence of internal strains which are no longer negligible as in the treatment of the alloy using the VCA, makes the calculations more difficult and the theoretical study actually performed can just give an idea of the magnitude of the effects of order and internal lattice relaxation. The self–consistent results allow to draw partial conclusions, namely that the VBO is the same in the three main crystallographic orientations for isovalent heterojunctions, and the additivity of anionic and cationic effects in $InP/Ga_{0.47}In_{0.53}As$. However, whether the energy–band discontinuities are a bulk or an interface property cannot be ascertained at this point.

Finally, in Ch. 6 I describe and discuss the linear response method. The investigation with this method of the same systems studied in Ch. 5 sheds light on several questions: it appears that for isovalent heterojunctions the energy–band discontinuities are indeed a bulk property within the limit of validity of linear response theory, whereas this is not true for the heterovalent systems in which the band offset

can be divided into a "bulk" and an "interface–specific" contribution, the latter being strongly dependent on the structural details of the interface. The validity of the linear response method is also analyzed in this Chapter, by discussing the role of quadratic and higher–order terms. The thesis is concluded with a general discussion about the contribution given by the present work to the investigation of the band–offset problem and with a description of open problems for future work.

Chapter 1

SEMICONDUCTOR QUANTUM STRUCTURES: MATERIALS AND PROPERTIES

The field of the semiconductor heterostructures is wide, and continuously enriched by the increasing complexity of the structures nowadays realizable. Among the large number of new interesting phenomena occurring in these artificial systems —of which only a brief representative list is given in the Chapter— the problem of the energy-band discontinuities plays a special role in basic interface physics, since it is one of the cleanest, simplest and most well-defined interface problems. The understanding of the basic physical phenomena occurring at the interface between two different materials is of fundamental importance, and can be studied in simple systems such as isolated heterojunctions, with a great advantage for computational purposes.

This Chapter introduces the general problem of the contact between two semiconductors, with some comments about the different types of interfaces, their structural characterization, and a list of the most widely used growth techniques.

1.1 Semiconductor heterostructures: general properties .

Semiconductor quantum structures: heterojunctions, quantum wells, superlattices

The field of the semiconductor heterostructures includes many and different systems, ranging from the simple *heterojunction* constituted by two thick slabs of different materials, which presents only one interface, to the double heterojunction showing quantum potential barriers for the charge carriers (*quantum wells*), up to the repeated heterojunctions (*multiple quantum wells* and *superlattices*). The improvement in the thin–film growth techniques such as Molecular–Beam Epitaxy (MBE)[1] and Metal Organic Chemical Vapor Deposition (MOCVD)[2] has in fact made possible to realize high–quality heterojunctions characterized by sharp interfaces, and many new artificial structures[3] based on sequences of alternating different semiconductor crystalline layers of varying thickness ($10 \sim 1000$ Å).

The physics of these new materials is nowadays the subject of intensive studies both for basic and applied science. Semiconductor heterostructures show physical phenomena of quantum mechanical nature —some of them even not possible in any "natural" bulk material—, and are largely used to realize new optical and microelectronic devices based on band–gap and wavefunction engineering: quantum–well lasers and other optical sources covering a large spectral range, high–speed modulation–doped field–effect transistor, superlattice photodetectors in high–efficiency solar cells are only some examples of the variety of devices which have been demonstrated. The combination of controlled variations in the composition, possible strain, thickness of the deposited layers provides the possibility of tuning electronic and optical properties, giving to the heterostructures a great flexibility suitable for the device design. For further details see for instance Refs. 4, 5, 6, and 7 and references therein.

Contrary to what happens at the interface between two differently doped slabs of the same material (*homojunction*), in case of two different materials (*heterojunctions*) the valence– and the conduction–band edges are discontinuous because of the difference in the two forbidden gaps (see Fig. 1.1).

The 3–D physical periodicity of the bulk material is broken in these structures. In a double heterojunction the band discontinuities give rise to a square shaped potential barrier for the charge carriers, as indicated in Fig. 1.2, confining electrons and holes according to a quasi–2–D density of states; the quantized confinement energies depend on the effective mass of the charge carriers, on the width of the confining layer, as well as on the band discontinuities. In case of thick barriers the associated wavefunctions are very localized, with small tails in the adjacent layers; in this case the system is called more precisely a *quantum well* (QW).

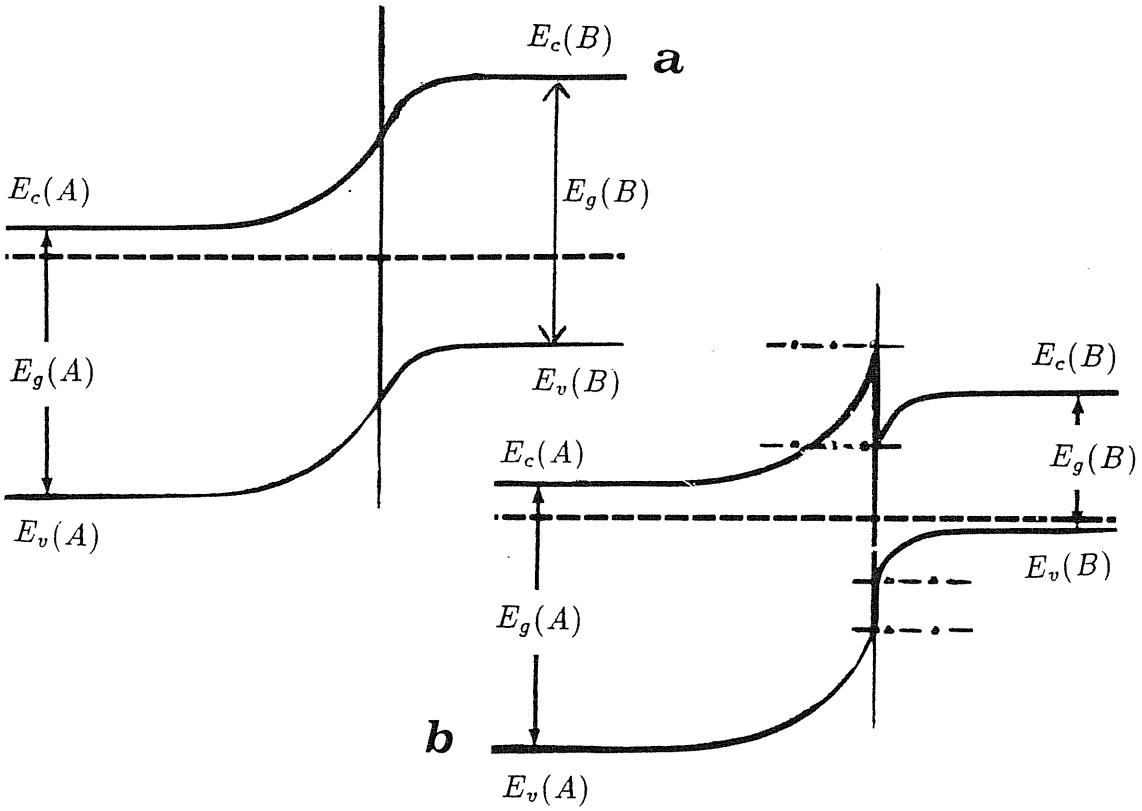

Figure 1.1. Spatial variation, along the growth direction, of the energy-band edges for homojunctions (a) and heterojunctions (b).

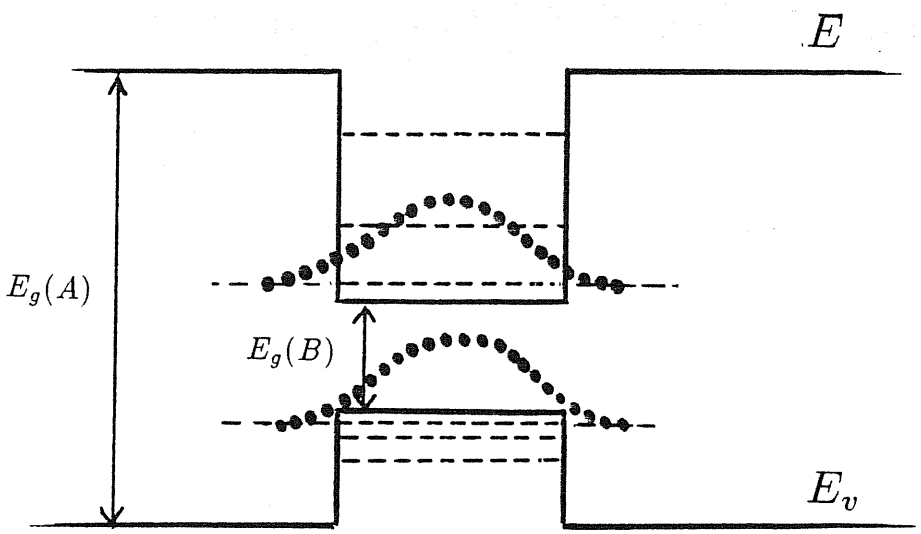

Figure 1.2. Energy-band barriers and quantized confined levels for electron and hole wavefunctions in a quantum well.

Multilayered structures, e.g. *multiple quantum wells* (MQW) and *superlattices* (SL), formed by repeated heterojunctions have different characteristics according to the length scale on which they are realized. In particular the name of *superlattices* (see Fig. 1.3) properly refers to the case in which the decay length of the wavefunction is comparable with the width of the barrier, such that tunneling phenomena can occur and the quantized levels spread into minibands.

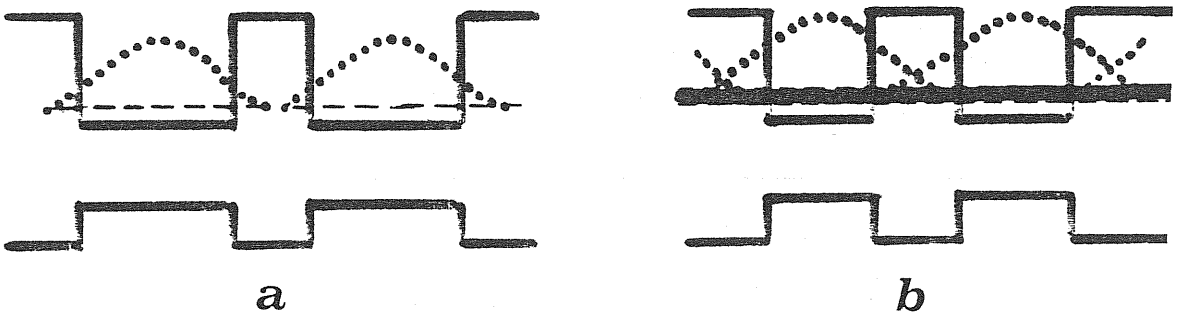

Figure 1.3. Spatial variation, along the growth direction, of the energy–band edges for a MQW (a) and a SL (b). Electronic quantized confined levels and minibands are indicated in (a) and (b) respectively, together with the corresponding wavefunctions.

In the semiconductor heterostructures two aspects are equally very important, i.e. the electron behaviour both in the growth direction (determining the transport properties), and in the potential wells parallel to the interface planes. Important parameters are the penetration depth of the electron wavefunction, the mean free path, the screening length.

Space–charge effects can modify the band edge structures that we have seen in the previous structures. The typical situation is that of a modulation doped GaAs/GaAlAs SL, where the free electrons from the ionized donors in the AlGaAs barriers are transferred to the undoped GaAs wells (see Fig. 1.4), where they form a high–mobility (of the order of 10^6 cm^2/Vs) 2–D electron gas, with interesting quantum properties (e.g. Quantum Hall Effect[8]). The accumulation of opposite charges on the two sides of the interface, and the resulting alternating sign of the charge distribution, causes a curvature of the band edges. The charge redistribution and the band bending is typical of long–length scales (isolated interfaces as in the heterojunctions or QW with thick layers), whereas it does not occur in those multilayered structures where the layer thickness is smaller or comparable to the screening length of the semiconductor. The possibility of realizing quasi–2–D systems of electrons and holes thus holds not only in multilayered structures but also at the simple heterojunction, because of the roughly triangular potential well binding the charge carriers (see Fig. 1.5).

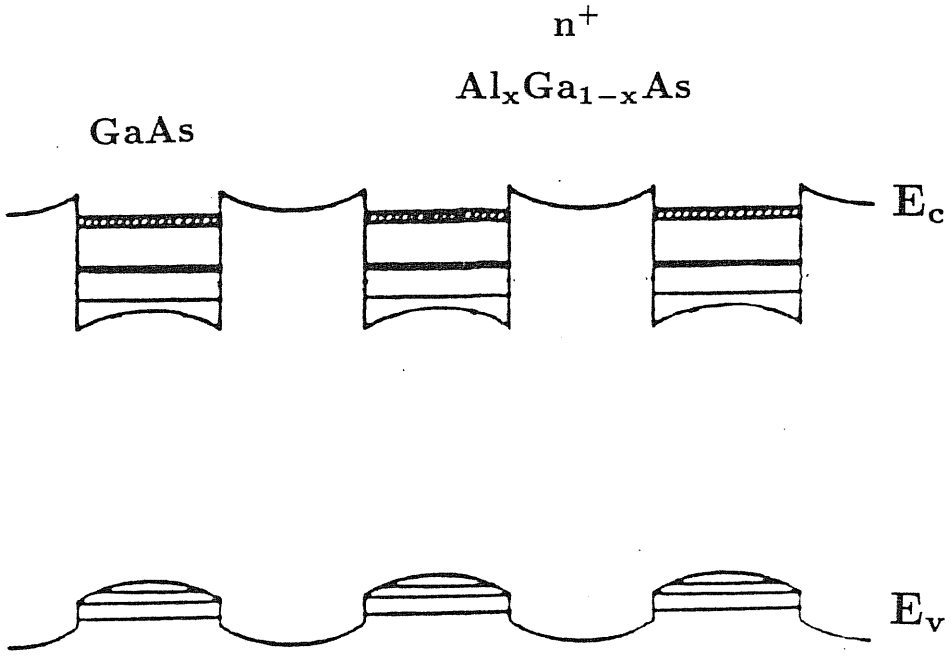

Figure 1.4. Spatial variation of the energy–band edges in GaAs/GaAlAs doped SL. The band bending results from the spatial separation of electrons from the ionized donors in the GaAlAs barriers.

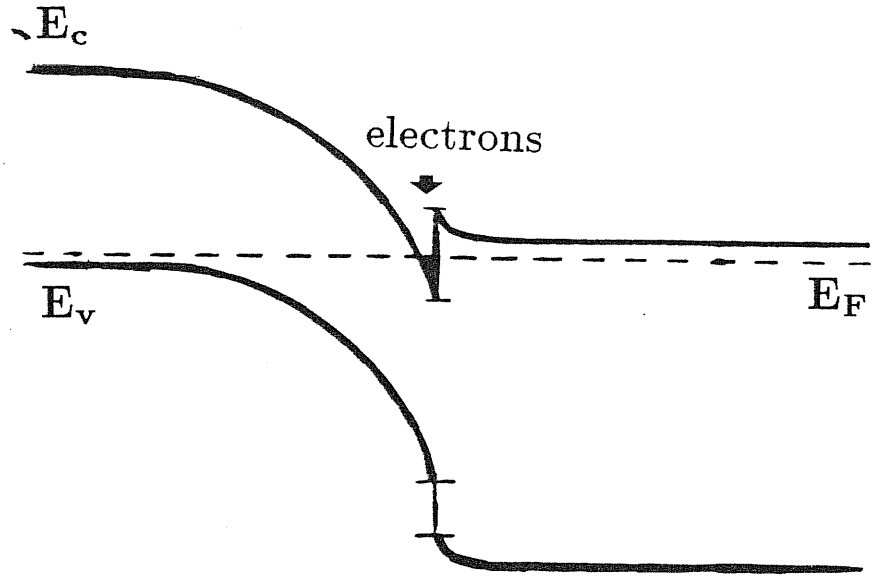

Figure 1.5. Band–bending due to space–charge effects at an heterointerface. The case of a quasi–triangular shaped potential well binding the electrons is shown.

1.2 Semiconductor interfaces: structural characterization and their classification

Classification of the semiconductor interfaces

It is useful to classify the semiconductor interfaces according to their structural or chemical characteristics. The most important parameter which determines the structure of the heterojunctions is of course the lattice constant of the two different semiconductors (see Tab. 1.1). A great number of technologically important heterojunctions are characterized by nearly identical lattice parameters and consequently by a particularly simple and ideal structure, usually referred to as "lattice–matched" heterojunctions. In particular, if all the bond lengths are the same, the heterojunction has a perfectly ideal structure and the presence of the interface is characterized only by a change in the atomic species filling the nodes of the crystal lattice, and by no distortions. The prototype case is GaAs/AlAs, whose experimental room–temperature lattice parameters are respectively 5.653 Å and 5.660 Å, i.e. they differ only by 0.2 %. A number of other important lattice–matched heterojunctions exists.

	a Å	Si	Ge	GaAs	AlAs	InAs	GaP	AlP	InP	GaSb	AlSb	InSb	ZnSe	ZnS	ZnTe	CdTe	HgTe
Si	5.43	0															
Ge	5.65	4	0														
GaAs	5.65	4	0	0													
AlAs	5.65	4	0	0	0												
InAs	6.08	11	7	7	7	0											
GaP	5.43	0	4	4	4	11	0										
AlP	5.43	0	4	4	4	11	0	0									
InP	5.87	8	4	4	4	3	8	8	0								
GaSb	6.08	11	7	7	7	0	11	11	3	0							
AlSb	6.08	11	7	7	7	0	11	11	3	0	0						
InSb	6.48	18	14	14	14	6	18	18	10	6	6	0					
ZnSe	5.65	4	0	0	0	7	4	4	4	7	7	14	0				
ZnS	5.40	1	4	4	4	12	1	1	8	12	12	18	4	0			
ZnTe	6.08	11	7	7	7	0	11	11	3	0	0	6	7	12	0		
CdTe	6.48	18	14	14	14	6	18	18	10	6	6	0	14	18	6	0	
HgTe	6.48	18	14	14	14	6	18	18	10	6	6	0	14	18	6	0	0

Table 1.1. Lattice parameters and relative mismatching percentage for various semiconductors with bulk diamond or zincblende structure in standard conditions of temperature and pressure. The lattice parameters are given in Å, the mismatch in units %.

The case of "lattice–mismatched" heterojunctions, whose typical example is Ge/Si, is more complicated: given a substrate, the material grown on it modifies its structure and adjusts itself to the substrate. In particular, the deposited material can assume along planes parallel to the growth surface the same structure of the substrate lattice, with stretching of the chemical bonds and a completely different structure with respect to the normal crystal; the lattice mismatch is thus accommodated with a uniform lattice strain in case of sufficiently thin layers. The typical diamond or zincblende structure of semiconductors is lost, and a distinction is usually made between the "in–plane" lattice parameter $a_{\parallel}$, that is the average lattice parameter along planes parallel to the interfacial plane, and the "inter–plane" lattice parameter $a_{\perp}$ (see Fig. 1.6): typically, $a_{\parallel}$ accommodates in order to be equal to the lattice parameter of the substrate, and $a_{\perp}$ is determined by the elastic properties of the material.(*)

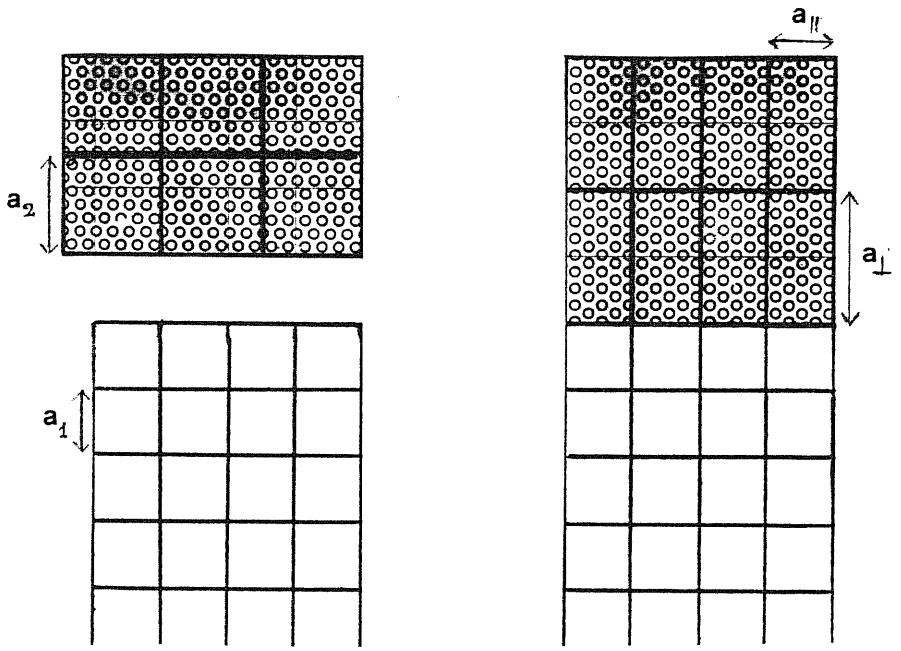

Figure 1.6. Interface formation between lattice–mismatched semiconductors.

(*) The flexibility of the new techniques for preparing interfaces has improved the possibility of producing complex strongly lattice–mismatched heterostructures, i.e. strained–layer superlattices (SLSL), with an accurate control during the growth in order to avoid defects such as misfit dislocations. It is important to notice here the great flexibility of such structures: in general, if a thin overlayer is grown on a substrate, the value of the "in–plane" lattice parameter is determined by the substrate and can be varied using different substrates. Practical growth of SLSL with dislocation–free interfaces is actually possible with layers under a certain critical thickness depending on the degree of lattice mismatch. New interesting features in the band edges —i.e. energy–band gap splittings and shifts studied for instance in Ref. 9— correspond to the increased structural complexity.

We can further distinguish "isovalent" and "heterovalent" interfaces, as indicated in Fig. 1.7. "Isovalent" heterojunctions are constituted by two elemental, or two binary or pseudobinary semiconductor compounds pertaining to the same group (both to group IV, or III-V, or II-VI): isovalent meaning that anions (cations) in the two materials have the same chemical valence. In particular, the isovalent "common–anion" heterojunctions are constituted by two binary or pseudobinary semiconductor compounds having the same anions, a typical system being GaAs/AlAs. The isovalent "no–common–ion" heterojunctions are constituted by two binary or pseudobinary semiconductors having both different anions and cations; in this class there are also heterojunctions, like for instance InAs/GaSb and InP/$Ga_{1-x}In_xAs$.

Conversely, the "heterovalent" heterojunctions are constituted by two semiconductors which belong to two different groups among the IV, III-V, or II-VI—group, the simplest and most studied case being Ge/GaAs. As already emphasized by other authors [10] such heterojunctions can present polar unstable interfaces with charge accumulation and unscreened electric fields; the interface is more complicated in such cases, and its properties are determined by e.g. the stoichiometry (i.e. the number of atoms of each type in the interface region), the order of the different types of atoms in the interface plane, and finally the displacements of the atoms from their ideal positions.

Interface formation:
growth techniques and structural characterization

The possibility of growing heterojunctions with good crystal quality is a quite recent technological achievement. Parameters which can favour one technique with respect to the others —at equal quality of the interface— are the background doping and the possibility of controlling the thickness of the deposited layer. One of the techniques first used is the liquid–phase epitaxy (LPE), introduce by Nelson [11] at the beginning of sixties and successively improved by Panish [12]. More recent techniques, largely used nowadays, are the metal–organic chemical vapor deposition (MOCVD) [2] and the molecular beam epitaxy (MBE) [1].

Whatever growth technique is chosen, the final quality of the interface can be affected by several factors, as for instance: i) the quality of the starting substrate, which is prepared with chemical polishing, vacuum annealing, and other surface techniques to avoid the effects of surface contamination on the grown crystal [13]; ii) the growing time, which determines the evolution and the final microscopical chemical structure of the interface [4]; iii) the substrate temperature during growth, which is very important to control the atomic interdiffusion: for instance it has been observed [14] that an abrupt interface Ge/GaAs with an epitaxial Ge overlayer is obtained when Ge is deposited over a GaAs substrate at 320°–360°C, whereas an amorphous Ge overlayer is obtained if the substrate is at room temperature, and a diffused interface is formed at higher temperatures; this observation also indicates

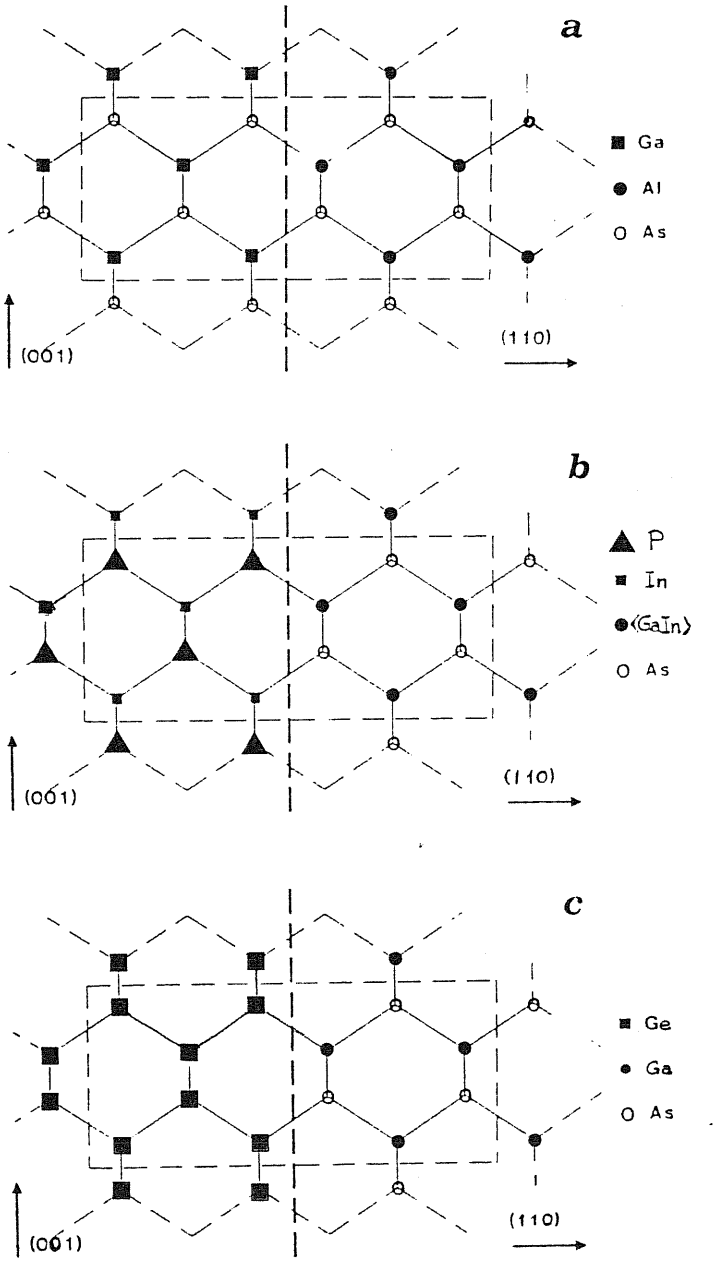

Figure 1.7. Examples of isovalent common–cation (a), isovalent no–common–ion (a), and heterovalent (c) lattice–matched heterojunctions.

that for two semiconductors to be potential constituents of an heterojunction, their temperature ranges of the crystalline phase must overlap.

Although a good improvement in the preparation techniques has been done, several deviations from the "ideal" character of the interface are possible, which may influence the actual charge distribution: chemiabsorption–bond dipoles, microdiffusion of atoms across the interface, the presence of defects and/or contaminants at the interface. It is important to point out the possible existence of such microscopic defects, and the fact that a complete knowledge of the actual structural situation at the interface is beyond our capabilities. Some progress in

this sense —to provide direct information about the microscopic structural and chemical properties of the interface also at the initial stage of formation— has already been done, with the combined efforts of some surface–sensitive techniques such as soft X–ray photoemission spectroscopy (SXPS), and other very recent techniques like transmission electron microscopy (TEM, see Fig. 1.8)[15], high–resolution X–ray diffraction (HRXRD), Rutherford backscattering, reflected high–energy electron diffraction (RHEED), ion channeling and ellipsometry.

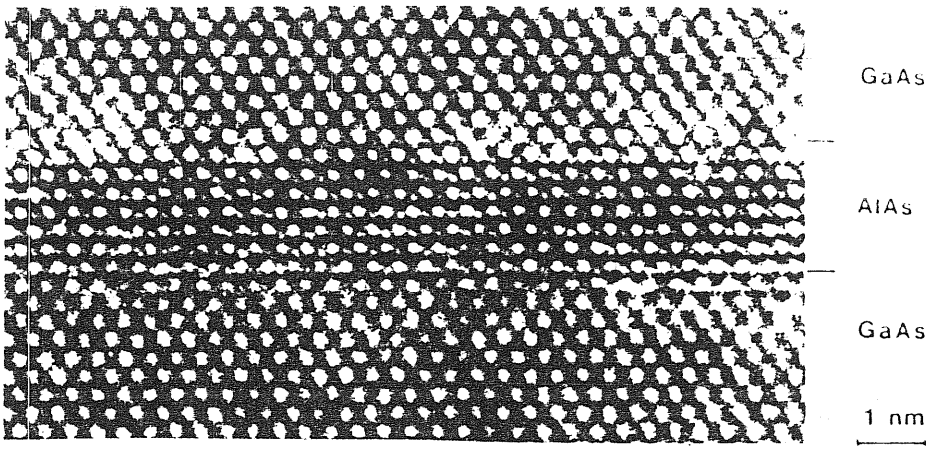

Figure 1.8. High–resolution TEM micrograph of GaAs/AlAs SL projected on (110) plane (from the cited Reference (b)).

Chapter 2

THE BAND–OFFSET PROBLEM: TRENDS FROM EXPERIMENTAL DATA

The band-offset problem —which I introduce at the beginning of the present Chapter— is of fundamental interest both for experimentalists and theoreticians, and combined efforts are necessary to its understanding. It is beyond the purpose of the present work to make a detailed and critical analysis of the experimental measurement techniques, and a collection of the existing literature: few lines are devoted to the most representative and nowadays largely used techniques.

The attention is rather focused on the general trends outcoming from the large amount of experimental data available, and in particular to what can be concluded about the main questions related to the lineup problem: validity of commutativity and transitivity rules, role of the interface structure. Experimental data show that for a wide class of systems (i.e. mainly isovalent heterojunctions) the band discontinuities show commutativity and transitivity properties, and do not depend on the interface orientation; whether other interface details, such as abruptness, dislocations, impurities... are important or not, it cannot be precisely ascertained by the existing experimental data.

2.1 Energy–band discontinuities

The band lineup, a fundamental parameter for any heterostructure.

The present study concerns the energy–band discontinuities, which are important parameters both for the isolated interface and for the repeated ones in the multilayered structures. The difference $\Delta E_g = E_g^B - E_g^A$ of the minimum forbidden gap in the two semiconductors is shared between the valence–band offset (VBO) and the conduction–band offset (CBO), such that VBO+CBO=ΔE_g. In the literature —regarding both experimental measurements and theoretical calculations— are reported either the values of VBO and CBO or their ratio, which also provides their values once ΔE_g is known. For computational reasons, since it is not trivial to calculate accurately the minimum forbidden gap, I refer in this work to the two band–discontinuities separately; in particular I focus the attention on the VBO, which can be calculated more accurately than the CBO. The various parameters and symbols are defined once and for all in Fig. 2.1; conventionally I consider *positive* a band discontinuity in the heterojunction A/B if the energy–band edge of semiconductor B is higher than the corresponding of semiconductor A.

The energy–band discontinuities and the related bending at the two sides of the interfaces can form potential barriers V_A and V_B for the charge carriers crossing the interface, playing a different role for electrons and holes. The interest is further restricted here to what happens in a limited region near the interface (1–10 Å), neglecting the band bending effect which requires a study of systems over a larger scale (10^2–10^4 Å).

The importance of the band lineup problem is twofold: on one side an accurate knowledge of the band offset values and the possibility of predicting them for a large number of systems is appreciable for technological purposes; on the other hand the description of what happens when two different semiconductors are put together is interesting in itself, being related to some fundamental questions, such as the mechanism which governs the band alignment or the possibility of referring the band structures of all semiconductors to a common energy scale.

Band discontinuities and interface characterization

The relative position of the energy bands at an interface between two semiconductors can vary dramatically by substituting the components of the heterojunction, as shown in Fig. 2.2 for four different possible situations.

The alignment of the energy bands may depend on both intrinsic and extrinsic, structural and chemical characteristics of the two semiconductors. Bulk lattice constants and bandgaps are examples of intrinsic factors, while extrinsic factors include, among others, chemical impurities, structural defects, atomic interdiffusion of the two materials.

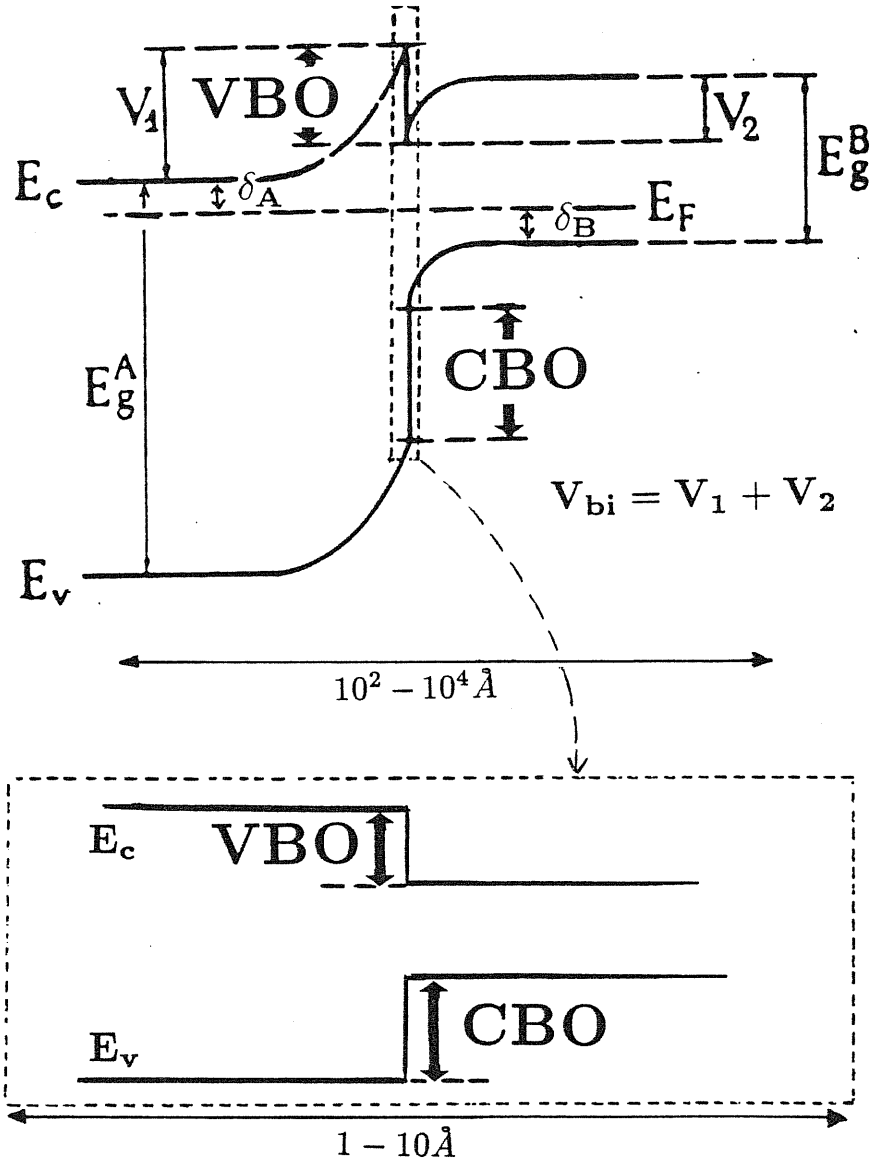

Figure 2.1. Schematic diagram of the energy bands at semiconductor–semiconductor heterojunction. E_F is the Fermi level; for each semiconductor: E_v, E_c, E_g are the valence–band, the conduction–band and the the band gap; the band–gap difference between the two semiconductors is shared between the valence–band (VBO) and the conduction–band discontinuities (CBO); V_{bi} the built–in potential barrier. The spatial range to which the calculations reviewed and the new ones presented in this work are referred is indicated in the magnified picture.

All these microscopic factors determine the charge distribution in the heterojunction and therefore can in principle influence the band lineup. It is essential to know which factors are really important and how much, i.e. whether the VBO is determined and can be thus explained and predicted only using the different bulk properties of the two materials, or if it is determined by specific interface details, and can be varied also for a fixed couple of semiconductors. In summary, there are three main questions to be investigated, as presented in Fig. 2.3:

i) the *commutativity*, i.e. whether $\mathrm{VBO}_{A/B} = -\mathrm{VBO}_{B/A}$. The expected answer may seem trivial, but the experimental test has to be performed following two

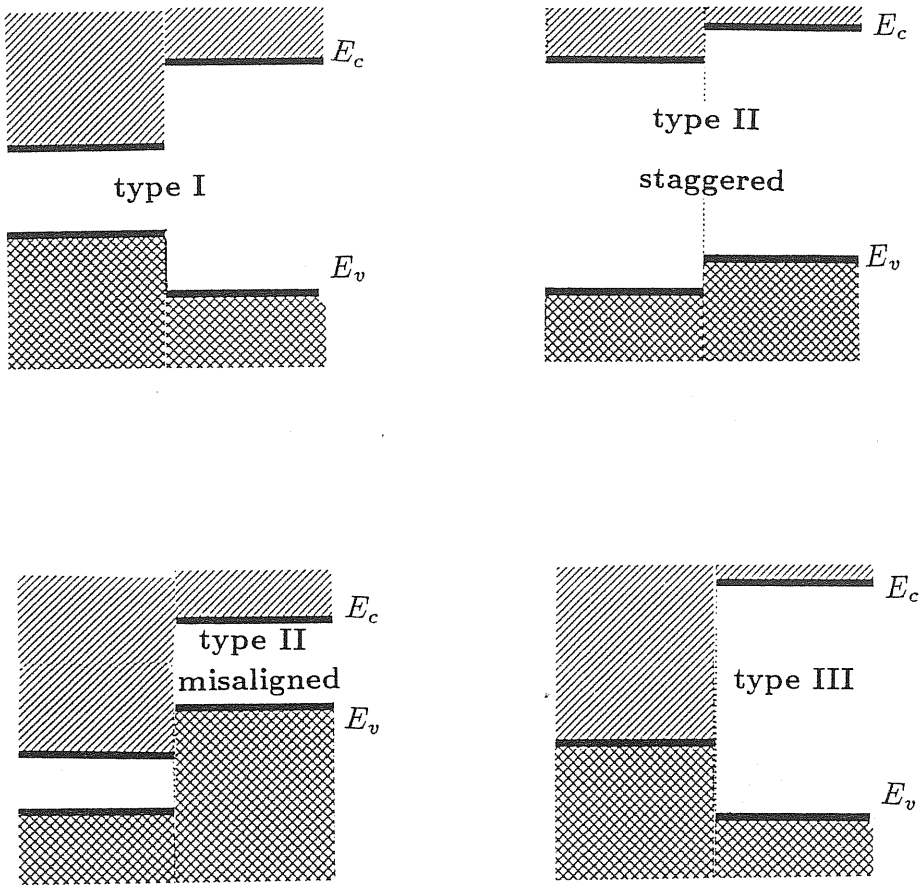

Figure 2.2. Energy–band discontinuities at different type of heterojunction interfaces.

different interface growth processes, starting with one semiconductor or the other as the substrate.

ii) the *transitivity* rule, i.e. whether in case of three heterojunctions A/B, B/C, and A/C, it occurs that

$$VBO_{A/B} + VBO_{B/C} \approx VBO_{A/C} \tag{2.1}$$

Note that if the VBO were a linear function of a given physical quantity ξ for the two semiconductors, i.e. VBO$= \xi_B - \xi_A$, then commutativity and transitivity would be consequences of that; viceversa, if such properties hold, this would suggest that the VBO is related to bulk quantities.

iii) the role of the *interface details*, and in particular whether the VBO does depend or not on the interface *orientation* and *abruptness*. Note that if the microscopic

interface details influence the band discontinuities, possible deviations from the "ideal interface" configuration would constitute a powerful tool to control and tune the band discontinuities according to the technological requests. The answer to these crucial questions is related to the understanding of the nature of the mechanism determining the band lineup.

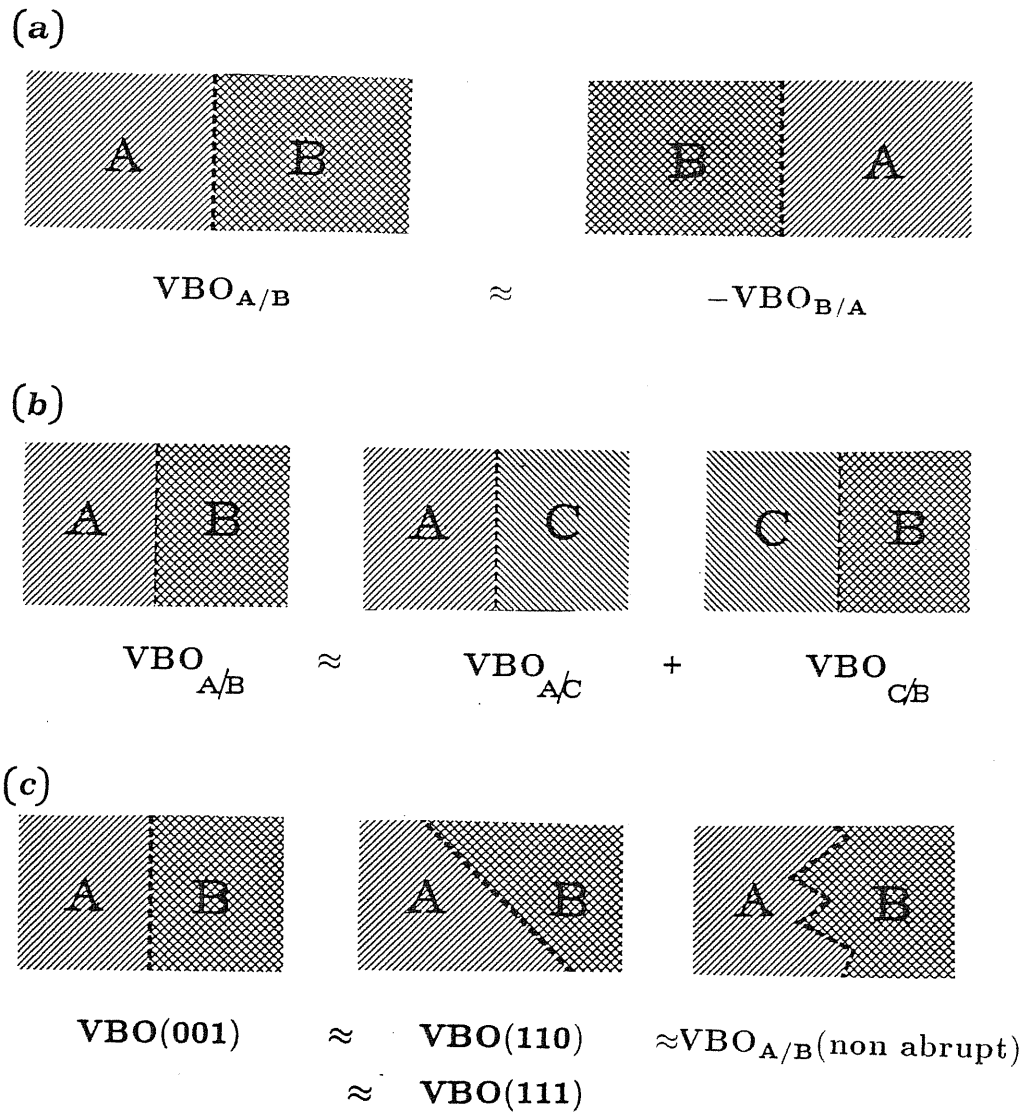

Figure 2.3. Schematization of the three main questions concerning the band-offset problem: commutativity (a), transitivity (b), role of interface orientation and abruptness (c).

2.2 Energy–band discontinuity measurements

Photoemission measurements

I give here a brief survey of some experimental techniques for band discontinuities measurements. More details can be found in Ref. 5 and references therein.

The photoemission spectroscopy is a very sensitive method for the band discontinuity measurements. The system, constituted by the substrate and a thin deposited overlayer, emits photoelectrons when bombarded with soft X–ray (XPS, X–ray photoemission spectroscopy) or ultraviolet photons, obtained from conventional or syncrotron radiation sources. A fundamental feature which makes the photoemission techniques particularly suitable for interface studies is the short mean–free path of the photoelectrons inside the two semiconductors, which is of course related to their kinetic energy: a kinetic energy of 50–150 eV corresponds to a minimum in the mean–free path less than 1 nm, so that detected photoelectrons with that energy come from the thin overlayer and therefore from a region very close to the interface.

The measured energy distribution of the photoelectrons roughly corresponds to the local density of states at the interface. A direct measurement of the energy–band discontinuity is often possible: if the VBO is not too small, its presence is directly revealed by a double–edge structure in the photoelectron spectrum, as indicated in Fig. 2.4.

If this is not the case, a less direct but more general method must be applied, by measuring the distance between the two valence band edges relative to the distance between core levels on each side of the interface (see Fig. 2.5).

The photoemission spectroscopy is the most suitable technique to monitor the evolution of the interface formation: from valence–band spectra taken from the covered–surface at consecutive overlayer thickness, one is able to observe at what coverage the overlayer semiconductor electronic features are fully developed, i.e. what is the extension of the interface effects. Two examples are shown in fig. 2.6: in both cases we observe that the bulk features of the overlayer semiconductor are completely recovered within a distance of 5–10 Å from the interface. Unfortunately, the high sensitivity of the photoemission measurements to the interface details does not correspond to a very good accuracy: the estimated accuracy is in general about 0.1 eV.

Other optical measurements

The optical properties of QW and SL allow alternative methods of measuring the energy–band discontinuities. The VBO and the CBO are two of the several parameters determining the quantized energy levels for electrons and holes, revealed by the presence of a particular series of peaks both in the infrared absorption

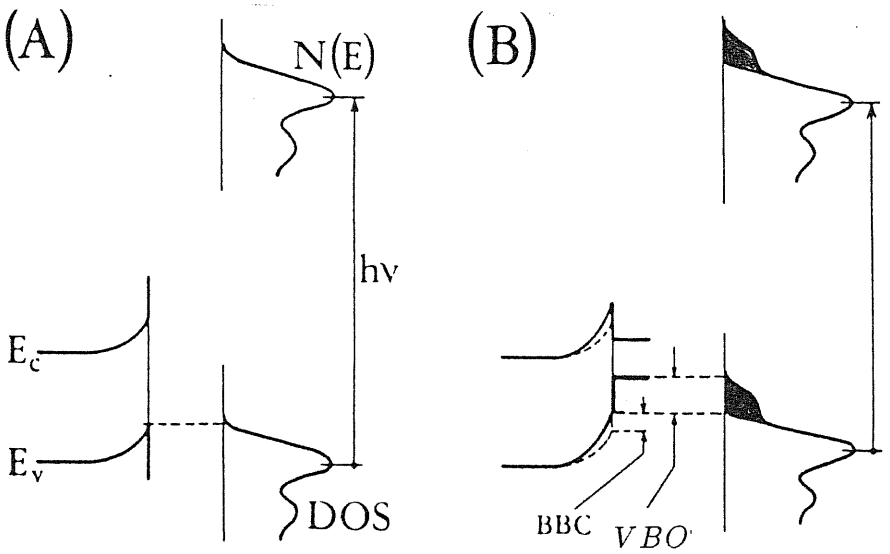

Figure 2.4. The two steps of a photoemission measurement of the band discontinuity: on the clean surface of the substrate (a), and when the overlayer has been deposited (b). After absorption of photons with energy $h\nu$, photoelectrons are created whose energy spectrum N(E) reflects the density of states of the substrate (a) or of the interface region (b), shifted in energy of an amount equal to $h\nu$ in both cases plus the band bending, different in presence of surface or interface. From the double–edge structure of the second spectrum the VBO can be directly measured; if it is not clearly visible, it is necessary to compare the positions of the leading edge of N(E) in (a) and (b), corrected with the different band bending. From Ref. 16.

spectra[17] and in the photoluminescence spectra[18]. Unfortunately the energy levels are not very sensitive to the band discontinuities,(*) although the experimental accuracy reachable in optical experiments is very high.

Transport measurements

Indirect measurements of the valence–band discontinuity are possible by estimating the built–in potential barrier, which is related to the VBO through the band gap of one semiconductor and the bulk Fermi level positions as shown in Fig. 2.1 and according the following equation:

$$VBO + V_{bi} + \delta_A + \delta_B = E_g^A \tag{2.1}$$

The built–in potential V_{bi} is usually determined by studying the capacitance–voltage (CV) characteristics[19]. If one reasonably assumes the relationship:

$$C \propto (V_{bi} - V)^{-1/2} \tag{2.2}$$

(*) The quantized levels of the charge carriers depend more strongly on the width of the potential well —and are therefore very sensitive to the thickness of the alternating layers— rather than on the band discontinuities.

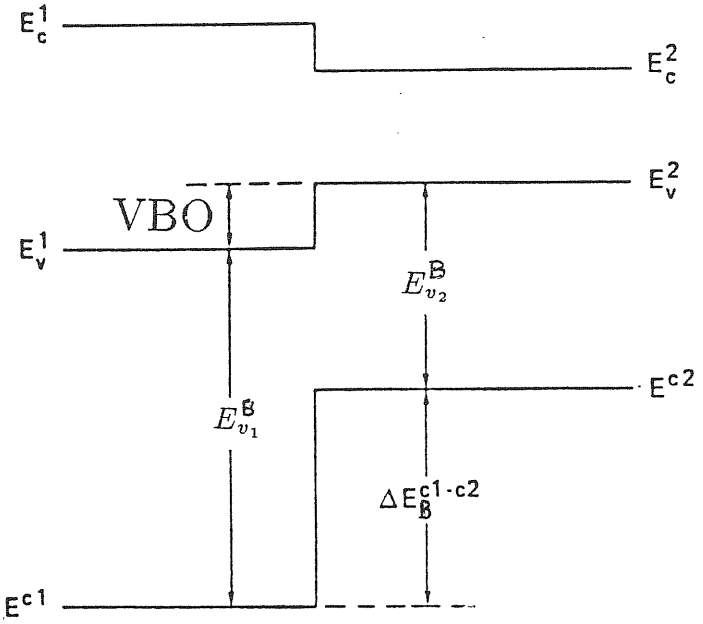

Figure 2.5. Schematic band structure at a semiconductor–semiconductor heterojunction, showing the distances E_v^B between the valence band edges and the position of the core levels.

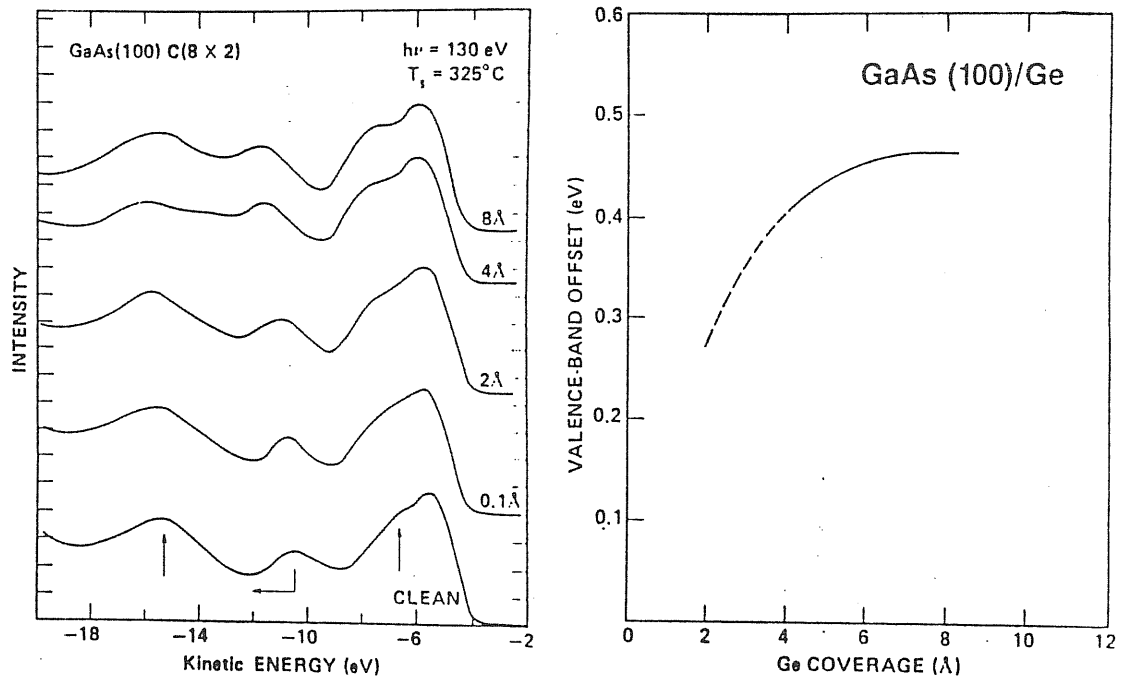

Figure 2.6. (a): Evolution of the valence–band spectra of GaAs with increasing Ge coverage taken at $h\nu$ =130 eV (from Ref. 5, p. 127). (b): Evolution of the positions of the valence–band edges of CdS and Si relative to the Cd 4d core levels as a function of the thickness of Si overlayer (from Ref. 4).

then a plot of C^{-2} versus V gives a straight line which intercepts the V–axis at V_{bi}. An alternative but in general less applicable method is the measurement of the current–voltage (IV) characteristics [20].

Admittance spectroscopy measurements

The admittance spectroscopy method also provides the determination of the band discontinuities through the determination of the built–in potential which limits the perpendicular transport at the interface, but it is different since it is a dynamic technique. In fact it is based on the thermal activation of carriers over the potential barrier formed by the heterojunction band offsets: a plot of the *ac* conductance versus the temperature shows a peak corresponding to the situation when a typical time RC of the junction (R is the series resistance of the sample, C the depletion layer capacitance) equals the inverse of the angular measurement frequency.

2.3 Trends and open questions

Experimental data base

A large number of reliable data is avaliable now, thanks to the continuous progress in the experimental techniques. As an example, I report in Tab. 2.1 the more reliable experimental values of the VBO from measurements obtained with different methods for the heterostructures studied in the present work, but the following discussion takes into account the existing data base for several other systems.

	$VBO(eV)$	*technique*	*authors*	*comments*
$AlAs/GaAs$	0.45	*charge transfer*	*Wang and Stern* (1985)	MBE, (001)
	0.40	*photoemission*	*Waldrop et al.* (1981)	*GaAs on AlAs* (110)
	0.15	*photoemission*	*Waldrop et al.* (1981)	*AlAs on GaAs* (110)
	0.38	*photoemission*	*Katnani and Bauer* (1986)	*commutativity*, (001)
$GaAs/Ge$	0.48	*photoemission*	*Waldrop et al.* (1982)	*Ge on GaAs*(111)*Ga*
	0.55	*photoemission*	*Waldrop et al.* (1982)	*Ge on GaAs*(001)*Ga*
	0.56	*photoemission*	*Waldrop et al.* (1982)	*Ge on GaAs*(110)
	0.60	*photoemission*	*Waldrop et al.* (1982)	*Ge on GaAs*$(\overline{100})$*As*
	0.60	*photoemission*	*Waldrop et al.* (1982)	*Ge on GaAs*$(\overline{111})$*As*
$InP/Ga_{0.47}In_{0.53}As$	0.37	$C-V$	*Forrest et al.* (1984)	
	0.24	*photoluminescence*	*Brunemeier et al.* (1985)	QW, (001)
	0.26	*photoluminescence*	*Temkin et al.* (1985)	QW *and* MQW *grown by* MBE
	0.00	$C-V$	*Steiner et al.* (1986)	
	0.38	*optical spectroscopy*	*Skolnick et al.* (1986)	QW *grown by* $MOCVD$
	0.33	*optical abs., photoluminescence*	*Westland et al.* (1987)	MQW *grown by* $MOCVD$
	0.35	*admittance spectroscopy*	*Lang et al.* (1987)	(001)

Table 2.1. VBO (in eV) for the different lattice–matched heterostructures studied in the present work. See for instance Ref. 21 for GaAs/AlAs, Refs. 22, 23, 24, and 25 for InP/GaInAs, Ref. 26 for Ge/GaAs; others in the review article by G. Margaritondo and P. Perfetti in Ref. 5.

It is not easy to estimate the accuracy and the reliability of the VBO measurements, since different authors are often in contrast on this point, and the spread of the data obtained with different techniques for the same material is rather large with respect to the quoted accuracies of the single measurements. A significant example is the one of $Al_xGa_{1-x}As/GaAs$: for a long time, say from the early seventies, it has been accepted the result of infrared measurements[17] giving a ratio of 15/85% for VBO/CBO and thus supporting the "common–anion rule" which predicts a very small VBO for heterojunctions with the same anion; other more recent optical data[18] as well as photoemission and other measurements have shown that the band–gap discontinuity is practically shared with the same weight between the valence– and the conduction–band discontinuities.

We should have in mind that the experimental accuracy achieved in measuring the quantities directly accessible is different in general with respect to the final accuracy of the estimate of VBO and CBO; the final result is obtained with some model to link the directly measurable quantities to the band discontinuities, and

hence it can be affected by some spurious contributions. According to some of the best experts,[5] the typical accuracy of the experimental measurements of the VBO is nowadays in general of the order of 0.15 eV; this does not exclude a better accuracy in some particular case.

As far as the reliability of these experimental data is concerned, it should be stressed that measurements performed on low–quality heterojunctions can give values of VBO which are not representative of any ideal abrupt epitaxial interface: many samples have interfaces which are not at all ideal and abrupt, but rather graded, or with a strong degree of atomic interdiffusion, presence of foreign atoms,epitaxial interface, epitaxial interface, internal lattice distortions and interfacial strains. A certain spread of the experimental data can thus be explained by the deviations from the abruptness, which are hardly detectable and can in principle strongly influence the energy–band discontinuities.

Experimental measurements and theoretical works

Which general features can be extracted from the large amount of experimental measurements? In particular, what do experiments say about commutativity, transitivity, and role of the interface details? *Commutativity* rule is in general satisfied: experimental tests performed for GaAs/AlAs and ZnSe/Ge[27,28] have shown that the deviations from it are in general negligible, and however always smaller than the accuracy of the single measurements. There is some evidence of *transitivity* for a lot of systems; tests[29] performed mainly for the group (GaAs, Ge, AlAs)[28] have given deviations from it of the order of 0.01–0.03 eV per interface, whereas other groups such as (GaAs, Ge, CuBr) show significant deviations from it ($\approx$ 0.2 eV per interface).[30] As far as the *interface orientation dependence* of the VBO is concerned, results are in contrast on this point, ranging for different systems from the case of GaAs/AlAs where the differences of the VBO in the three main crystallographic orientations are negligible to the case of heterovalent interfaces such as Ge/GaAs where the result strongly depend on the interface orientation (see Tab. 2.1). Again, there is an evidence of *linear behaviour* of the VBO in those systems including an alloy, such as $Ga_{1-x}Al_xAs/GaAs$ (see Fig. 2.7): the VBO is found to be a linear function of the composition x.[31]

In summary, experimental measurements suggest that the energy–band discontinuity between two semiconductors can be expressed with a good approximation as a linear function of some bulk quantities for a large class of systems, mainly for isovalent heterojunctions, whereas deviations from commutativity and transitivity tests of the order of 0.2 eV —and however surely greater than the experimental accuracy— are shown in other systems.

What can the experiments say about the importance of other interface details, i.e. the possible deviations from the "ideal interface" configuration? The existing experiments alone cannot ascertain this crucial question, i.e. whether these details

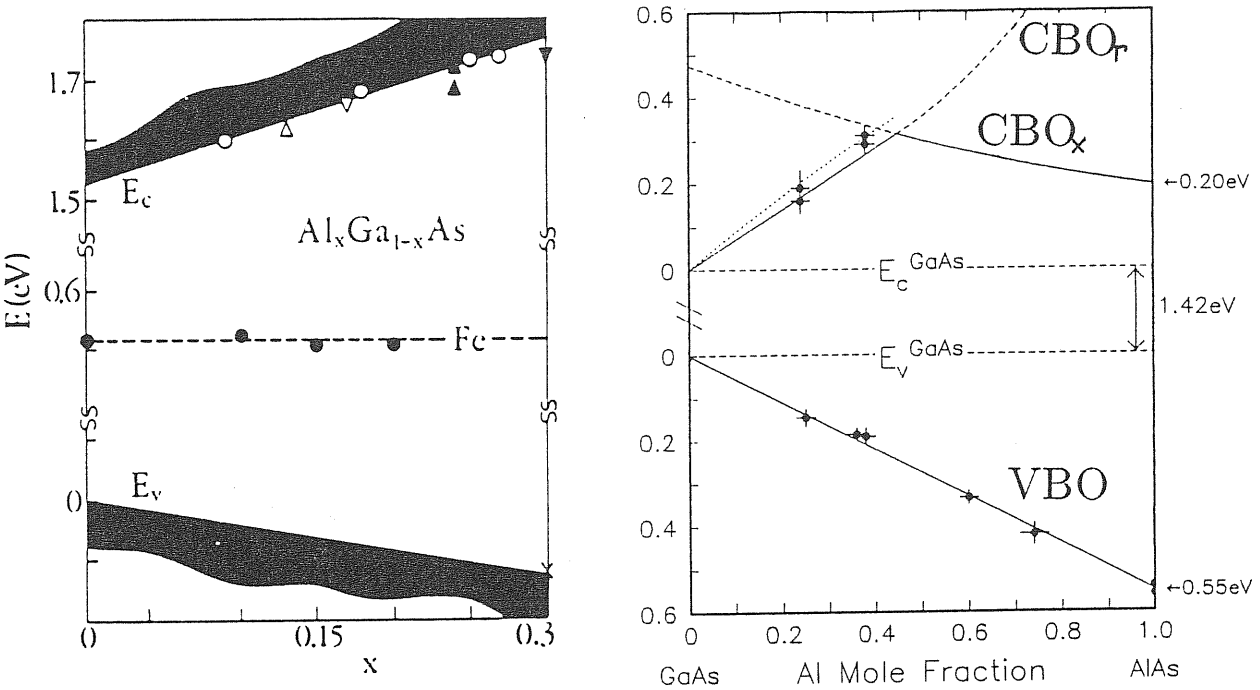

Figure 2.7. (a) Position in energy of the band edges of $Ga_{1-x}Al_xAs$ with respect to the Fe(2+) deep impurity level, as a function of the composition parameter x. From these results, a linear behaviour of VBO and CBO as a function of x and a CBO of 64% of the band-gap difference is given for $Ga_{1-x}Al_xAs$/GaAs interfaces (from Ref. 5, p. 86); (b) Energy-band alignment for the same heterojunction, proposed by J. Batey and S.L. Wright [31] (dots indicate experimental data).

influence or not the VBO: the knowledge of the microscopic structural and chemical details at the interface is still beyond our capabilities, so that we neither know to which specific interface configuration some measurements are referred, nor we can predict it with calculations; moreover the accuracy of the available experimental data is often comparable to the effects one is looking at.

However some experiments indicate that the band discontinuities can be modified by acting on the interface chemically. Some attempts of modifying the VBO by a controlled contamination of the interface have been done by doping the interface region [32] or introducing an entire monolayer of foreign atoms [33,34]: these doping effects have no influence at all in cases such as GaAs/Ge, whereas produce very large changes in band lineups ($\approx$ 0.5 eV for the VBO) for instance in case of Si/SiO_2 contaminated with hydrogen; a change up to several tenths of an eV has been recently found in GaP/Si and in ZnSe/Ge interfaces modified with the introduction of Au-diffusive intralayers. [35]

In summary, a clear and complete understanding of the nature of the energy-band discontinuities is not yet achieved, and many questions are unanswered. A theoretical approach to the band-offset problem has become and is still necessary not only to discriminate between the different experimental data and to determine

the reliability of the measurements,(*) but expecially to give a general systematic understanding of the physical mechanisms responsible of the band offset. Moreover, theoretical calculations —based on certain geometrical and structural assumptions— also could provide predictions for heterojunctions not yet experimentally grown or not well characterized, and would be very useful to focus new physical properties, giving perspectives of future applications.

The physical understanding of the band lineup problem is the crucial question, but a lot of efforts have still to be done both from experimentalists and theoreticians also to strongly improve the accuracy of the values obtained: for technological purposes an accuracy of the order of 1 meV is in fact necessary.

As a final challenge, there is some evidence that reducing the interface problem to a one–dimensional problem is too much simplified: some experimentalists are trying to explore the interfaces, with the help of the new syncrotron radiation technologies, in their complete three–dimensional structure, without any spatial average, and to understand if they are, as many of them seem to be, laterally inhomogeneous.

(*) However, the main goal of calculations could not be in any case to reproduce the experimental data with excellent agreement, since a realistic description of the interface is not yet possible even using the most sophisticated and most accurate theories available at present.

Chapter 3

THE BAND–OFFSET PROBLEM: THEORETICAL FORMULATION AND CRITICAL REVIEW OF EXISTING MODELS

A theoretical approach to the band–offset problem is necessary to answer questions that cannot be ascertained by experiments only. In principle one might expect that, being the heterojunction problem well defined, accurate theoretical calculations of the band discontinuities are quite simple to be carried out. This is not the case for the many uncertainties about the interface structure, but even if one assumes to know all the crystallographic and stoichiometric details needed and is able to make accurate calculations, the understanding of the problem does not find immediate solution, and each possible theoretical approach gives a partial contribution.

Theoretical works can be classified mainly into two different sections. The first one (discussed in the next Chapter) includes accurate numerical calculations —first-principles ab initio— giving information also on the electronic charge distribution at the interface; all theories of the band lineup other than such ab initio calculations can be described as "model" theories, in that they make severe approximations describing the interface. These theories, which I am going to review in the present Chapter, have the advantage of being easier to be performed, can be applied to a wide range of materials, and tend to focus on the physical mechanisms which are believed to govern the phenomenum rather than on the exact description of all the physical quantities involved.

3.1 The theoretical approach

The problem of the reference bulk level and the absolute energy scale

From a precise theoretical point of view, the band offset across the junction between two semiinfinite crystals A/B is defined in terms of the *local density of states*, $n(\epsilon, \mathbf{r})$. Far from the interface on the two sides of the junction, $n(\epsilon, \mathbf{r})$ is a periodic function of $\mathbf{r}$, and its spatial average in those regions yields the bulk densities of states of the two materials, $g_A(\epsilon)$ and $g_B(\epsilon)$. The VBO is *exactly defined* as the difference between the valence–band edges of the two densities of states calculated as above described.

Different theoretical approaches have been proposed to calculate the band offset; most of them avoid the calculation of the local density of states and use directly the *integrated density of states* of the two materials, which is a more accessible quantity. In terms of the integrated density of states the band offset is determined by a rigid shift of the two bulk band structures (whose separate determination is a standard problem of electronic structure calculations), and the problem of band lineup is thus to determine such shift.

The use of the integrated bulk density of states poses however a problem, since for instance in the usual framework of density–functional theory (DFT) the density of states of the bulk infinite system is defined only up to an arbitrary constant in the energy. One usually does not worry about it if he is interested on the band structure of only one material, but this constitutes the critical point in the heterojunction problem, since the band structures of the constituting semiconductors are not "naturally" put on a common energy scale.

What is the reason of this indetermination? Can it be overcome? The common description of the single–particle energies in DFT treats them as "removal" and not "absolute" energies: as in an infinite system there is no "elsewhere" to remove an electron, it would follow that single-particle eigenvalues are ill-defined quantities. This argument is not completely true. In fact, were the interactions finite-range, the removal energy from a macroscopic but *finite* sample would *not* depend on shape or surface effects and would have therefore a well defined thermodynamic limit: in other words we could extract from a finite system information for the infinite one without using any specific feature about the "finiteness" of the sample. It is precisely the long-rangeness of the Coulomb interaction which makes removal energies depend on the detailed structure of the surface[36] and hence ill-defined in the thermodynamic limit.

Many model theories have implicitly or explicitly assumed that the VBO is determined only by bulk properties, and hence neglected interface specific effects and determined a "reference level" for each material: a long debate has been done on the best criterion to fix it. Actually there exists a prescription to determine an absolute

energy level to which refer the band structures of an infinite bulk semiconductor also in the framework of DFT[37,38]: provided it can be described as made up of neutral "building blocks", the thermodynamic limit is well defined, and does not depend on the specific features of the finite sample from which it is extracted.

An absolute energy scale to which refer the energy bands of semiconductors could thus be constructed, and once all the semiconductors were placed on it, the relative positions of their energy bands and hence the band discontinuity at an heterojunction would be automatically determined. Let us make in fact the assumption that a reference level E_r exists and is a well defined property of the bulk semiconductor. Then the position of the valence–band maximum relative to E_r is:

$$E_v^r = E_v - E_r \tag{3.1}$$

If we consider now the heterojunction A/B, we must line up the reference levels of the two semiconductors:

$$\Delta E_r = E_r(A) - E_r(B) = 0 \tag{3.2}$$

so that we obtain for the valence–band offset:

$$VBO = E_v^r(B) - E_v^r(A) \tag{3.3}$$

Note that this picture is valid only for ideal heterojunctions where the constituting semiconductors are *far away* from each other or however *not interacting*, i.e. where the interface details are not important. Commutativity and transitivity properties (see Eq. (2.1)), which are satisfied in many cases according to the experimental data, are direct consequences of Eqs. (3.1–3.3), as well as the interface–orientation independence.

In view of the experimental data, which clearly show that the VBO depend on the interface orientation in some heterojunctions, the energy scale so defined seems wrong. Some model theories have modified the concept, conserving the "universality" for all the semiconductors, but not the "absoluteness" inside the same material: semiconductors constituting the heterojunction are more properly considered as semiinfinite systems, and for each one several different reference levels are established, according to the specific surface orientation. This approach is not appropriate if one considers real surfaces, with physical features very different with respect to those of the ideal surfaces constituting the interface; it can be interesting if model surfaces ideally joined to constitute the interface are considered. Note that this picture is still in agreement with the commutativity and transitivity properties.

The effects of the interaction of the two surfaces put in contact and the role of the interface details are taken into account semi–empirically by some model theories: they start from the picture of an absolute energy scale with a rigid alignment of the

band structures, and then, by modeling the rearrangement of the charge density at the interface, they determine the additional potential shift which give the final energy–band discontinuities; notice that this interface–dependent additional shift breaks the concept of the absolute energy scale. Interface details are also taken into account by the self–consistent *ab–initio* calculations which give an accurate description of the electronic distribution at the interfaces, and which hence do not assume *a priori* the existence of a common absolute energy scale; in such case the results obtained will indicate whether the idea of a common reference energy scale is valid and can be recovered, and will eventually suggest other models.

Band offsets vs. electrostatic potential lineup

We will see now how the relevant informations for the band offset problem can be extracted from the electronic structure calculations: I face the question following a general line, by focusing some concepts common to many different theoretical approaches, both model and ab–initio theories.

We have identified the problem of the band offset with the determination of a shift of the band structures of the two materials, when the integrated bulk density of states is used. The average electrostatic potential for the electrons is a convenient reference level for the band structures, since, even if one is not able to determine it in an infinite bulk material, its difference across the interface between two semi–infinite solids is, on the contrary, well defined.

It is easy to realize that the potential lineup is in general different from zero, even in the equilibrium crystallographic and electronic structure. Let us consider the heterojunction along the growth direction as made by charged planes, and start in particular from a "false" homojunction constituted by the identical material on the two sides: locally the solid is not neutral, and can be described as a continuum of infinite microscopic planar capacitors; they globally neutralize each other, the average electric field is zero, and the average electrostatic potential is constant. If we consider now an heterojunction, because of the different charge distribution in the two materials at the interface there is a microscopic capacitor whose effects are not neutralized by the others: while the electric field outside is zero, the average electrostatic potential jumps from one constant value on the left side of the capacitor to a different value on the right side.

What one can thus exactly describe is the relative position of the band edges in the heterojunction by adding separately the two bulk band structures to the electrostatic potential lineup —as indicated in Fig. 3.1— rather than the absolute position of the two bands separately.

The energy–band discontinuity can thus conveniently be written as a sum of two different contributions(*), one strictly dependent only on the different bulk properties

(*) I stress that the splitting of the VBO into different contributions is arbitrary and not uniquely

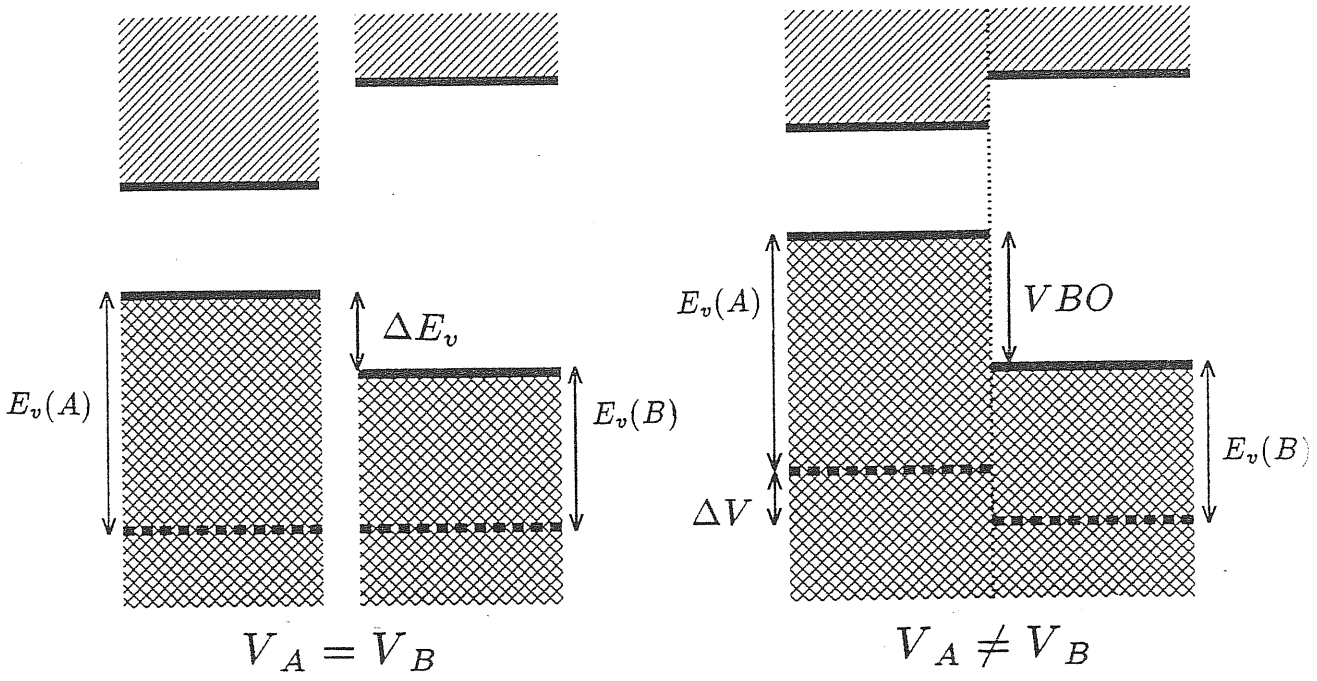

Figure 3.1. Valence-band offset, VBO, vs. electrostatic potential lineup, ΔV. ΔE_v is the difference between the bulk valence edges, referred to their own average electrostatic potentials.

of the materials forming the heterojunction, and the other in principle on the actual structure of the interface. For the valence–band offset (VBO) —for instance— we have:

$$VBO = \Delta E_v + \Delta V, \tag{3.4}$$

where ΔE_v is the difference of the bulk band edges referred to their own average electrostatic potential, and ΔV is the electrostatic potential lineup. All the interface features are contained in the macroscopic electrostatic quantity ΔV, whereas ΔE_v is obtained by two distinct bulk calculations. In practice in the present calculations the electrostatic potential is the one generated by the valence electron distribution neutralized by a uniform background.

Within this scheme, one part of the work (the calculation of the energy band structures of the bulks) is a standard solid–state problem. The new specific problem is the calculation of the electrostatic potential lineup: this requires a description of the charge distribution at the heterojunction, which can be rough in model theories or accurate in self–consistent *ab–initio* calculations.

---

defined; some model theories in fact do not make use of the concept of electrostatic potential lineup, or they use some alternative distinctions.

Planar and macroscopic averages

The question is now how the electrostatic potential lineup is related to the three–dimensional distribution of the charge density, and to the corresponding electrostatic potential obtained via Poisson equation. As an example, we show in Fig. 3.2 the contour plots of the self–consistent electronic valence charge for the GaAs/AlAs (001) heterojunction: there are two equivalent interfaces, one at the figure center and one at the figure borders, but on this scale the presence of interfaces is hardly detectable.

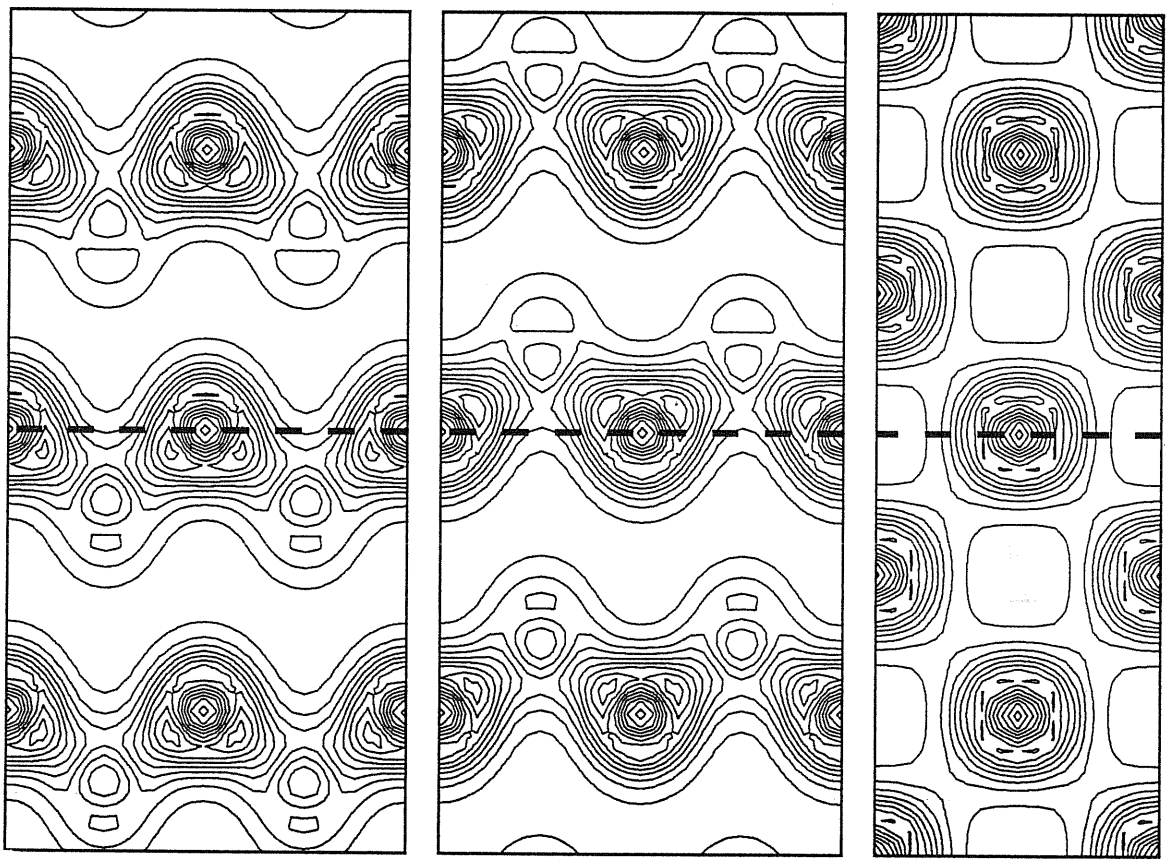

Figure 3.2. Charge density of GaAs/AlAs (001) heterojunction: the plots are centered at an interface anion; GaAs down; AlAs up. Left: (110) plane; middle: $(1\bar{1}0)$ plane; right: (010) plane.

The physical quantities $f(\mathbf{r})$ we are interested in —such as the electron density $n(\mathbf{r})$ or the electrostatic potential $V(\mathbf{r})$— are periodic in the planes perpendicular to the growth direction (z axis). Since the most interesting feature is the z dependence of these three-dimensional quantities, it is convenient to define $\overline{f}(z)$ as the xy planar average of $f(\mathbf{r})$: $\overline{f}(z) = \frac{1}{S}\int_S f(x,y,z)dxdy$, where S is the area of the two–dimensional unit cell on the xy plane. From the three-dimensional electronic density one thus gets the one-dimensional density $\overline{n}(z)$ and potential $\overline{V}(z)$ shown in Fig.

3.3(a) (in that case $S = \frac{a_0^2}{4}$).(*)

The function $\overline{f}(z)$ is nonperiodic in the interface region, and goes asymptotically into two different periodic functions (having the same period L in the example shown here of a lattice–matched heterojunction) far from the interface. Any interface effect is due to the *difference* between these periodic functions and is barely visible on this scale, being almost masked by bulk oscillations; for instance, the constant shift of the average electrostatic potential, approximately sketched in Fig. 3.3(a) with a dotted line, is very small with respect to the amplitude of the bulk oscillations.

In order to get rid of bulk effects and blow up interface features a new procedure has been recently proposed[38], which avoids any definition of an ideal interface and its use as a reference. Such a procedure is an application of the concept of macroscopic average which is well known in classical electromagnetism.[39]

Any macroscopic quantity $f^{(macro)}(\mathbf{r})$ is related to its microscopic counterpart $f^{(micro)}(\mathbf{r})$ through a convolution:

$$f^{(macro)}(\mathbf{r}) = \int w(\mathbf{r} - \mathbf{r}') f^{(micro)}(\mathbf{r}')\, d\mathbf{r}', \tag{3.5}$$

where $w(\mathbf{r})$ is real, nonzero in some neighborhood of $\mathbf{r} = 0$ and normalized to unity over all the space[39]. It is a trivial matter to show that such an averaging commutes with the space and time differentiation occurring in Maxwell equations. The choice of the filter function w is largely arbitrary; however each macroscopic problem has its own appropriate lower limit of relevant lengths and this sets the size of the w function to be used[40]. When studying electrostatics in crystalline materials, periodicity allows to take length scale as small as $\Omega^{1/3}$, where Ω is the unit cell volume.

As far as the semiconductor heterojunction problem is concerned, matters are simple for interfaces between two lattice-matched materials, where the above defined characteristic function w can be chosen as material-independent. Macroscopically averaged quantities (charges, fields, potentials) show no microscopic oscillations on either side of the interface and recover the macroscopic limit in the two bulks. Conversely, deviations from the bulk macroscopic values define the interface region: for instance the one relative to the electronic charge density defines unambigously an *interface dipole*.

Before the concept of macroscopic average was introduced into the heterojunction problem[38], there has been a long-standing confusion about the nature of the interfacial dipolar charge distribution which gives rise to the potential lineup, as it will appear in Sect. 3.1. The macroscopic average, on the contrary, allows to define an interface dipole and to blow up any genuine interface features using the interface

(*) Notice that, according to the definitions used here, the planar and macroscopical averaged quantities have the same dimensions as the original quantities.

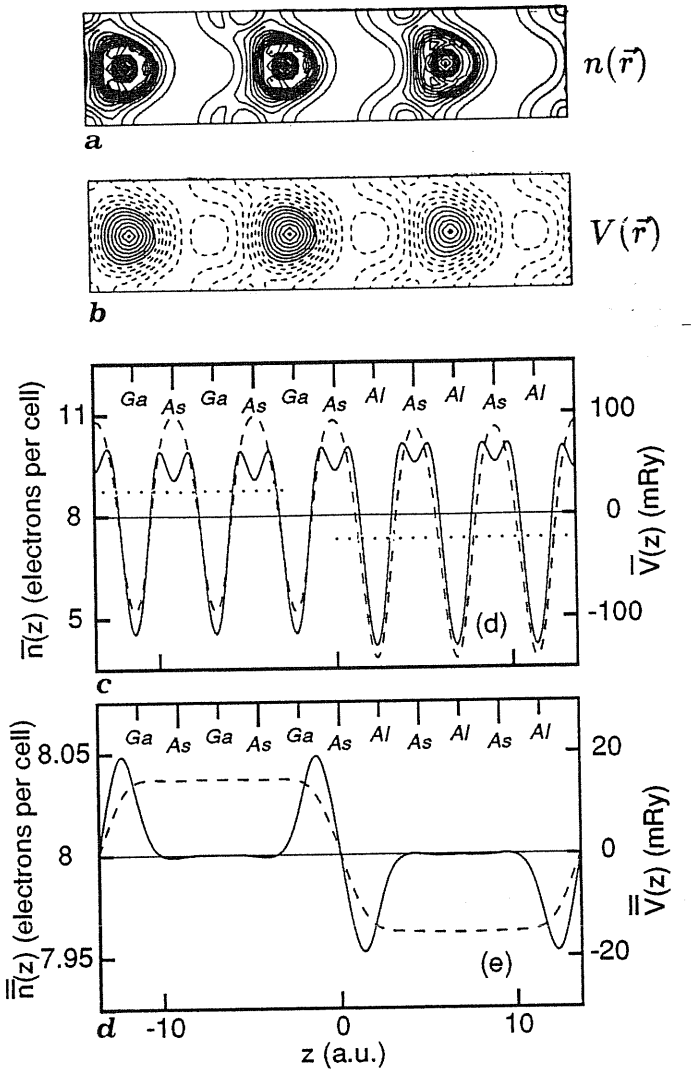

Figure 3.3. Contour plots of the charge density (a) and the electrostatic potential (b) on the $(1\bar{1}0)$ plane for GaAs /AlAs (001) heterojunction; planar (c) and macroscopic (d) averages of the charge density (solid lines) and the electrostatic potential (dashed lines) of GaAs /AlAs (001) heterojunction: the dotted line in (c) indicates approximately the average electrostatic potential in the two bulk regions.

charge as the only ingredient. The interface charge profile so obtained is related, via the Poisson equation, to the potential lineup.

In practice, once we have the planar average of the quantity under study, the next step to obtain the macroscopic average is to pass the function $\bar{n}(z)$ (or $\bar{V}$) through the (one-dimensional) filter defined by the convolution:

$$\bar{\bar{n}}(z) = \frac{1}{L}\int_{z-L/2}^{z+L/2} \bar{n}(z')\, dz' = \frac{1}{L}\int \Theta\Big(\frac{L}{2} - |z - z'|\Big)\bar{n}(z')\, dz', \qquad (3.6)$$

where Θ is the step function and L is the period of planar averaged bulk quantities (e.g. $L = \frac{a_0}{2}$ for the GaAs/AlAs (001) heterojunction, illustrated here). The results are shown in Fig. 3.3(b). As for the charge, the figure shows, on a magnified scale,

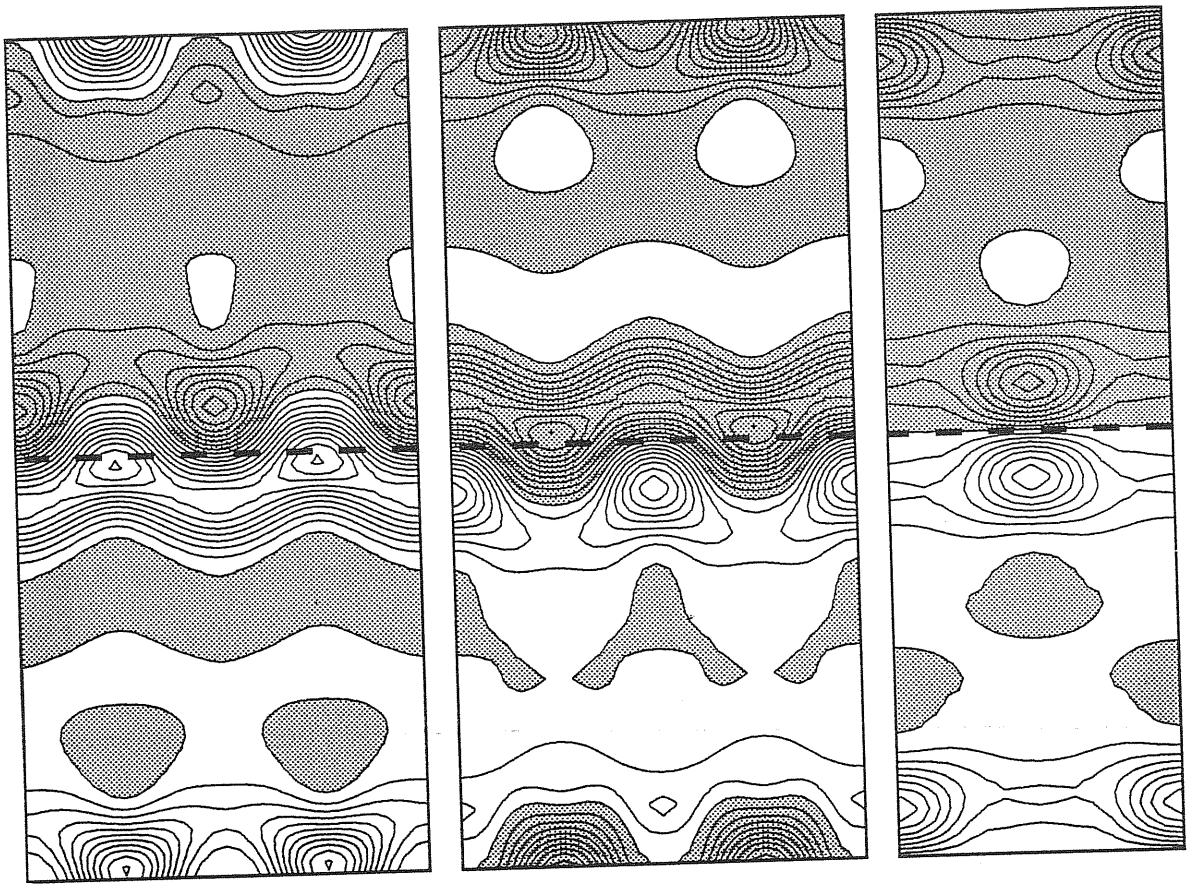

Figure 3.4. Three-dimensional macroscopic average of the electron density in GaAs/AlAs (001) heterojunction, projected onto the (110), (1$\bar{1}$ 0), and (010) planes. A characteristic function of the bulk Wigner-Seitz cell is used in this case as a filter function. The shaded area indicates regions where the density is lower than eight electrons per cell. GaAs down, AlAs up.

a dipole profile, localized in the neighborhood of the interface. The electrostatic potential varies across the interface from a constant value to a different one, the lineup Δ being related to the dipole moment of the charge profile through:

$$\Delta V = 4\pi e^2 \int z\big(\bar{\bar{n}}(z) - n_o\big)\, dz, \tag{3.7}$$

where n_o is the average electronic density of the two bulks (eight electrons per cell, in the case of elemental or binary semiconductors, whose unit cell is a zincblende cell). It is important to notice that Eq. (3.7) allows to define unambiguously the concept of *interface dipole* for any interface, and also, in particular, for a surface.

The macroscopic averages shown so far have been built in two steps: first the (x,y) planar average and then the one-dimensional convolution. It is easy to realize that this construction is equivalent to the three-dimensional convolution of Eq. (3.5), where the filter function $w(\mathbf{r})$ has been taken as the characteristic function of a suitably shaped (slab-adapted[36]) unit cell of volume SL. Other choices are possible, all giving the same macroscopic values but different resolution and different shapes of the interface charge. Previously I have shown the planarly averaged interface profile; suppose instead one wants to investigate on which *sites* this dipole charge is located. To this aim, (x,y) resolution is needed: the natural and unbiased choice for the filter function w is then the characteristic function of the bulk Wigner-Seitz cell.

Starting again from the three-dimensional charge of Fig. 3.2, one obtains the macroscopically averaged charge shown in Fig. 3.4 using a magnified spacing between contours. This kind of macroscopic average goes—as the previous one—to the constant value n_o in the two bulk regions; it deviates from n_o in the neighborhood of the interface, but this deviation is no longer (x,y) independent. Instead it is periodic in the (x,y) planes and shows very clearly that the dipoles responsible for the lineup are localized at the interface anions (and nearby bonds) for this (001) geometry. Such a localization can be defined only to a resolution of $\Omega^{1/3}$, intrinsic to the macroscopic average itself.

3.2 Existing models

The "common–anion rule"

Before discussing the several models theories based on the concept of the "reference level", I just mention a completely different and very empirical model, which is known as "common–anion rule". It simply states that the valence–band discontinuity between two semiconductors with the same anion is *very small*, the reason being that the states at the valence–band maximum have primarily the anionic contribution. It was thus believed for many years that the band–gap difference in GaAs/AlAs is primarily absorbed by the CBO; recent precise measurements (see Tab. 2.1) have proved it to be wrong, and have shown that the physical picture which is at the origin of such old and well established rule is too much simplified and misleading in predicting band lineups.

Electron affinity rule: Anderson, Freeouf and Woodall, Harrison, Frensley and Kroemer

The "common–anion rule" is just a qualitatively general assertion; quantitative predictions can be extracted from other model theories and empirical rules involving a minimum of computation, since they reduce the problem of band lineup to few fundamental factors. A large group of model theories assumes that the vacuum level (let us call it E_0) outside the surface of each semiconductor is the best reference level to line up the bands in the two semiconductors: it is a well defined external reference, constant throughout the junction in the absence of dipole layers (see Fig. 3.5). These theories are collectively referred to as "electron affinity rule" theories, although historically this name is for a particular one due to Anderson[41]: the conduction–band discontinuity is given by the difference in the electron affinities of the two semiconductors. Once the band gap is known, it is equivalent to formulate the model for the VBO: Φ is the ionization potential (positive) for each semiconductor; the relative position of the valence–band edge is $E_v^r = -\Phi$, so that

$$VBO = -\Delta\Phi = -(\Phi_B - \Phi_A) \tag{3.8}$$

As long as the two semiconductors are separated by a vacuum region, as assumed in Fig. 3.5, this picture is correct; the problem is that dramatic changes can occur when the two semiconductors are put in contact. However, if the picture remains valid and the valence–band maxima in the two semiconductors fall at an energy $-\Phi$ with respect to the vacuum level, this theory would correctly predict the band lineup.

Moreover the ionization potential is not a true "bulk" reference level, but it rather depends on the orientation and in general on the structure of the crystal surface. The first non trivial problem is *which* ionization potentials have to be used.

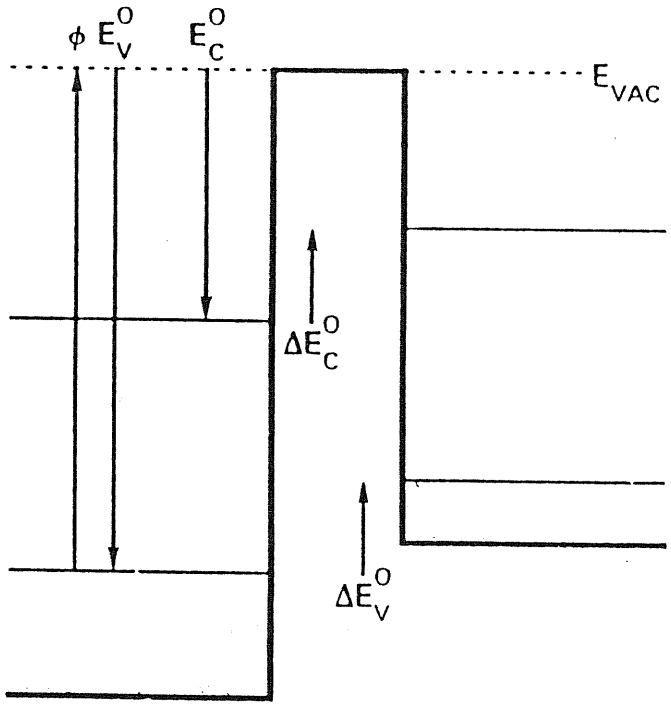

Figure 3.5. Schematic diagram of the electronic structure of two semiconductor surfaces, separated by a small vacuum region. The bold dashed line is the potential associated with the two semiconductors, the dotted line is the vacuum level, and the horizontal solid lines show the valence and conduction bands; symbols are explained in the text.

The experimental values are not appropriate: even if one considers the value for the surface oriented as the corresponding interface, the real surface is not the best choice, since the structural surface details which strongly influence the ionization potential can be completely different at the interface or irrelevant for the heterojunction problem. Reconstructions and relaxations for instance are not present when the two surfaces are brought together, but instead different atomic displacements from the ideal structure can occur because new kind of chemical bonds are present at the interface.

It seems necessary to go beyond the empirical use of the experimental affinities or ionization potentials, as made for instance by Freeouf and Woodall[42]; the large spread of data for electron affinities implies that even the choice of their experimental values is very difficult, and the partially good results obtained for some systems would not provide a significant test for the validity of the affinity rule itself. Electron affinities or ionization potentials have to be *calculated* for the appropriate ideal semiconductor surface.

The electron affinity rule works exactly if the interface charge density would be equal to the superposition of the charge densities of the two surfaces for which the ionization potentials are calculated; such "ad hoc" reference surfaces must be

calculated for each pair of semiconductors, and this of course breaks the concept of an absolute bulk reference level.

The differential charge density δn between the true interface charge distribution and the superposition of the two reference surfaces originates an "interface dipole" determining a different shift on the average electrostatic potential in the two bulks and hence the possible variation of the relative position of the band edges; if that the differential charge δn is very small this electrostatic shift is negligible and the electron affinity works well.

Harrison[43] proposes a LCAO tight–binding approach, taking as the vacuum level the one determined by the free atoms. Although he does not make use explicitly of any reference surface, referring to the free–atom terms is equivalent to considering the reference surface as superposition of isolated atoms. The predictions obtained in this tight–binding framework agree with the experimental data within 0.2–0.4 eV; the inclusion of the relativistic effects does not seem to improve much this kind of agreement.

Frensley and Kroemer[44] point out that if the total charge density can be represented as a superposition of spherical charges, and if these are sufficiently localized to guarantee a negligible charge density in the interstitial positions, then the average electrostatic potential in the interstitial positions approximates well the electrostatic potential at infinity for the surface made up by the superposition of such spherical charge densities, and hence this interstitial average potential is a good reference level. The position of the valence–band edges with respect to such reference level is calculated using a self–consistent pseudopotential approach for the bulk semiconductor. The severe approximations made in the actual implementation of the model are perhaps responsible of the poor success of the model itself: a refinement of the model might give good results. In any case it is a very convenient and computer–time saving approach, since only a single calculation for the bulk potential is needed, while others require "to build up" the (model) heterostructure.

Tersoff and the interface quantum dipoles

The basic idea of Tersoff[45]'s model is related to the affinity rule, but it emphasizes the role of the interface dipole,(*) originated by the difference between the charge distribution in the real heterojunction and in the "reference" heterojunction formed by rigid juxtaposition of the two reference surfaces. The idea is that whenever two different materials are brought together, there is a charge flow from the less electronegative material to the more electronegative one.

(*) The term "interface dipole" is used to indicate both the charge density and the associated electrostatic potential.

To explain his theory, Tersoff starts with the metal–metal interface: in a metal the Fermi level is well defined, and the workfunction measures the electronegativity of the system. When two metals are brought together, charge flows until the two Fermi levels are lined up, and this corresponds to the formation of an interface dipole equal in magnitude to the difference between the two Fermi levels; the final self-consistent lineup is independent on the starting reference surfaces, and is simply obtained by aligning the two Fermi levels. Note that this serves as a true "bulk" reference level. In case of contact between two highly insulating and non interacting solids, such as rare–gas solids, there will be a very small charge flow starting from the ideal "reference heterojunction" configuration.

For a semiconductor–semiconductor interface the situation is somewhat more complicated and stays midway between these two extremes. Tersoff proposes the existence of a "charge neutrality level" characteristic for each semiconductor and playing a role analogous to the Fermi level for metals. I try first of all to explain the concept from the point of view of macroscopic electrostatics, and then the microscopic origin of the charge rearrangement.

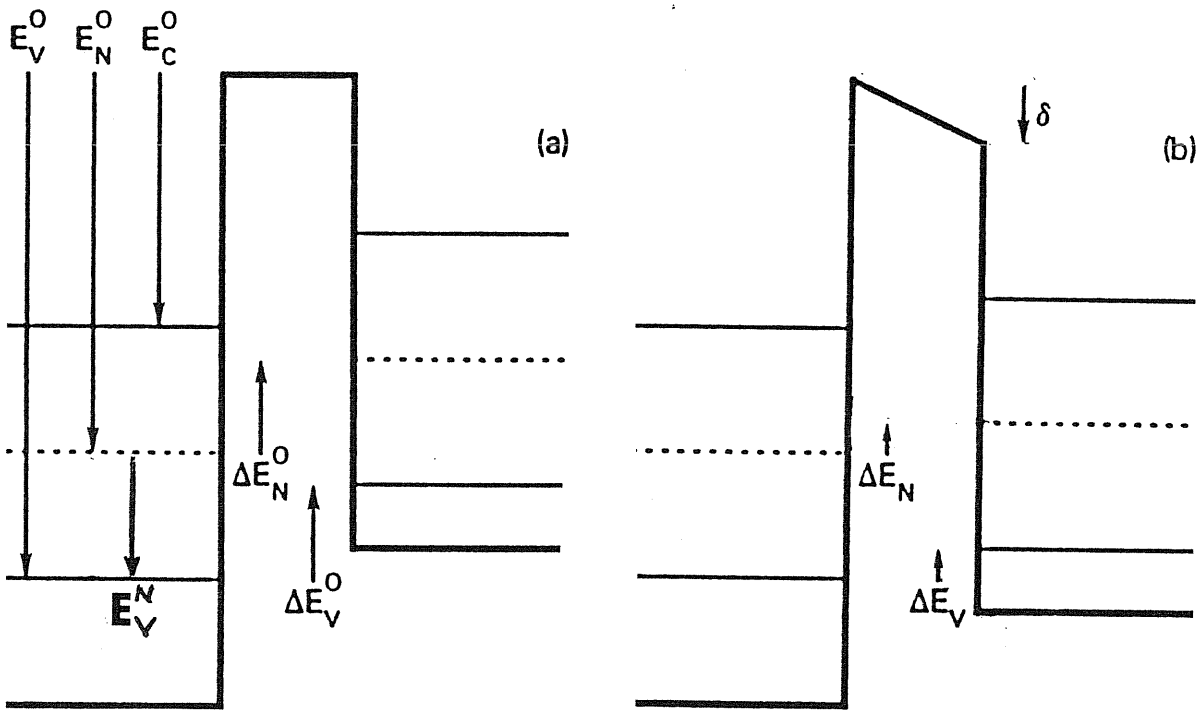

Figure 3.6. Relationship between interface dipole and alignment of the band structures. (a): for two semiconductors separated by a small vacuum region, or in contact but non interacting, or again before the charge density rearrangement occures (b): for two semiconductors put together, after the charge rearrangement. Symbols are explained in the text.

The classical electrostatics which controls the phenomenum is explained in Fig. 3.6. Fig. 3.6(a), which refers to two semiconductors separated by a vacuum region or non interacting, is actually similar to Fig. 3.5, except that the introduction of the neutrality level here allows to express all the quantities also with respect to it.(*) The superscripts *0* and *n* indicate respectively that quantities are measured with respect to the vacuum level and to the neutrality level; absence of any superscript refers to the situation where charge rearrangement has already occurred. δ is the interface dipole which tends to reduce the relative difference between the two neutrality levels, and it is proportional to the final difference:

$$\delta = -\alpha \Delta E_n \tag{3.9}$$

Looking at Fig. 3.6(b) the VBO can be expressed as:

$$VBO = \Delta E_v^0 + \delta \tag{3.10}$$

and rearranged with a little bit of algebra in the form:

$$VBO = \frac{\alpha}{1+\alpha} \Delta E_v^n + \frac{1}{1+\alpha} \Delta E_v^0 \tag{3.11}$$

In case of infinite susceptibility α we find $VBO \rightarrow \Delta E_v^n$, i.e. the metallic case with the alignment of the neutrality levels, whereas neglecting the dipole gives $VBO \rightarrow \Delta E_v^0$, the non–interacting case with the alignment of the vacuum levels and the affinity rule. It is instructive to notice the correspondence between Eq. (3.10) and the one I usually refer, namely Eq. (3.1), with ΔE_v and ΔE_v^0 having in practice the same role: ΔE_V^0 is a pure "bulk" contribution, (the difference between the two band edges referred here to the respective zero level and to the zero of Hartree potential in my language); δ is the electrostatic contribution which is in principle interface–dependent.

It is clear from this model that *the electrostatic lineup* —to use the language defined in Sect. 3.1— *is an essential part of the VBO*: at this point I can stress that the idea of referring to the vacuum level is not intrinsically wrong, but the problem is to take into account the discontinuity of the vacuum level at the interface.

How in practice can the interface dipole be evaluated, or how can be evaluated the quantities ΔE_n or ΔE_v^n? After the macroscopic point of view, we should understand now the microscopic nature of the interface dipole and of the neutrality levels. This is related to the gap states: bulk electronic states in one semiconductor which fall energetically in the band gap of the other tunnel into the latter, and this tunneling is responsible for the interface dipole. The reference situation for the

(*) This should suggest us a connection between the various reference level theories.

definition of the interface dipole is the alignment of the neutrality levels, as shown in Fig. 3.6(a); but the relationship can be usefully reversed, and the neutrality levels defined in such a way that their alignment correspond to charge neutrality at the interface; the gap states which can guarantee this condition are those —at some effective midgap energy— crossing over from mainly valence- to conduction-band character, and they can be determined with some Green's function technique. On the basis of macroscopic electrostatic arguments (macroscopic screening of the dipole) Tersoff shows that $\alpha \approx 10$ typically, thus according to Eq. (3.11) the case of heterojunctions is near to the metallic case, and the simple alignment of the neutrality level (without calculating explicitly neither ΔE_v^0 nor the self-consistent dipole δ) provides a good quantitative estimate of the VBO.

Compared to the previous works, the Tersoff's model has the simplicity of the other reference-level model theories but it includes the discontinuity of the vacuum level on the two sides of the junction, which is essential for an adequate comprehension of the problem and for a quantitative predictive power of the model itself.

Van de Walle and Martin

Van de Walle and Martin [37] have proposed a model which is again in the spirit of the "reference level" theories, takes into account the discontinuity of the vacuum level across the interface, and moreover is very well defined and conceptually clear. They use the superposition of *neutral* atomic charge densities: the potential outside them goes asymptotically to a zero level which is the vacuum level, and once these neutral objects are put together to form the solid, the average electrostatic potential in the solid is uniquely defined, being linear with respect to the charge density.(*) At this point the concept of reference surface is useless, since the presence of a surface would not modify the average electrostatic potential. However, this reference level is really a well defined "bulk" property, with all the consequences that I have discussed before: i.e. this model predicts a VBO strictly independent on the interface orientation (also for heterovalent interfaces where, according to the present results, this seems not to be true), and satisfies the transitivity rule within its numerical accuracy (about 0.05 eV). The total accuracy of the model can be estimated considering that the agreement with the self-consistent ab-initio calculations (see next Chapter) is for a large number of systems better than 0.1 eV, and always within 0.25 eV.

Why does this model work quite well, even it is not able to reproduce the bulk charge density? The difference between the self-consistent bulk charge density and the corresponding superposition of free atoms is a function with approximate bulk

(*) The *total* potential of the solid, which includes exchange and correlation contributions, is not linear in the charge density, but this is not a problem since these contributions can be added *a posteriori* being of "bulk" nature.

periodicity; the difference between the self–consistent heterojunction charge density and the correspondent superposition of free atoms has properly the periodicity of the supercell used to describe the interface, but if it is dominated by a component with bulk periodicity, as it is the case, it does not affect much the electrostatic potential lineup.

Baldereschi, Baroni and Resta

On the same line of Van de Walle and Martin, Baldereschi, Baroni and Resta[38] have proposed a model in which the heterojunction is well described by "building blocks" derived from the bulk charge densities of the constituents, using crystal symmetry. For lattice-matched common-anion heterojunctions, they decompose the bulk electronic densities into Wigner-Seitz cells (WSC) centered on the cations and with fractions of anions at four corners, as shown in Fig. 3.7. They are neutral and by symmetry do not have dipole nor quadrupole moments; i.e. the potential of WSC's is sufficiently short range not to produce macroscopic effects across the interface.

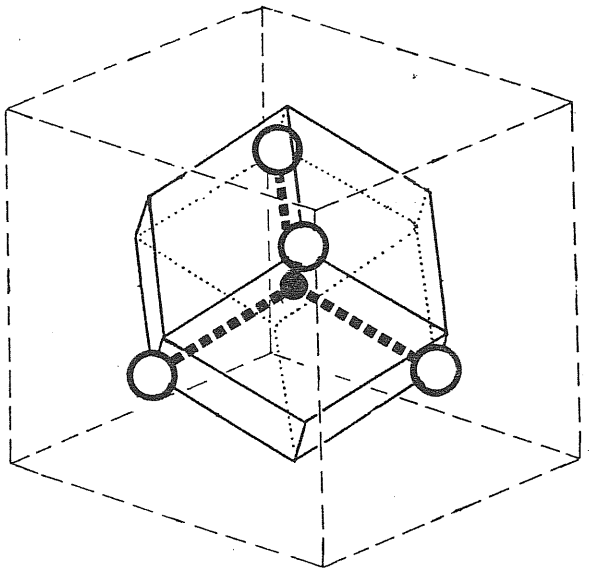

Figure 3.7. Zincblende Wigner–Seitz unit cell (WSC) centered on a cationic site: there are four anions at the corners with weight $\frac{1}{4}$ for the geometry of the cell.

The same holds obviously for the WSC obtained from the difference of the two bulk WSC's, $\Delta\text{WSC}=(\text{WSC}_B-\text{WSC}_A)/2$, and for the one obtained as the average, $\langle\text{WSC}\rangle=(\text{WSC}_B+\text{WSC}_A)/2$; each of the two bulks is also exactly described by superposing with the proper sign (+/− for B/A) the ΔWSC's to a periodic reference crystal which is the "average" bulk constituted by the juxtaposition of $\langle\text{WSC}\rangle$'s. This picture may suggest an alternative possible approach to the problem: I stress in fact that, in order to extract the relevant informations for the electron charge distribution

and the electrostatic potential, it is sufficient to model the *difference* between the two sides of the heterojunction and *not* the complete heterojunction, since any periodic reference crystal defined in this or in similar ways does not contribute with macroscopic effects to the interface phenomenology.(*) For computational purposes the two procedures are equivalent: in any case the model interface electronic density or its difference from the reference periodic crystal is obtained simply by rigid juxtaposition of WSC's —which cover all the space without overlapping— according to the cationic type at the center of the cell (proper sign for ΔWSC's). The three-dimensional model electron density has small discontinuities at the boundaries of the WSC's at the interface, while the corresponding potential is continuous. The results obtained by applying the model are satisfactory, although a small electronic rearrangement must occur in order to ensure at least charge continuity. The agreement of the model with first-principles supercell calculations is extremely good, at least for GaAs/AlAs —where it has been tested— as it can be seen in Fig. 3.8:(†) the spatial extent of the interface regions and the shape of the dipolar charge distributions are well reproduced, and the VBO is within 0.01 eV, i.e. the numerical accuracy of the calculations.

The formulation of the model is useful for understanding the nature of the VBO and in particular the electrostatic potential lineup: its origin has to be searched in the different electronic distributions around the *As* nucleus when the latter is surrounded by different nearest neighbors: this is indeed indicated by the ΔWSC's, which are the building blocks actually contributing to the interface dipole.

Apart from this general observation, from the practical point of view of a predictive power, this model has advantages and disadvantages with respect to the Van de Walle and Martin's model: on one side, it *exactly* reproduces the electronic density in the bulk, provides a more accurate description of density profiles at the interface, and it is one order of magnitude more accurate and practically coincides with first-principles calculations; on the other side it is less general since in its present form it only applies to common-anion (or common-cation) lattice-matched heterojunctions. Extensions of this simple model to more general interfaces are possible: I will present in the last Section of the present Chapter a possible reformulation of it and its application to the more general case of no–common–ion lattice–matched heterojunctions.

(*) The macroscopic average of a periodic reference crystal is a constant, and the possible interface dipole present are due to the deviations from it.

(†) The first–principles interface calculations are the subject of Ch. 4 and Ch. 5, but some results are anticipated here to discuss the validity of the model under study; however, the density–functional self–consistent *ab–initio* pseudopotential method is also used here to obtain the bulk ingredients.

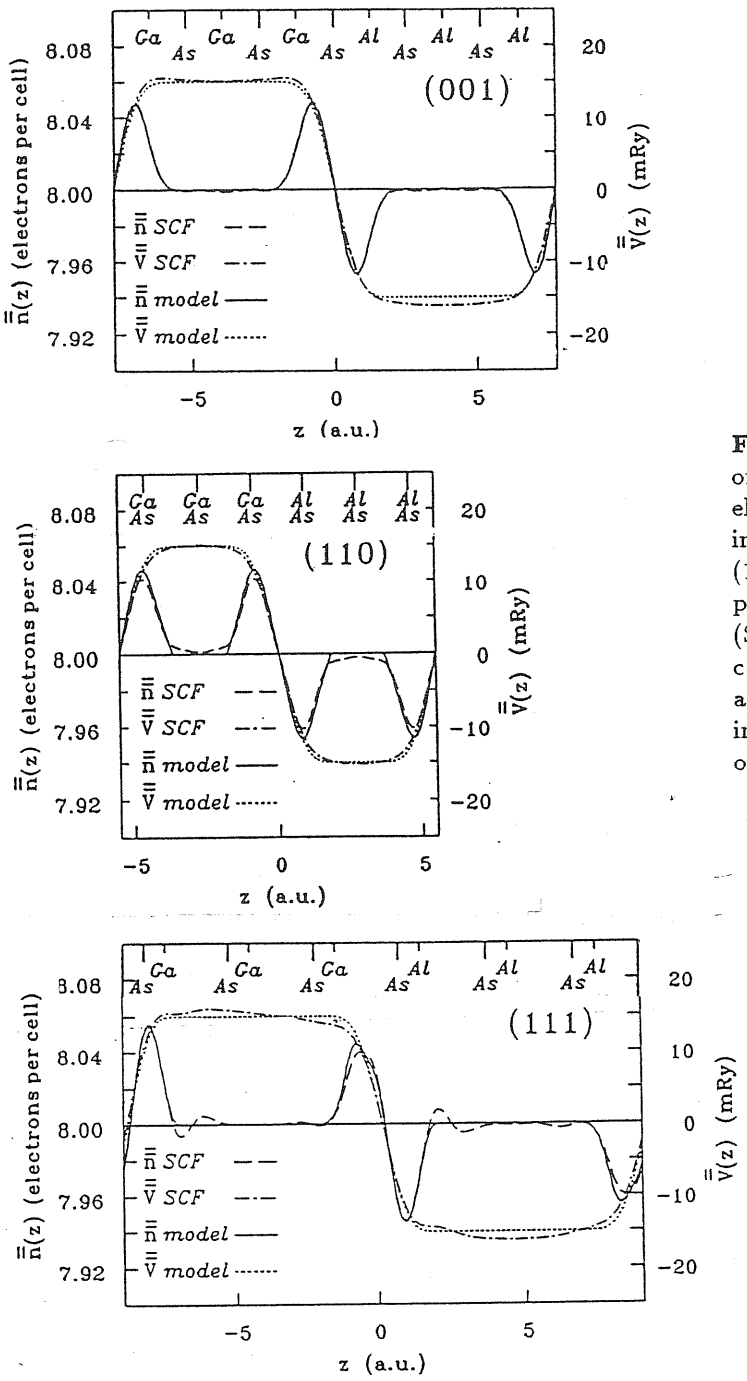

Figure 3.8. Macroscopic averages of the electron density and the electrostatic potential for GaAs/AlAs interface along (001), (110), and (111) orientation, obtained from first-principles self-consistent calculations (SCF) and from the WSC model calculations. Each panel refers to a system with two complementary interfaces, one at the figure center and one at the figure borders.

3.3 Other theoretical calculations

Beyond the neutrality–level rule: Tejedor and Flores's theory

An approach similar to the one of Tersoff in the basic ideas of neutrality levels and charge transfer is proposed by Tejedor and Flores[46]. For each heterojunction they perform separately three self–consistent tight–binding calculations: i) one for each of the two bulks separately, whose band structures are then put in a relative position according to the empirical electron–affinity rule: this is taken as initial unperturbed situation to calculate the initial lineup δE_n^0 between the charge neutrality levels; ii) one for the complete interface to obtain the final band offset and hence a final difference δE_n. The initial difference δE_n^0 is screened by a flow of charge from one semiconductor to the other up to the final value $\delta E_n = \mathcal{S}\delta E_n^0$. The screening factor $\mathcal{S}$ is obtained by comparing the calculated δE_n and δE_n^0; it is directly related to the susceptibility α appearing in the Tersoff's model: $\mathcal{S} = (1+\alpha)^{-1}$.

The authors propose to calculate it from the dielectric behaviour of the two semiconductors A and B in the heterojunction: to this purpose they define an "effective interface dielectric constant" ϵ^{eff} for each semiconductor and rewrite the screening factor in terms of such dielectric constants:

$$\mathcal{S} = \frac{1}{2}\left(\frac{1}{\epsilon_A^{eff}} + \frac{1}{\epsilon_B^{eff}}\right) \tag{3.12}$$

By comparing different screening factors calculated self–consistently for several systems and writing them according to eq. (3.12), the authors then give an estimate of the corresponding effective interface dielectric constants ϵ^{eff}. The authors also perform an "ab–initio" calculation of such effective dielectric constants —along a given direction only (one–dimensional model)— for a pure bulk or an interface geometry; the only results which allow comparison reported by the authors are for GaAs, i.e. $\epsilon_{\mathrm{GaAs}}^{eff} \approx 7.5$ from the pure bulk calculation, ≈ 6.2 from a (110) interface geometry (roughly speaking corresponding to the response of the semiinfinite crystal), ≈ 5.3 from fitting of the self–consistent screening factors.

A general limit of the present approach is that it has been developed originally for one dimension (both to calculate simply the charge neutrality level and to obtain the dielectric response of the materials under study), and its generalization to the three–dimensional case is something tricky and cannot say anything about the general aspects of the problem, such as for instance the dependence of VBO on the interface orientation and in general the bulk or interface features of the neutrality levels. Tejedor and Flores only argue that the charge–neutrality level can slightly depend on the interface orientation, having found a non negligible dependence of the VBO on the interface orientation in some systems with nonpolar interfaces, such as CdTe/HgTe.

Other calculations: Pickett, Louie and Cohen; Priester, Allan and Lannoo; Kunc and Martin

Other than properly called "model theories" are those which explicitly employ a specific interface calculation, i.e. a self–consistent study of the charge distribution, but that are not *ab–initio*.

I just mention here the pioneering work of Pickett, Louie, and Cohen,[47] with the use of empirical local–pseudopotentials. It is very difficult to determine the accuracy of such calculations; however there is generally a poor agreement with experiments for interfaces between a II–VI semiconductor and a less polar one.

A charge–dependent tight–binding calculation of band–offset has been done by Priester, Allan and Lannoo[48]. The tight–binding matrix is determined by using the parameters obtained from a fitting of the band structures for each bulk material; intra-atomic terms are determined only up to an unknown additive constant within each separate material. They assume that the possible charge transfer at the interface and the consequent Coulomb potential induced affect only the diagonal matrix elements of the Hamiltonian. The approximation introduced (local charge–neutrality condition) is on the same spirit of the minimization of the charge–neutrality levels in the Tersoff's model and in the Tejedor and Flores's theory. The authors find that the screening in the materials studied (GaAs, AlAs, Ge, ZnSe, InAs, GaP) is very efficient and three planes on each side at the interface are enough to take into account the interfacial charge transfer. The application to the case of the non–polar Ge/GaAs (110) interface in particular provides results in good agreement with the experimental data.

The work of Kunc and Martin[10] —again with the use of empirical pseudopotentials— is expecially remarkable as a pioneering study of the polar interfaces: they study in particular the Ge/GaAs (001) interface, and observe that in order to avoid charge accumulation in the interface plane and thermodynamical instability only two non-equivalent interfaces must be taken into account, i.e. with a mixed Ga–Ge or As–Ge plane. They also find that the average VBO between these two possible (001) interfaces equals the one for the nonpolar (110) case.

3.4 Discussion and limits of model calculations

Generalized or alternative formulations of the Wigner–Seitz–cell model

Among the models previously described, the Wigner–Seitz model proposed by Baldereschi, Baroni, and Resta has the merit of being well defined, simple in the application and very accurate at least for the case of GaAs/AlAs where it has been tested, but its applicability is limited to the common–anion (common–cation) heterojunctions. I have tried to generalize it to the more general heterojunction where both anions and cations are different in the two materials, i.e. C_1A_1/C_2A_2. The idea is to decompose the difference between the charge densities in the two bulks into WSC's centered both on the anions and on the cations, by isolating the two contributions. To this purpose I consider the two original bulks and the other two (even if not really existing) obtained by interchanging the ionic types of the two original ones: C_1A_1, C_2A_2, C_1A_2, and C_2A_1, at the lattice parameter of the heterojunction. The difference between the two original bulks can be exactly described as the juxtaposition of two types of WSC's, one centered on the cationic sites, i.e. in symbols:

$$\begin{aligned}\Delta WSC_C &= \frac{1}{2}\left(\Delta WSC_{C_1A_1/C_2A_1} + \Delta WSC_{C_1A_2/C_2A_2}\right) \\ &= \frac{1}{2}\left(\frac{WSC_{C_1A_1} - WSC_{C_2A_1}}{2} + \frac{WSC_{C_1A_2} - WSC_{C_2A_2}}{2}\right)\end{aligned} \tag{3.13}$$

and the other on the anionic sites:

$$\begin{aligned}\Delta WSC_A &= \frac{1}{2}\left(\Delta WSC_{C_1A_1/C_1A_2} + \Delta WSC_{C_2A_1/C_2A_2}\right) \\ &= \frac{1}{2}\left(\frac{WSC_{C_1A_1} - WSC_{C_1A_2}}{2} + \frac{WSC_{C_2A_1} - WSC_{C_2A_2}}{2}\right)\end{aligned} \tag{3.14}$$

The weight factor $\frac{1}{2}$ accounts for the double filling of the whole space. The two bulks are *exactly*(*) obtained by juxtaposition over the whole space of the two types of ΔWSC's —centered on the anionic sites and on the cationic sites, and with the proper sign, always positive or negative for a chosen bulk— on the periodic reference crystal obtained as the average of the two bulks, as already described. In symbols:

$$C_1A_1 = \frac{1}{2}(C_1A_1 + C_2A_2) + \sum_C \Delta WSC_C + \sum_A \Delta WSC_A \qquad (3.15(a))$$

$$C_2A_2 = \frac{1}{2}(C_1A_1 + C_2A_2) - \sum_C \Delta WSC_C - \sum_A \Delta WSC_A \qquad (3.15(b))$$

(*) From Eqs. (3.13–3.15) it comes out immediately that the contributions due to the two fictitious "mixed" bulks cancel each other.

In a similar way the heterojunction is described by juxtaposition on the same reference periodic crystal of the two types of ΔWSC's with their proper sign, changing in the interface region; the criterion for the choice of the sign is undoubted, depending on the ion present at the atomic site where the WSC has to be centered. Notice that the original formulation of the WSC model is immediately recovered once C_1=C_2 or A_1=A_2 (common–anion or common–cation heterojunctions).

Because of the intrinsic structure of the model, so described the interface features are limited in a small region close to the heterojunction, and bulk features are suddenly exactly recovered (see Fig. 3.9).

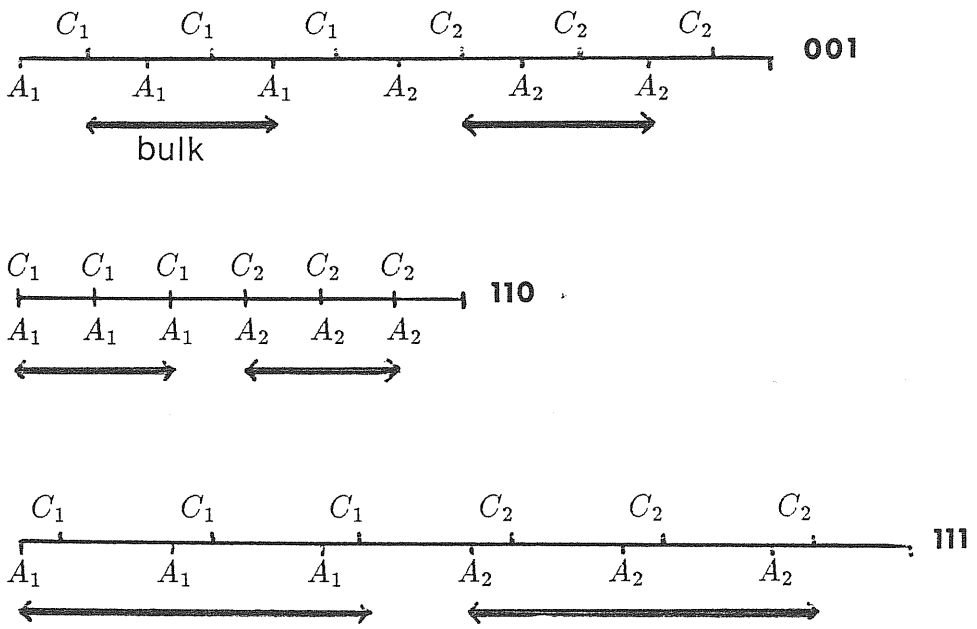

Figure 3.9. Schematic representation of the extension of the bulk and the interface regions and of a single WSC in (001), (110), and (111) heterojunctions. For the (111) heterojunction the two in principle non-equivalent interfaces are shown.

In case of no–common–ion heterojunctions for a given growth sequence (e.g. C_1A_1/C_2A_2) there exist in principle two different interfaces according to the termination (anionic or cationic plane) of the substrate: ...$A_1C_1A_2C_2$... or ...$C_1A_1C_2A_2$... Actually these two interfaces are equivalent, in that they correspond to the same electrostatic potential lineup; in the supercell describing the heterojunction they are both present, one in the atomic sequence described, and the other in the inverse sequence. This equivalence is rigorously an *identity* in the present model, because of the linear superposition of the building blocks. The case of (111) oriented interface in no–common–ion heterojunctions is slightly more

complicated, since two *in principle* really non-equivalent interfaces exist according to whether the transition from the substrate to the overlayer is when the chemical bonds are parallel to the growth direction or not. It is not difficult to check that the linear character of this model (linear superposition of the building blocks) implies not only the *commutativity* and the exact equivalence of complementary interfaces in case of (001) and (111) directions, but even the equivalence of the two different interfaces in (111) orientation: in each case the difference between the two interfaces —in principle non equivalent— can be simply thought of as a rigid shift of all the building blocks centered for instance on the cationic sites.

I apply here the model to the case of InP/$Ga_{0.47}In_{0.53}As$ interface, for which I anticipate some SCF results for comparison. In this case the GaInAs alloy is treated in the virtual crystal approximation (VCA), taking all the cations equal to the virtual one X=$\langle Ga_{0.47}In_{0.53} \rangle$ and thus using the zincblende structure. Unfortunately the results for such system are not as satisfactory as those for GaAs/AlAs, when compared with the self–consistent calculations: very small differences in the electronic charge distribution, hardly detectable on the scale of the planar averages, cause significant differences for the electrostatic potential lineup. The value obtained with the model in the three main crystallographic orientations is always underestimated by about 30% with respect to the SCF value ($\Delta V^{mod.} = -0.17$ eV versus $\Delta V^{SCF} = -0.25$ eV, with the same convergence parameters), as it can be seen in Fig. 3.10.

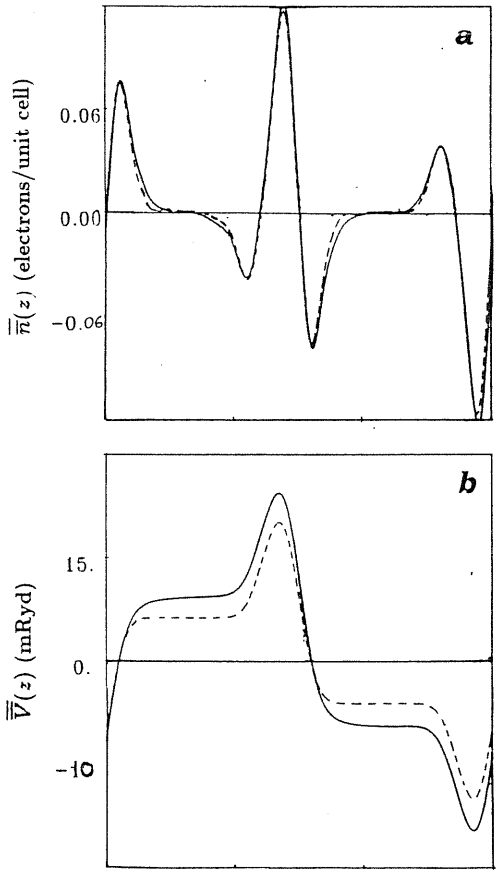

Figure 3.10. Macroscopic averages of the electron density (a) and the electrostatic potential (b) for InP/$Ga_{0.47}In_{0.53}As$ (001) heterojunction obtained from first–principles self–consistent calculations (SCF) (solid lines) and from the WSC model calculations (dashed lines). Each panel refers to a system with two complementary interfaces, one at the figure center and one at the figure borders.

At this point it may seem fortuitous the success of the model when applied to the simple case of GaAs/AlAs; the reason why the model works so well in one case and not in the other cannot be understood completely now, but many contributing factors can be identified, e.g. the strong chemical difference between the anions, which carry the largest part of the electron density.

I have made several attempts to define alternative building–block models, with the hope to do something more general and a little more accurate, but the results were not particularly satisfactory.

a) For instance I have tried to define as building blocks of the bulks the Wigner–Seitz BCC cells (volume $\frac{1}{2}$ of the FCC cell) centered on atomic or interstitial sites: the filling criterion to construct the heterojunction was the kind of atom present for BCC cells centered on atomic sites and a weighted average according to the nearest neighbor atoms for those centered on the interstitial site. This model appears more flexible than the WSC one since it allows in principle a larger extension of the interface region, but on the other side it completely neglects any interface detail since it simply reproduces the interface region as a uniform average of the two bulks (see Fig. 3.12).

b) Another attempt was to use even smaller building blocks, i.e. the Wigner–Seitz FCC cells centered on the middle of the chemical bonds or of their extension in the interstitial regions (volume $\frac{1}{8}$ of the unit WSC); in this case the filling criterion to construct the heterojunction was a weighted average according to the types of atoms or of interstitial sites joined with the bond or its extension. The model again allows a larger extension of the interface region, and with respect to the previous one it seems more flexible in describing the electronic structure at the interface; however, the degree of complication is increasing as long as the building blocks become smaller and smaller, and no improvements are obtained in the results at least for the cases where it has been applied.

c) I also tried an alternative generalization of the original WSC model, considering each WSC divided into four segments, according to the four ions at the corners; each segment is uniquely identified by the bond contained inside: if it is characteristic of one of the two bulks, then the segment can be immediately extracted from the pure bulk, whereas, if it is not the case, some *ad hoc* WSC should be considered in order to extract the segment of interest. Such a model is not convenient since it requires some interface–specific inputs (such *ad hoc* constructed FCC WSC's); conversely, it has the merit of being applicable to heterovalent interfaces.

Very approximate estimates of electrostatic potential lineup can be done analytically following these models: one can calculate the total electronic charge contained in each building block, consider it as concentrated at the center or smeared out on the correspondent atomic or interstitial planes, and then calculate analytically

the electrostatic potential lineup due to such charge distribution. This procedure may be useful just for a first prediction of the electrostatic potential lineup between two given semiconductors, once one has tabulated the electronic charges contained in the building blocks for each of them.

The role of the model theories: successes and limits

The various attempts of establishing an absolute energy scale can be seen as proposals of different "relative reference frames". Some theories refer to a reference level "external" to the system, e.g. the ionization potential or the electron affinity, others to a reference level "internal", as for instance the neutrality level in the Tersoff's model.

Some theories emphasize the importance of the charge distribution at the interface, and some others completely neglect it. The physical concepts, such as "reference heterojunction" or "reference surface", "interface dipole", "zero–dipole condition", "charge transfer", are in many theories ill–defined; the concept of "interface dipole" for instance was always related to a suitably but arbitrarily chosen reference charge distribution, and for this reason different theories emphasizing or neglecting the role of interface dipoles have given equivalent results.

As far as the description of the interface features is concerned, and in general the contribution of the model theories to the physical understanding of the problem, I think it is important to further examine here the WSC model in its original formulation and successively in its generalization, since it qualitatively correctly describes, although with few ingredients, what happens at the interface. It is actually extremely important that this model allows and well reproduces the *polarization of the interface ions* surrounded by different nearest neighbors. To better understand this mechanism I refer to the simple case of GaAs/AlAs, where only the interface As anions are polarized: the situation is visible in Fig. 3.11, where also the correspondent picture obtained with full SCF interface calculations is reported for comparison. It is thus clear why any other attempt to define different building blocks, neglecting the possibility of polarization of the interface ions has been completely unsuccessful. Fig. 3.12 shows for instance the differential plot of the charge density at the GaAs/AlAs heterojunction, obtained with the model using Wigner–Seitz BCC cells as elementary bricks, with respect to the charge density of the average virtual crystal: comparison with Fig. 3.11(a) and (c) indicates that the relevant mechanism responsible of the interface dipole, i.e. the polarization of the interface anion, is completely neglected by such model.

Perhaps the empirical optimization of the "reference level" theories rewritten by Katnani and Margaritondo[49] is the best version of them and of all the model theories from the point of view of the accuracy: they use the VBO measured by soft x–ray synchrotron radiation photoemission and a least–squares method to estimate the absolute positions of the valence band edges, taking the germanium valence band as

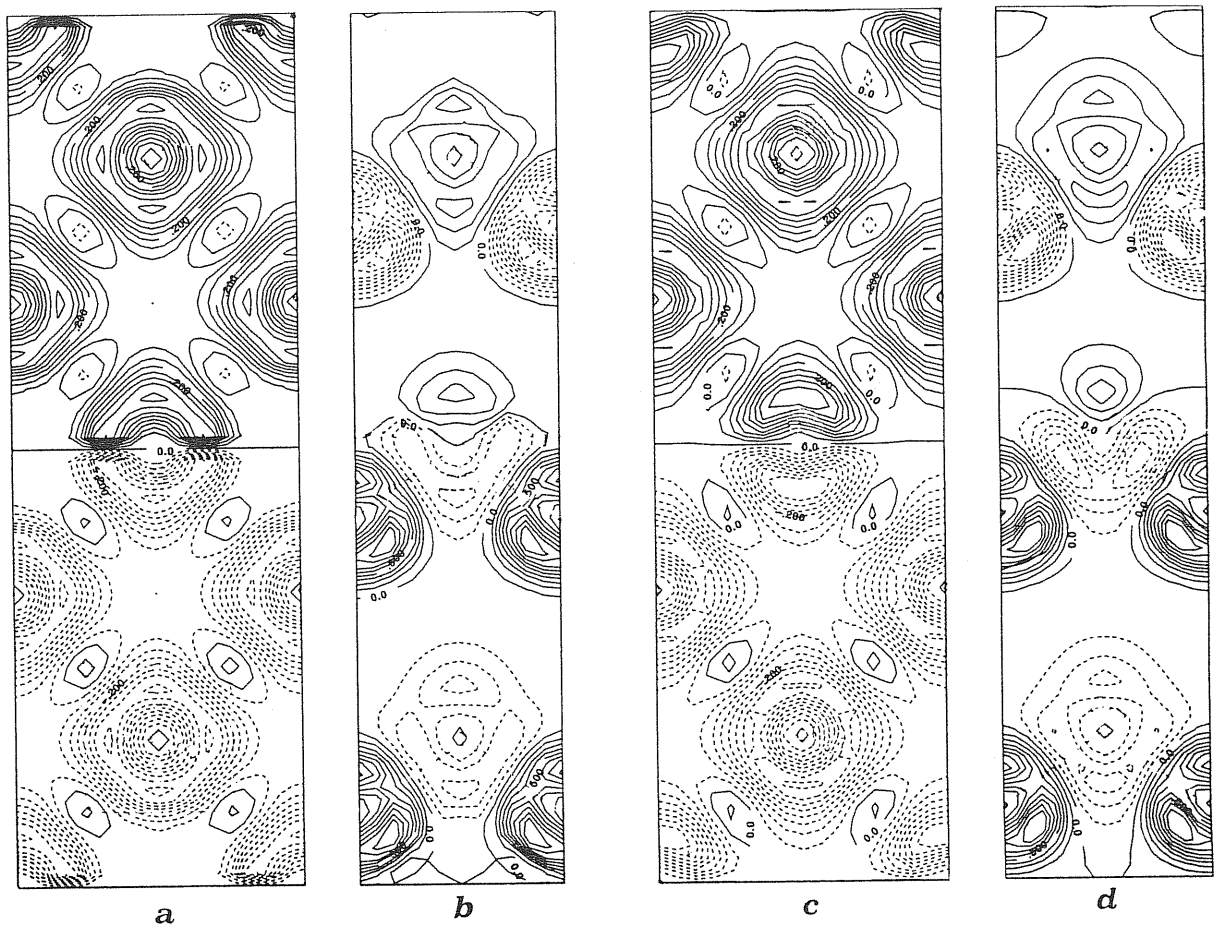

Figure 3.11. Differential plot on the (010) and (110) planes of the electron density distribution of GaAs/AlAs (001) heterojunction with respect to that of a reference periodic crystal, obtained according the WSC model (a,b) and with full SCF interface calculations (c,d). The reference periodic crystals used in the two cases are slightly different: in (a,b) the electron density is the average of those of two bulks, according to the procedure indicated in the description of the model, whereas in (c,d) it is obtained self-consistently for a fictitious crystal where the cations are equal to the virtual one $\langle Ga_{0.5}Al_{0.5}\rangle$.

the arbitrary zero. The results are shown in Tab. 3.1. Unfortunately the completely empirical nature of such a theory does not contribute to the physical understanding of the problem.

Successes and failures of the oldest and simplest model theories have however contributed to focus finally one problem: a detailed investigation of the interfacial charge distribution cannot be *a priori* neglected. In almost all the existing model theories —also where the physical concepts are well defined— the description of the interface region is quite rough, and the accuracy of the results poor. More generally, up to now the crucial question whether the band offset is a bulk or a specific–interface property has actually not yet found a precise answer.

Perspectives beyond the model theories

The importance of the model theories must not be underestimated, since on

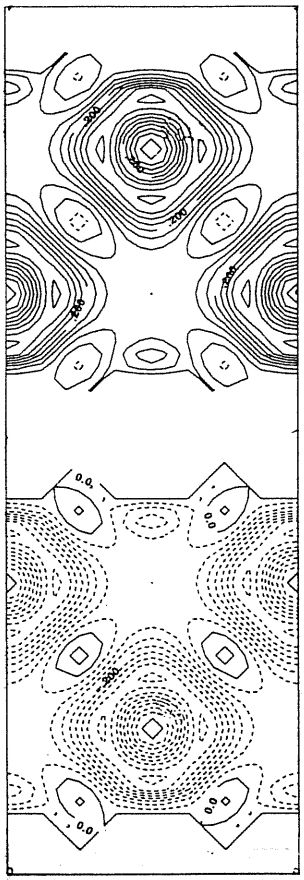

Figure 3.12. Differential plot on the (010) plane of the electron density distribution of GaAs/AlAs (001) heterojunction with respect to that of the reference average periodic crystal, obtained according the model using Wigner–Seitz BCC cells. The discontinuities in the electron density due to the rigid juxtaposition of the building blocks are clearly visible.

Material	Empirical position in energy of the valence-band edge (in eV)	Material	Empirical position in energy of the valence-band edge (in eV)
Ge	(Reference)	CdS	−1.74
Si	−0.16	CdSe	−1.33
α-Sn	+0.22	CdTe	−0.88
AlAs	−0.78	ZnSe	−1.40
AlSb	−0.61	ZnTe	−1.00
GaAs	−0.35	PbTe	−0.35
GaP	−0.89	HgTe	−0.75
GaSb	−0.21	CuBr	−0.87
InAs	−0.28	GaSe	−0.95
InP	−0.69	$CuInSe_2$	−0.33
InSb	−0.09	$CuGaSe_2$	−0.62
		$ZnSnP_2$	−0.48

Table 3.1. Empirical estimate of the positions of the valence–band edges in several semiconductors with respect to Ge, according to the fitting scheme of Katnani and Margaritondo[49]: the difference between two terms gives a first–order estimate of the discontinuity VBQ

one side it is very useful to have some simple recipes for a rapid estimates of the band discontinuities in several systems, and, on the other side, I do not exclude that a further progress in the model theories will shed more light on the mechanisms governing the band offsets.

It appears clear the necessity of a large theoretical and experimental effort to ascertain some still obscure points; as far as the theoretical approach is concerned, it must be improved in particular with its roles of i) predictive power and ii) conceptual scheme for the understanding of the experimental evidences and in general of the phenomenum. To this purposes it may be useful both a systematic study and a more precise investigation of some prototype systems.

These are the general motivations which urge to perform more accurate self–consistent ab–initio calculations. In particular one should try:

i) to overcome the limits of applicability to restricted classes of heterojunctions, which are present in several models;

ii) to prove the validity of the various model theories, to understand the implications of the approximations involved and to estimate *ab–initio* their intrinsic accuracy limit: a comparison restricted to the experimental work or among the model theories themselves cannot solve this point;

iii) to go beyond the crude and simple description of the interface which is the only possible in the model theories: in principle it should be possible, starting from a given structural and stoichiometric configuration at the interface, to give accurately the electronic configuration and hence the band discontinuities; intrinsic limits are due to the incomplete knowledge of the atomic–scale crystallographic structure in complicated systems (not only interfaces, but also alloys...);

iv) to analyze quantitatively the possible effects on the VBO due to different structural details (orientation dependence, atomic displacements at the interface, deviations from abruptness, atomic order and possible internal relaxation in the bulk region of the two materials, introduction of isolated foreign atoms or thin intralayers in the interface region); this would help the general understanding of the problem and the identification of the main factors affecting the VBO, and would also give an explanation of the discrepancies between different experimental accurate data;

v) to treat in the same way all the semiconductors, without any fitting of experimental data as it occurs in the self–consistent calculations which make use of empirical pseudopotentials; the *ab–initio* procedure would give to the theory predictive power.

Chapter 4

FIRST–PRINCIPLES SELF–CONSISTENT CALCULATIONS

Most of the model theories I have examined before rely on information on the individual bulks alone, and do not provide a description of the actual electron distribution at the interface. In order to obtain a realistic picture of interface effects one has to perform calculations in which the electrons are allowed to readjust to the specific environment determined by the different ionic species at the heterojunction. Density functional theory (DFT) provides a fundamental theoretical framework to face this problem, with the reduction of the many–body problem to the single–particle problem.

I present in this Chapter the general features of the DFT approach together with some general tools necessary to solve in practice the single–particle problem (e.g. local–density approximation, pseudopotentials, plane–wave basis set, special–point technique) and some others specific (e.g. supercells) for an actual application of the method to the heterojunctions.

Even if the development of super–computers has only recently allowed full self–consistent interface calculations, a lot of work has been already done and it is reviewed in the present Chapter. It supports some trends (i.e. interface orientation independence and transitivity of the VBO for a wide class of systems) appeared from the existing experimental data, but new first–principles calculations will be necessary for a complete understanding of the band–offset problem.

4.1 Formulation of the basic theory and methods

Density functional theory

The Schrödinger equation describing a system of N electrons in an external potential $V(\mathbf{r})$ is

$$H\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = (T + V + U)\Psi = E\Psi \tag{4.1}$$

Even if one is interested in the ground state only, it is practically impossible to find a solution of it, since he has to treat a wavefunction of 3N coordinates. A revolutionary idea, formally justified by Hoemberg and Kohn [50] was the introduction of the electron density $n(\mathbf{r})$ —which is a function of three coordinates only— as the basic variable: they demonstrated that the external potential $V(\mathbf{r})$ (apart from an irrelevant additive constant) is uniquely determined by the ground state density $n(\mathbf{r})$, and then any property of the system can be considered as a functional of n rather then of V. Among these properties, the ground state Ψ. One can introduce a variational principle directly related to the density $n(\mathbf{r})$. In fact it is well defined a universal functional

$$F[n] = \langle\Psi|(T + U)|\Psi\rangle \tag{4.2}$$

which contains the kinetic energy and the electron–electron interaction contribution to the total energy of the system. For a particular external potential $V(\mathbf{r})$ we then introduce the functional

$$E_V[n] = \int V(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n] \tag{4.3}$$

which has the minimum value, corresponding to the total energy of the system, when $n(\mathbf{r})$ is the ground–state charge density.

Separating in the functional $F[n]$ the Hartree term, due to the classical electron–electron interaction, and a kinetic term $T[n]$ corresponding to a system of non interacting electrons with the same density, one rewrites Eq.(4.2) as:

$$F[n] = T[n] + \frac{e^2}{2}\int d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[n] \tag{4.4}$$

where $E_{xc}[n]$ is a residual functional which contains the so–called "exchange–correlation energy" contributions, formally well defined but unknown. The solution of the Schrödinger equation for the system of N electrons (Eq.(4.1)) is reduced to the minimization of the total energy functional $E_V[n]$ with respect to $n(\mathbf{r})$ under the condition $\int n(\mathbf{r})d\mathbf{r} = N$. This leads to a single–particle equation, first derived by Kohn and Sham [51]:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) + \mu_{xc}\right)\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}) \tag{4.5}$$

where

$$V_H(\mathbf{r}) = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}', \tag{4.6}$$

$$\mu_{xc} = \frac{\delta E_{xc}}{\delta n(\mathbf{r})}, \tag{4.7}$$

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2, \tag{4.8}$$

and the sum covers the lowest N energy eigenstates of Eq.(4.5).

Local–density approximation

The application of DFT to a real system requires an approximation for the exchange–correlation energy $E_{xc}[n]$ and its functional derivative $\mu_{xc}(\mathbf{r})$. The most commonly used approximation is the so called local–density approximation (LDA) which assumes that the exchange–correlation energy per particle is locally equal to the one of an ideal uniform electron gas with the density equal to the density of the real system in the point considered:

$$E_{xc}^{LDA}[n] = \int \varepsilon_{xc}\left(n(\mathbf{r})\right) n(\mathbf{r}) d\mathbf{r} \tag{4.9}$$

where $\varepsilon_{xc}(n)$ is the exchange–correlation energy per particle of the uniform electron gas; eq.(4.10) gives directly the exchange–correlation potential:

$$\mu_{xc}^{LDA}(\mathbf{r}) = \frac{\delta E_{xc}^{LDA}[n]}{\delta n(\mathbf{r})} = \frac{d}{dn}(\varepsilon_{xc}(n)n)|_{n=n(\mathbf{r})} \tag{4.10}$$

The LDA is exact in the limit of uniform density, and it is a good approximation in the case of sufficiently slowly varying density. Application to a number of different systems, i.e. atoms, molecules and solids, has shown that it gives good results even if the density is not slowly varying.[52] Up to now we have seen how in the framework of DFT the many–body problem is reduced to the treatment of one–body equations; for their solution powerful tools have been developed.

Reciprocal–space formalism and plane wave basis set

In several cases a formulation of the density functional theory in reciprocal space[53] is advantageous.

The actual implementation of Kohn–Sham equations (Eq. (4.5)) requires the choice of a finite basis set to develop the wavefunctions. It is very convenient to use as a finite basis set all the plane waves (PW) corresponding to a kinetic energy up to a certain cutoff:

$$|\mathbf{k}+\mathbf{G}\rangle = \frac{1}{\Omega} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \qquad \text{with} \quad \frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{G})^2 \leq E_{cut} \tag{4.11}$$

where k is in the Brillouin zone (BZ) and G is a reciprocal space vector. In this way the accuracy can be improved by increasing the kinetic energy cutoff and consequently the dimension of the basis set. In particular the expressions of the kinetic energy and of the electrostatic energy are straightforward in this formulation.

Pseudopotentials

An accurate description of core electrons would require the use of very attractive potentials, singular at the origin; the corresponding wavefunctions show strong oscillations and need a large PW basis set to be described. In a solid, however, the core electrons are localized as in the isolated atom configurations; they are well separate in energy from all the valence states which are responsible of the formation of the chemical bonding and which determine the majority of the properties of the solid.

The basic idea of the pseudopotential theory [54] is to avoid an accurate description of the core electrons; weaker (not singular) effective potentials, taking into account the orthogonality between valence and core states, are used to describe the valence pseudowavefunctions, which are smooth and require a reasonably small PW basis set.

Basically the pseudopotential concept has been developed and implemented in two different ways: in the first one, very simple atomic pseudopotentials are empirically fitted to reproduce experimental energy bands,[55] but the validity is of course limited to a precise atomic configuration and such pseudopotentials are not able to describe the same atom in other environments. In the other way the pseudopotentials are generated using only theoretical calculations on atoms,[56,57] without introducing any fitting of experimental band structures or other properties (they are thus referred to as "first–principles ab–initio" pseudopotentials). Recently developed pseudopotentials, in particular those tabulated by Bachelet, Hamman and Schlüter[57] —which I have actually used in my calculations—, have the following properties:

1) real and pseudo eigenvalues for occupied states are the same;
2) real and pseudo valence wavefunctions (and hence real and pseudo charge densities) agree beyond a chosen "core radius" r_c;
3) the integrals from 0 to r of the real and pseudo charge densities for each valence state agree for $r \geq r_c$ (norm conservation);
4) the logarithmic derivatives of the real and pseudo wavefunctions and their first energy derivatives agree for $r \geq r_c$.

Note that from the property 3) follows that, by the Gauss's theorem, the electrostatic potential produced outside r_c is identical for real and pseudo charge distributions. Contrary to all–electron potentials, such pseudopotentials are not singular at the origin, and they are "soft" enough to be well described in a Fourier analysis with a reasonable number of plane waves: this is very important when one has to deal with several atoms, as in the case of heterojunctions.

Another important aspect of such pseudopotentials is that —since they do not contain any empirical data— they provide the appropriate tool for calculations giving reliable predictions in different fields, whereas with the empirical potential the predictive power is limited to cases very similar to the ones used to extract the fitting parameters. Moreover all the elements are treated in the same way, and this is also very important in the study of the heterojunctions where one treats together different kinds of materials.

The common used pseudopotentials are non–local, in the sense that they contain a term which explicitly depend on the angular momentum l of the wavefunction on which it acts:

$$V(\mathbf{r}) = V^L(\mathbf{r}) + V^{NL}(\mathbf{r}) = V^L(\mathbf{r}) + \sum_l V_l(\mathbf{r})P_l \tag{4.12}$$

where the superscripts L and NL mean *local* and *non–local* respectively, and P_l is the projector on the angular momentum l.

Heterojunctions and supercells

The heterojunctions are actually described by periodically repeated supercells, containing two complementary interfaces. I want to stress that supercells allow a reciprocal space formulation of the problem and the use of PW, otherwise not possible because of the loss of translational symmetry.

The supercells actually describe a multilayered structure, and the length scale may be of the order of magnitude of the one of SL. But as far as the energy–band discontinuities are concerned, the experience has shown that the relevant effects due to the presence of a sharp neutral interface are confined in a small region, and bulk features are completely recovered within few atomic units from the interface (see for instance Fig. 2.6): this implies that the main interface features can be studied using supercells with a reasonably small number of atoms, and the isolated interface configuration is well represented, provided that the two adjacent interfaces are sufficiently spatially separated in order not to interact.

Special–points technique

The integral of a periodic function over the BZ is in practice reduced to a discrete sum over a set of $\mathbf{k}$ points. There are different techniques to generate a set of representative $\mathbf{k}$ points.[58,59] We follow here the mean value point technique introduced by Baldereschi[59], then generalized by Chadi and Cohen[60], Monkorst and Pack[61], and recently recovered by Froyen[62] and Dow[63] and applied to the case of superlattices. Let us consider a function $f(\mathbf{k})$ which is totally symmetric and periodic in $\mathbf{k}$–space. It can formally be expanded into symmetrized linear combinations of plane waves of symmetry Γ_1 as follows:

$$f(\mathbf{k}) = f_0 + \sum_{m=1}^{\infty} f_m G_m(\mathbf{k}) \tag{4.13}$$

where

$$G_m(\mathbf{k}) = \sum_{\mathbf{R}\in star} e^{i\mathbf{k}\cdot\mathbf{R}} \tag{4.14}$$

and the star is that of the equivalent lattice vectors related to each other through the operations of the lattice point group. Since

$$\frac{\Omega}{(2\pi)^3}\int_{BZ} G_m(\mathbf{k})d\mathbf{k} = 0 \quad \text{for} \quad m = 1, 2, ...\infty \tag{4.15}$$

we have that

$$\langle f\rangle = \frac{\Omega}{(2\pi)^3}\int_{BZ} f(\mathbf{k})d\mathbf{k} = f_0 \tag{4.16}$$

If it existed a point $\mathbf{k}^*$ such that

$$G_m(\mathbf{k}^*) = 0 \quad \text{for} \quad m = 1, 2, ...\infty \tag{4.17}$$

we would immediately have

$$f_0 = f(\mathbf{k}^*) \tag{4.18}$$

Such a point does not exist, but we can find one satisfying Eq. (4.17) for $m = 1, 2, ..$ up to a certain finite N. In general one selects a certain set of k points, say $\{ \mathbf{k}_1^*, \mathbf{k}_2^*, \ldots \mathbf{k}_n^* \}$ and relative weights w_i (normalized to 1), such that:

$$\sum_{i=1}^{n} w_i G_m(\mathbf{k}^*{}_i) = 0 \quad \text{for} \quad m = 1, 2, ...N \tag{4.19}$$

so that a good approximation to the average $\langle f\rangle$ is given by:

$$\langle f\rangle \approx \sum_{i=1}^{n} w_i f_m(\mathbf{k}_\mathbf{i}^*) \tag{4.20}$$

Such sets of k points form uniform grids covering the BZ, the choice depending on the desired accuracy.

SCF cycles and diagonalization techniques

The Kohn–Sham equations (Eq. (4.5)) must be solved numerically in a self–consistent cycle. The first iteration requires a trial potential: several choices are possible, e.g. a ionic potential screened by the dielectric function of a free–electron gas, or the potential corresponding to the superposition of free-atom charge densities. Different mixing schemes to update the potential are also possible.

A very technical but important point is the diagonalization of the matrix $\langle\Psi|H|\Psi\rangle$, very expensive as far as the computer time is concerned: standard diagonalization schemes require a number of operations scaling as N^3, where N is the dimension of the matrix. Substantial improvements have been obtained with non standard diagonalization schemes: e.g. if only the first M eigenvectors of a very big matrix are required, the application of some non standard diagonalization schemes that scale as N^2M allow to significantly reduce the consumed computer time.

4.2 Existing first–principles SCF calculations

Van de Walle and Martin

A systematic study of the band–offset at semiconductor interfaces has been done by Van de Walle and Martin[37]: they perform SCF supercell calculations simulating abrupt interfaces with all the atoms in their ideal zincblende or diamond structure positions. In order to extract some general features of the lineup mechanism, they study systematically a number of lattice–matched heterojunctions between group IV, III-V, and II-VI compounds. They extract the average level of the supercell SCF potential (sum of ionic, Hartree, and exchange–correlation potentials) in the bulk regions on the two sides of the heterojunction;(*) they obtain finally the VBO by adding the band structures calculated for the bulk materials to the potential shift. Spin–orbit splitting effects on the bulk valence–bands are added *a posteriori* from experimental data. I report in Tab. 4.1 their results for the systems studied here, together with those obtained by other authors.

	(a)	(b)	(c)	(d)	(e)	(f)
$AlAs/GaAs$	0.37	0.53	0.42	0.50	0.54	0.45
$GaAs/Ge$ (110)	0.63	0.46			0.45	0.56
$AlAs/Ge$ (110)	1.05	1.03			1.07	1.07
$InP/Ga_{0.47}In_{0.53}As$	0.35					

Table 4.1. Self–consistent interface calculation for the VBO (in eV) performed by different authors. The various columns refer to: (a) Van de Walle and Martin[37], (b) Christensen[64], (c) Wei and Zunger[65], (d) Massidda, Min, and Freeman[66], (e) Lambrecht and Segall[67], (f) Bylander and Kleinman[68].

(*) They do not use the concept of macroscopic average, but calculate the average potential in the central layer of the two bulk regions.

They estimate a numerical accuracy of 0.05–0.10 eV on the final result, the main sources of inaccuracy being the kinetic energy cutoff (6 Ryd for supercell calculations, 12 Ryd and in some cases 18 Ryd for bulk calculations), and the small number of atoms (8 or 12) in the supercell; they argue that possible errors introduced by the use of LDA both in the bulk band structure calculations and in the reference potentials obtained from supercells calculations do not significantly affect the VBO of the systems studied, as long as the semiconductors constituting the heterojunction are similar.

The authors also perform DFT–total energy calculations for GaAs/AlAs in order to derive the minimum–energy structure: they find that the ideal structure (i.e. all the atoms occupying zincblende or diamond structure sites) "is very close to the minimum–energy configuration, with very small forces acting on the atoms", inducing possible changes in the atomic positions smaller than 0.06 a.u.. The authors assert that in general in non–polar interfaces displacements of the order of 0.1 a.u. may occur, but their effects on the VBO are within the numerical accuracy of the results.

Let me make here two comments:

i) I agree with this sentence as far as the GaAs/AlAs is concerned, and in general for all the heterojunctions constituted by elemental or binary semiconductors having all the bond–lengths equal. This is not true in general, as for instance in InP/GaInAs, where different bond lengths are present: in such a case, the determination of the actual atomic positions is very delicate, and the assumption of ideal bulk geometry is a serious approximations and a source of uncertainty on the final result. Moreover, looking at atomic displacements of the order of 0.1 a.u. has no significance, since the fictitious theoretical lattice–mismatch induced by the use of pseudopotentials is sometimes even greater, of the order of 0.1–0.2 a.u..

ii) The second comment concerns the choice of the kinetic energy cutoff in performing the band calculations for the bulk materials: the authors say that "it is not critical for deriving the valence–band lineups", but I will show conversely that VBO for GaAs/AlAs system is very sensitive to this parameter.

An important trend which has come out from the systematic work of Van de Walle and Martin is a general independence of VBO on the interface orientation for "a wide class of nonpolar interfaces", as well as the transitivity rule, which is satisfied to better than 0.06 eV, i.e. within the numerical accuracy of the calculations. Taking into account these trends, Van de Walle and Martin suggest the possibility of deriving the lineups by determining an absolute reference level for each semiconductor, and lining up the bulk band structures according to these reference levels: the model–solid theory proposed by the same authors was illustrated in Ch. 3.2.

Christensen

Alternative first–principles self–consistent calculations for lattice–matched semiconductor heterostructures are those performed by Christensen and cooworkers[64]. They use the relativistic linear–muffin–tin–orbital (LMTO) method applied in the supercell geometry. More precisely, the procedure followed can be schematized in these steps: i) they use the atomic–sphere approximation (ASA) (i.e. space filling is obtained with slightly overlapping spheres centered on atomic positions and containing spherical potentials); ii) they perform supercell self–consistent calculations in the framework of DFT–LDA; iii) they extract from each of the two bulk regions in the supercell the central–layer potentials and places these on a bulk structure; iv) in this structure they perform a single bulk–band calculation ("frozen–potential calculation") for each of the two materials constituting the heterojunction, and finally v) they determine the VBO by subtracting the two bulk valence–band maxima. Possible non–sphericity of "atomic" potentials in the heterojunction and intra–atomic polarization (which is very important in the heterojunction problem, as discussed in Sect. 3.4), are taken into account by non–equivalent empty spheres centered in the tetrahedral interstitial positions.

The small basis set needed in LMTO calculations allows computations for supercells bigger than those used in PW pseudopotential method. In particular, Christensen considers cells with up to seven layers on each side (i.e. 28 atoms) and studies the variation of the VBO with respect to the size of the supercell. He finds that in general "a 5+5 supercell is sufficiently large to produce a reliable offset value". He finds important to include the outermost cation d–like states, which hybridizing with the valence–band maximum influence the calculated VBO value, and the higher conduction d states, which has opposite effects; the final VBO is the average of two calculations separately performed for each heterojunction, with and without the higher d states. These LMTO calculations show that generally the VBO is independent on the interface orientation; transitivity rule is satisfied, a part from significant deviations found for the group including CdTe, HgTe and CuBr semiconductor compounds.

The model previously proposed by the same author (see Ch. 3) does not include the effects of charge rearrangements at the interface. In order to estimate the magnitude of these effects and to test the validity of his model, the author calculates the quantity ΔE_{dip}, defined as the difference between the VBO obtained from the model and from the supercell SCF calculations: these dipole corrections are negligible e.g. for GaAs/AlAs ($\approx$ 0.03 eV), and large if the ionicity of one of the two semiconductors is large (e.g. $\approx$ 0.3 eV for Ge/GaAs). The author then extracts from the supercell calculations the "electron transfer" $\Delta N^{L\rightarrow R}$ from the left side to the right side of the interface, and notes —apart from a few exceptions— a trend indicating that $\Delta N^{L\rightarrow R}$ is larger when ΔE_{dip} is.

Moreover, the charge transfer $\Delta N^{L\to R}$ is generally proportional to the difference ΔV_{MT} between the average muffin–tin potentials on the two sides of the heterojunction. This idea of relating electron transfers as differences between the individual semiconductor potentials, as well as the one relating deviations from the transitivity rule to interface–specific features, are promising, but no further consequences are derived. Moreover, the concept of interface dipole is here ill–defined: the author considers in fact an "interface layer" formed by the last layer of the left–side semiconductor and the first one of the right–side, and then gives a physical meaning to the "excess charge" (with respect to neutrality) in the so–defined interface layer, as illustrated in Fig. 4.1. This is, in my opinion, an arbitrary quantity, since a thicker interface layer, considered for instance from the central layer on the left side up to the one on the right side, necessarily guarantees charge neutrality and no excess charge can be extracted from it. However, the author associates the presence of this excess charge with the deviations from the proportionality between ΔV_{MT} and $\Delta N^{L\to R}$ (he distinguishes between "normal" and "anomalous" charge transfer according if there is proportionality or not), and with the formation of localized interface states. This explains, according to the author, the anomalous results for some heterojunctions containing CuBr, CdTe and HgTe compounds; he argues that "there are certain semiconductor compounds that produce interface specific electronic states that influence the offset", and concludes that "the question whether the band offset is a bulk property ... cannot be answered in general".

Wei and Zunger

The role of d–orbitals in the VBO is emphasized also by Wei and Zunger[65] in a work which is limited to common–anion heterojunctions. They estimate the absolute positions of the energy bands with respect to the deep impurity levels.

They calculate self–consistently the band structures of ZnTe, CdTe, HgTe, AlAs and GaAs, treating core states relativistically and using the general–potential linear augmented plane–waves method in LDA.

They divided the VBO into the "bulk" contribution ΔE^{b}_{VBM} (the correspondent of ΔE_v) reflecting the disparity between the valence–band maxima in the two bulks, and the "interface specific" contribution ΔE^{is}_{VBM} (like ΔV) which reflects chemical events at the interface, but it is not clear the role of ΔE^{is}_{VBM}, being referred to the ill–defined concept of interfacial charge–transfer (screening) effects. The authors argue that the difference between the valence–band edges in two common–anion binary semiconductors should contain the contribution of *cation* orbitals; they observe from self–consistent band calculations an hybridization of cation d character in the Γ_{15v} valence–band maximum state, which can in fact exist because in the tetrahedral symmetry both the cation d and anion p states have a Γ_{15} symmetry component and hence can interact through the mixed potential matrix elements. The authors conclude that the omission of cation d–orbitals rather than interface specific effects

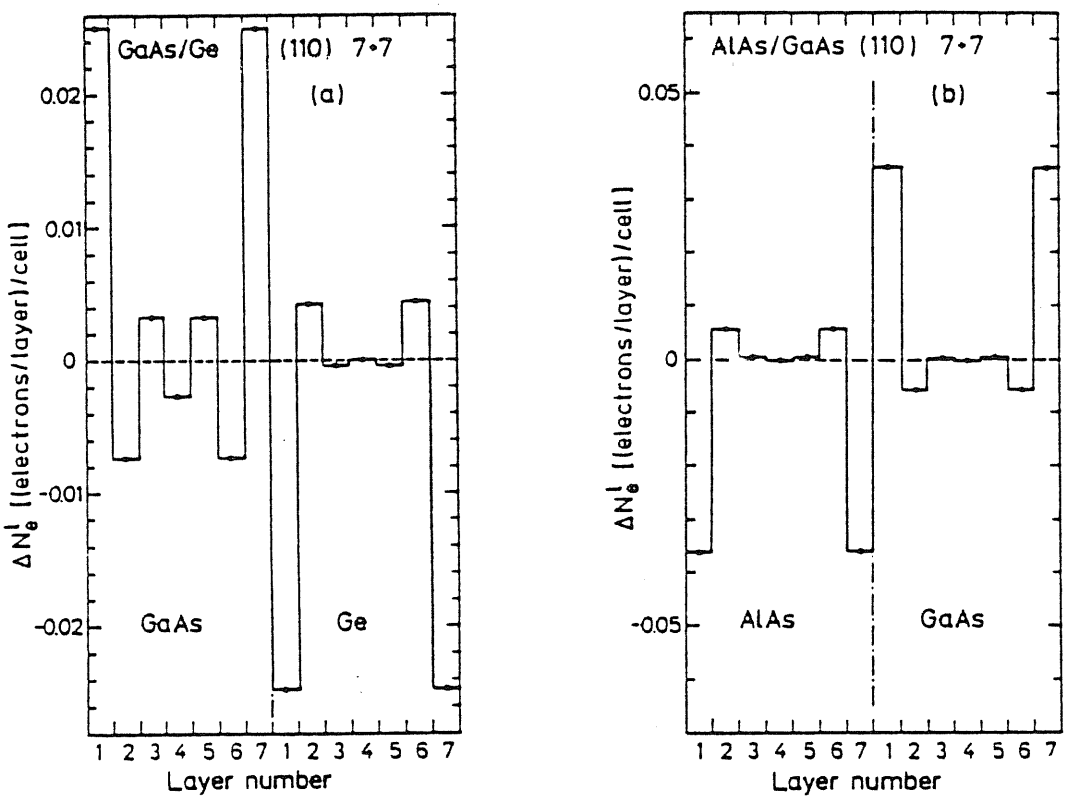

Figure 4.1. Excess electrons on each atomic layer of the 7+7 supercells calculations performed by Christensen for GaAs/Ge (a) and AlAs/GaAs (b).

is responsible of the partial disagreement of the tight–binding calculations with the experimental data and of the failure of the so–called "common–anion rule" which would predict small values of VBO.

Massidda, Min, and Freeman

The work of Massidda, Min, and Freeman[66] on $(GaAs)_n(AlAs)_n$ (001) superlattices contributes to the debate about the role of the shallow cation d–states. The authors perform self–consistent full–potential linearized augmented–plane–wave (FLAPW) calculations on the two bulk semiconductors and on the supercell; band energies are calculated semirelativistically, whereas the core states are calculated fully relativistically.

They use the core levels as reference energies to determine the relative alignment of the valence–band edges, similarly to the procedure followed in the X–ray photoemission measurements: the binding energies of the selected core levels relative to the top of the valence bands ($E^b_{c_1}$ and $E^b_{c_2}$) are derived from the bulk band structures, whereas the superlattice calculation gives the binding energy differences (ΔE_B) of the same core levels on the two sides of the interface, so that $VBO = E^b_{c_1} - E^b_{c_2} - \Delta E_B$. The authors find VBO=0.50±0.05 eV using anion core levels as reference energies, and notice that this result is sensitive to the different treatment of the Ga $3d$ states (i.e. as core or band states).

Studying the variation of the VBO when passing from $n = 1$ to $n = 2$ and 3, they find that the $n = 1$ value give an offset of $\approx$0.11 eV lower, and relate this to the interface charge redistribution. More precisely they calculate the dipole potential $\Delta V_{dip.}$ from the first moment —in one dimension— of the charge density planar average with respect to the As–interface plane: they find $\Delta V_{dip.} \approx 0.14$ eV when the first moment is calculated in a region extending from the last As–plane on the left side to the first one on the right side with respect to the interfacial one, and relate it to a a charge transfer induced by the electronegativity difference. However, the charge transfer is still ill–defined, being obtained by subtracting from the supercell charge density the correspondent bulk charge density on the two sides of the junction, and no relationship is established between $\Delta V_{dip.}$ and the VBO.

Lambrecht and Segall

Lambrecht and Segall[67] carry out supercell band–structure calculations starting from the LMTO method with the ASA. The ingredients are similar to those in the work of Christensen, but here self–consistency treatment is very limited. The authors start from standard bulk band–structure calculations of the individual semiconductors; the individual bulk solid ASA potentials are on an absolute energy scale due to a well defined choice for charge density partitioning: the difference between them gives a first contribution to the VBO, but then self–consistent cycles are required to obtain the supercell potential starting from the bulk ASA potentials. This procedure —which is performed here in a "frozen shape" approximation, allowing only a constant potential shift layer by layer— causes a charge transfer which again has no precise physical meaning because of the arbitrariness of the reference situation (the authors have well in mind this point), but however provides a correction to the VBO. The "dipole correction", obtained by subtracting to the so–calculated VBO the ASA term, is essentially dielectrically screened.

Bylander and Kleinman

Other self–consistent semirelativistic *ab initio* pseudopotential calculations for interfaces between GaAs, AlAs and Ge have been performed by Bylander and Kleinman,[68] who have investigated also their formation enthalpy. They construct fully relativistic norm–conserving s, p, and d ionic non–local pseudopotentials for Ga, Al, and As, and they use a Gaussian orbital expansion of the wavefunctions.

The supercell calculation provides the crystal potential average values in the two bulk regions(*), and the valence–band offset is obtained by adding to them the energy

(*) The authors show that the final result for VBO does not change sensitively if the average values of the crystal potential are taken from the central slab cell (bulk unit cell) or from the single central plane; however, they consider the average of the VBO calculated in the two ways as the best estimate.

of the top of the valence band at Γ relative to the average crystal potential in the bulk semiconductors.

The authors then consider the difference between the planar average of the supercell charge density and that of its constituents, and refer to this discontinuous differential charge density profile as the interface dipole; they extract from it a Coulomb potential difference $\Delta_{dip.}$, to which is added the difference between the bulk valence–band top edges calculated fixing the arbitrary zero of Coulomb potential as the averaged potential from cutted slabs of bulk charge densities.

The authors claim an extremely high accuracy of their results —something of the order of 1 meV—, giving for the VBO the values of -0.447 eV for GaAs/AlAs (001) and (110), -0.558 eV for GaAs/Ge, and -1.073 eV for AlAs/Ge, thus finding a failure of the transitivity rule that they attribute to the "large diamond–zincblende interfacial mismatch".

4.3 Open problems

Conclusions from the existing theories

Some limits due to the excessive simplicity of most of the model theories have been overcome in SCF ab–initio calculations. Let me summarize here some achievements:

i) There is no restriction on the class of materials that can be treated, the limits being those of the standard bulk calculations with some further computational problems due to the large number of atoms and electrons in the supercells that have to be used.

ii) The results of self–consistent calculations performed with different methods for the same system are not the same; however the quoted accuracy is often better than the experimental one, and the spread of the theoretical results is smaller for a given interface. Because of this, the first–principles self–consistent calculations provide a reliable reference for model calculations, even where experimental data do not exist, or are very contradictory.

iii) Once the interface structure and stoichiometry are fixed, a full interface self–consistent calculation provides the electronic structure of the heterojunction in all its details.

iv) In principle it is possible to treat atomic displacements or substitutional defects at the interface; moreover, the structural equilibrium configuration of the interface can be predicted from first–principles. The estimate of the effects of the interface details on the VBO is thus possible, although only Van de Walle and Martin have presented results in this direction.

v) Finally, it is clear the predicting power of ab–initio self–consistent calculations, which avoid any empirical input.

Some general trends come out expecially from the accurate and systematic work of Van de Walle and Martin: for instance the transitivity rule is found to be generally satisfied within the quoted accuracy of the results; however the controversy is not yet solved (see e.g. Bylander and Kleinmann for GaAs/Alas/Ge). Another important point is the independence of the VBO on the interface orientation: this seems to be a general property of all isovalent heterojunctions, although some authors do not agree on this point: however, the claimed deviations are very small. The calculations suggest that the deviations from this "linear" property has to be searched in the interface–specific features (see e.g. Christensen[64]).

Open questions and necessity of new first–principles calculations

The crucial question whether the VBO is a bulk or an interface property has not yet received a precise answer. The trends resulting from the existing first–principles self–consistent calculations suggest that for a wide class of systems the

band discontinuity should be related to differences between quantities intrinsic to the bulks, but it should still be understood "to what" and "why"; furthermore, the quantitative effects on the VBO of the various interface details have still to be investigated.

A systematic work including a lot of different materials at this point is no longer necessary; it seems rather more instructive to study some prototype systems in order to isolate for which classes of materials some properties hold and others do not, and thus to identify the important physical parameters. Note that it is convenient to this purpose the precise separation of the VBO into a pure bulk quantity (ΔE_v), and another one (the electrostatic potential ΔV) which contains all the possible interface–specific contributions. The accurate analysis of the nature of the electrostatic potential lineup (e.g. what is the mechanism which makes it different from zero, how is it influenced by atomic displacements near or far from the interface) is the subject of the new first–principle calculations that I will present in the next Chapter.

Chapter 5

NEW FIRST-PRINCIPLES RESULT

I explain in the first part on the Chapter the technical ingredients used in the present original work, discussing in some details the parameters chosen and the approximations made.

Results are then shown for some prototype systems which represent different classes of heterojunctions, giving a first hint of the essential parameters governing the mechanism of the band discontinuities. It comes out that the band offset at isovalent interfaces (both common-anion and no-common-ion) is equal for the three main crystallographic orientations, and the same holds separately for the cationic and anionic contributions in which —in the general case of heterojunctions with no common ions— the VBO can be exactly decomposed. For heterovalent interfaces, the offset depends on the details of the atomic arrangement at the interface; however, once this arrangement is known, the interface-dependent contribution to the offset can be calculated "on the back of an envelope" from such simple (and experimentally accessible) quantities as the bulk lattice constant and dielectric constants of the constituents.

5.1 Actual implementation of SCF methods for real systems

SCF methods — LDA scheme

For all the systems under study I calculate the electronic structure using the DFT–LDA self–consistent pseudopotential supercell method described in its general features in the previous Chapter. The exchange–correlation effects are treated using Ceperley and Alder[69] data and the analytic expression of Perdew and Zunger; the norm–conserving pseudopotentials tabulated by Bachelet, Hamman, and Schlüter[57] are used. I describe in the following some more technical details used for the new calculations presented here.

Periodically repeated supercells used in the present work

Periodically repeated supercells suitably chosen are used to study differently oriented interfaces. For instance the (001) oriented heterojunction is described by a tetragonal supercell, the longest side corresponding to the growth direction (z axis) of the heterojunction. In Fig. 5.1 are reported three isovolumic supercells used to describe (001), (110), and (111) oriented heterojunctions; they all contain 12 atoms, i.e. three double layers of each of the two materials. I have checked that, for most of the systems studied here and constituted by semiconductors with a bulk diamond or zincblende structure, a 3+3 supercell is thick enough in order to reproduce bulk features in the region midway two adjacent interfaces, also for the (110) oriented supercell which is —keeping fix the number of atoms— the shortest one in the growth direction.

To give a first idea of the size effects I show in Fig. 5.2(a) and (b) the differential plots of the electron density of the (001) 3+3 supercell describing GaAs/AlAs heterojunction with respect to the two correspondent bulks: it is evident that the regions midway between two adjacent interfaces are bulk regions; similar informations can be extracted by looking at the planar averages in Fig. 5.2(c). In order to check more quantitatively how much the size of the supercell affects the accuracy of the final result, I have performed and compared the calculations for (001) 2+2, 3+3 and 4+4 tetragonal supercells for the InP/$Ga_{0.47}In_{0.53}As$ heterojunction, as can be seen in Fig. 5.3: whereas the 2+2 supercell allows only a rough estimate of the electrostatic potential lineup, the result from the 3+3 case to the 4+4 case remains the same within 0.01 eV.

I finally just mention here that in some cases it can be necessary to improve the size of the supercell, mainly i) when the bulk structure of the materials constituting the heterojunction is more complicated, as in the case of InP/$(GaAs)_1(InAs)_1$ (001), where the bulk unit cell of the material on the right side is a tetragonal one, and ii) in

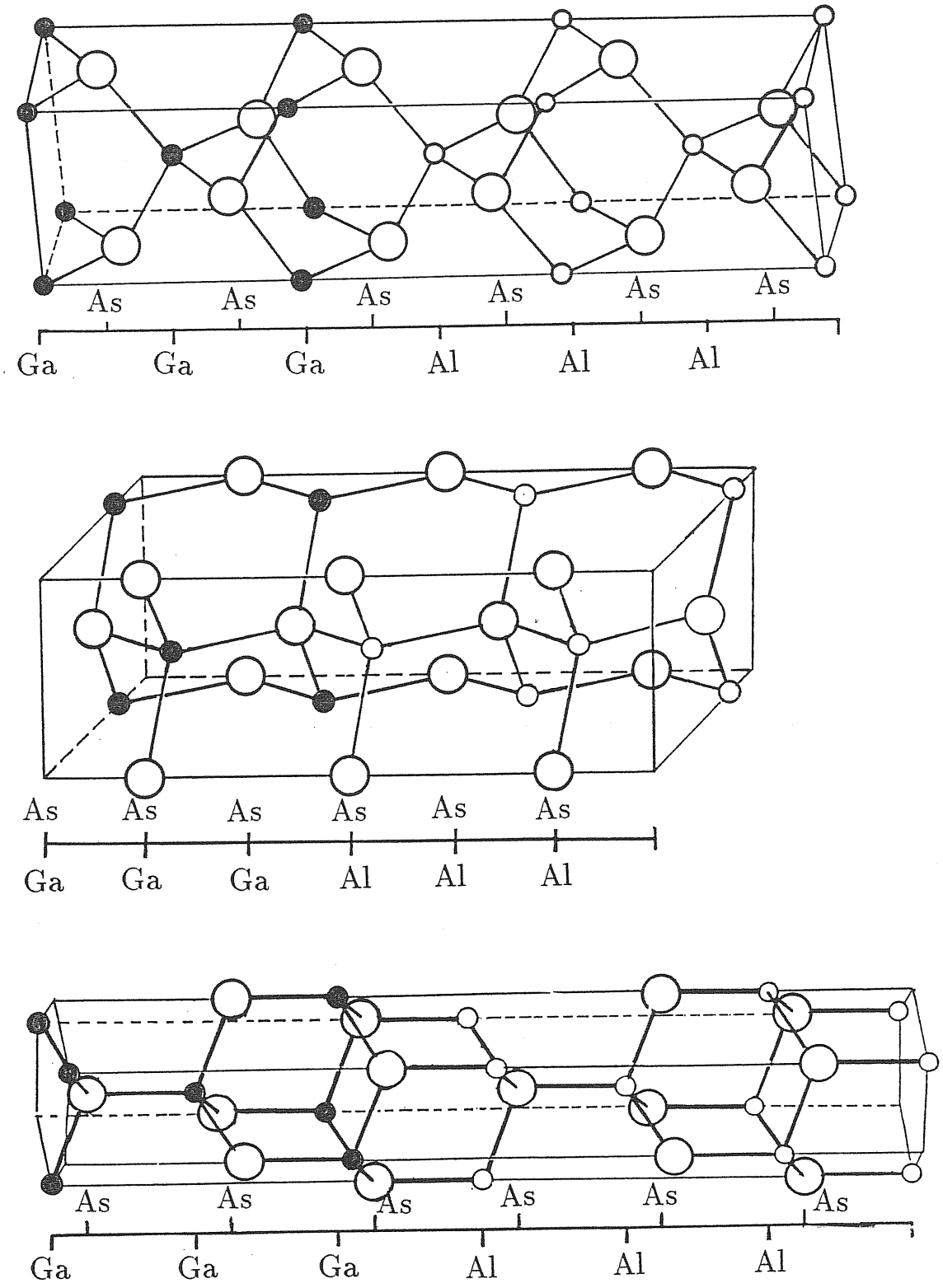

Figure 5.1. Supercell used for (001), (110), and (111) oriented heterojunctions, and the corresponding basis vectors; the case on GaAs/AlAs is considered to exemplify the distribution of the different ionic types in the anionic and cationic sublattice.

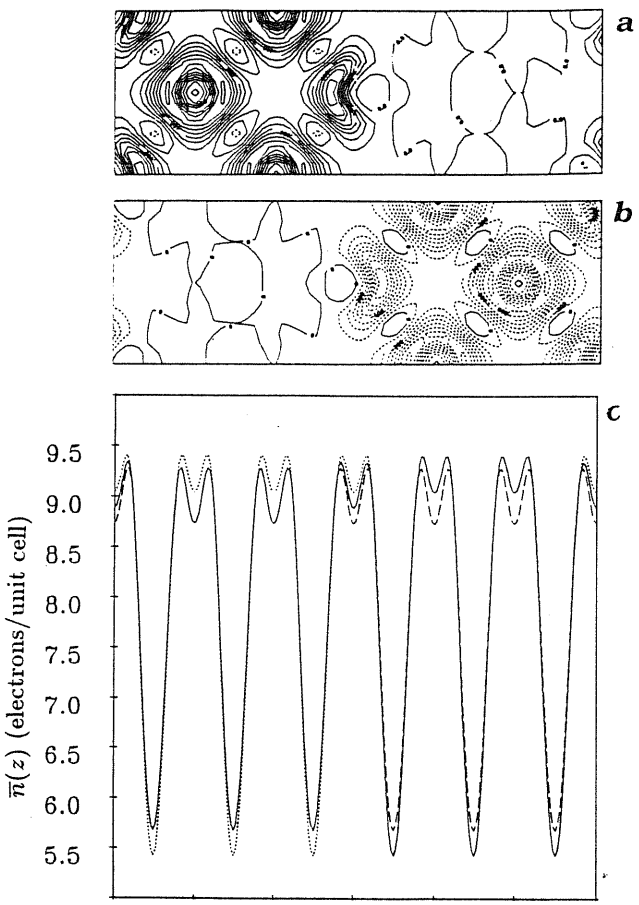

Figure 5.2. (a): differential plots obtained by subtracting to the electron density of the heterojunction the one of GaAs (a) and AlAs (b) in (001) 3+3 supercell; (c): planar averages of the electron density of GaAs/AlAs heterojunction (solid line) and of the two pure materials GaAs (dashed line) and AlAs (dotted line) in the same geometry.

case of heterovalent interfaces, such as Ge/GaAs, where the presence of the interface has longer–range effects.

Random alloy approximated with VCA

To conclude the discussion on the methods used to describe an heterojunction, I consider the case in which one of the materials constituting the system is an alloy, choosing as a prototype the case of InP/$Ga_{1-x}In_xAs$. The virtual crystal approximation (VCA) can be used to describe the alloy, by defining a virtual cation X=$\langle Ga_{1-x}In_x \rangle$, which is a properly weighted average of the two cationic species, and considering all the atoms occupying the unrelaxed positions of a zincblende.

I point out that the VCA approximation is better justified when the bulk lattice parameters of the pure materials constituting the alloy are equal, as in the case of $Ga_{1-x}Al_xAs$ but not of $Ga_{1-x}In_xAs$. The real situation in case of an alloy formed by lattice–mismatched components is nowadays well understood, thanks to EXAFS measurements which allow to investigate the atomic–scale structural properties, such as nearest–neighbor distances: internal lattice distortions are present in order to allow the individual bond lengths to remain close to the ones of the corresponding pure bulk materials, while the average lattice parameter (as deduced from usual X–ray diffraction data) is close to a linear interpolation between the two bulk values. We report in Fig. 5.4 the situation for $Ga_{1-x}In_xAs$.[70]

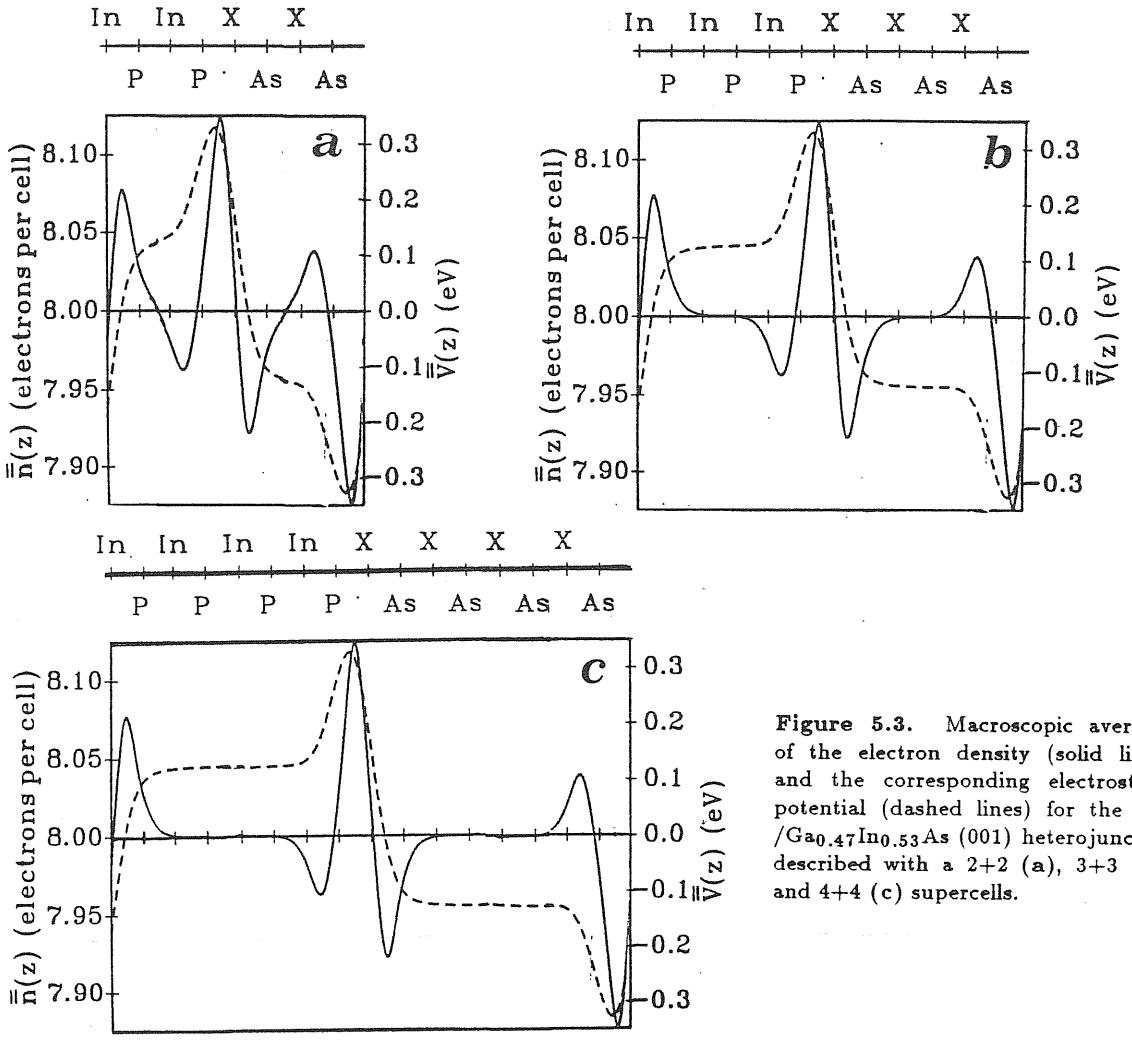

Figure 5.3. Macroscopic averages of the electron density (solid lines) and the corresponding electrostatic potential (dashed lines) for the InP /$Ga_{0.47}In_{0.53}As$ (001) heterojunction described with a 2+2 (a), 3+3 (b), and 4+4 (c) supercells.

Kinetic energy cutoff: convergence studies

Two more technical details will be discussed, i.e. PW and k special points, since they control the accuracy of the results. Let us first consider the problem of the number of plane waves, i.e. of the kinetic energy cutoff used. I take into account both the problem of supercell calculations, from which I extract the self–consistent charge density distribution in the heterojunction and consequently the electrostatic potential lineup, and the problem of the bulk calculations to obtain ΔE_v.

It is well known that the band–gap values are very sensitive to the size of the basis set used, depending on the material, its lattice parameter, its dielectric constant. It is also known that LDA predicts too small values for energy gaps, and moreover some semiconductors are metals in LDA at the theoretical equilibrium lattice parameter. Fortunately the valence bands, to which my attention is devoted, converge faster than the conduction bands when varying the kinetic energy cutoff; moreover, the use of the experimental lattice parameter usually avoids problems of closure of the gap and hence of "metallization" of the system. I studied the variation of the bulk valence band top edges and their difference with respect to the energy cutoff for GaAs/AlAs

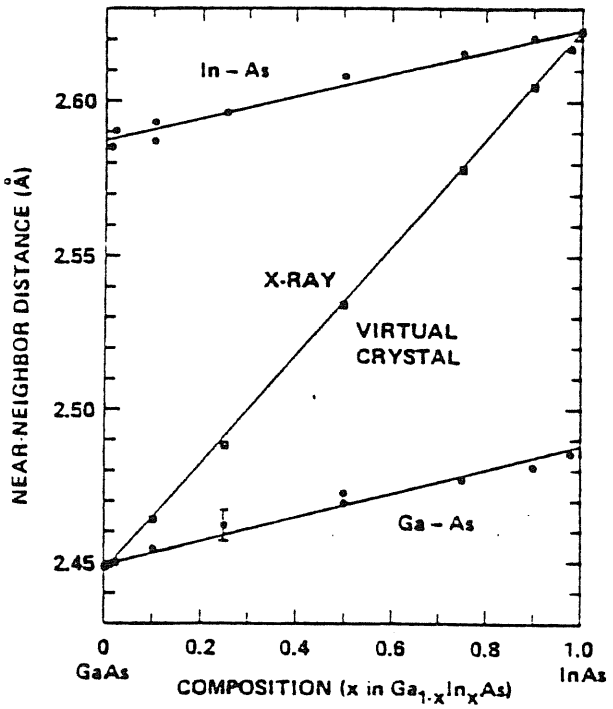

Figure 5.4. Experimental values of the nearest-neighbor distances and of the average lattice parameters for $Ga_{1-x}In_xAs$ alloy at different values of the composition x. The experimental lattice parameters of pure GaAs and InAs are respectively 10.5 a.u. and 11.5 a.u., and the experimental lattice parameter of the alloy is with an excellent approximation the linear interpolation between the two bulk values.

and for $InP/Ga_{0.47}In_{0.53}As$ systems, and I report in Tab. 5.1 the results. We see that in the case of GaAs/AlAs the valence top edges decrease as the cutoff increases, but the relative difference obtained at 14 Ryd differs by few meV from the fully converged result at 20 Ryd; for this system a kinetic energy cutoff of 14 Ryd —i.e. about 150 plane waves per atom— in all the calculations guarantees an accuracy of the order of 10 meV for the results.

Unfortunately this is not the case for $InP/Ga_{0.47}In_{0.53}As$, where the same accuracy would be obtained with a kinetic energy cutoff of at least 18 Ryd. Such a high cutoff is not a serious problem for the bulk calculations, where the unit cell used is small, but it is a problem for supercell calculations; hopefully, since the interface features that are obtained from a supercell calculation are mainly described by the long–wavelength Fourier components, that is by the small $\mathbf{q}$–vectors, one might expect that the electrostatic potential lineup is not so sensitive to improvements of the energy cutoff. As a first estimate of the accuracy obtained by using 12 Ryd$^{(*)}$, I compare the electron distribution in the two bulks InP and $Ga_{0.47}In_{0.53}As$ separately, calculated at 12, at 16 and 20 Ryd, and I report in Fig. 5.5 the contour plots of the relative differences between the two extreme cases in a plane containing the bonds.

$^{(*)}$ Note that the lattice parameter of $InP/Ga_{0.47}In_{0.53}As$ is 11 u.a. versus 10.5 of GaAs/AlAs, and 12 Ryd cutoff already correspond to about 140 plane waves per atom.

	$E_{cut}(Ryd)$	12	14	16	18	20	22
$AlAs$	E_v	5.162	5.125	5.118	5.127	5.117	5.107
	E_c	7.980	7.598	7.513	7.515	7.500	7.489
$GaAs$	E_v	5.185	5.173	5.160	5.170	5.160	5.150
	E_c	6.673	6.223	6.145	6.148	6.133	6.120
	ΔE_v	0.024	0.048	0.042	0.043	0.044	0.044
InP	E_v	3.769	3.798	3.813	3.818	3.817	3.818
	E_c	5.168	5.103	5.022	4.969	4.893	4.888
$Ga_{0.47}In_{0.53}As$	E_v	4.236	4.260	4.258	4.250	4.237	4.238
	E_c	4.807	4.777	4.704	4.675	4.642	4.642
	ΔE_v	0.467	0.461	0.446	0.432	0.421	0.421

Table 5.1. Variation with the kinetic energy cutoff of the valence band top edges of the bulk constituent and of the difference ΔE_v for the systems GaAs/AlAs and InP/$Ga_{0.47}In_{0.53}As$; the energy cutoff is in Ryd, the energy band eigenvalues in eV.

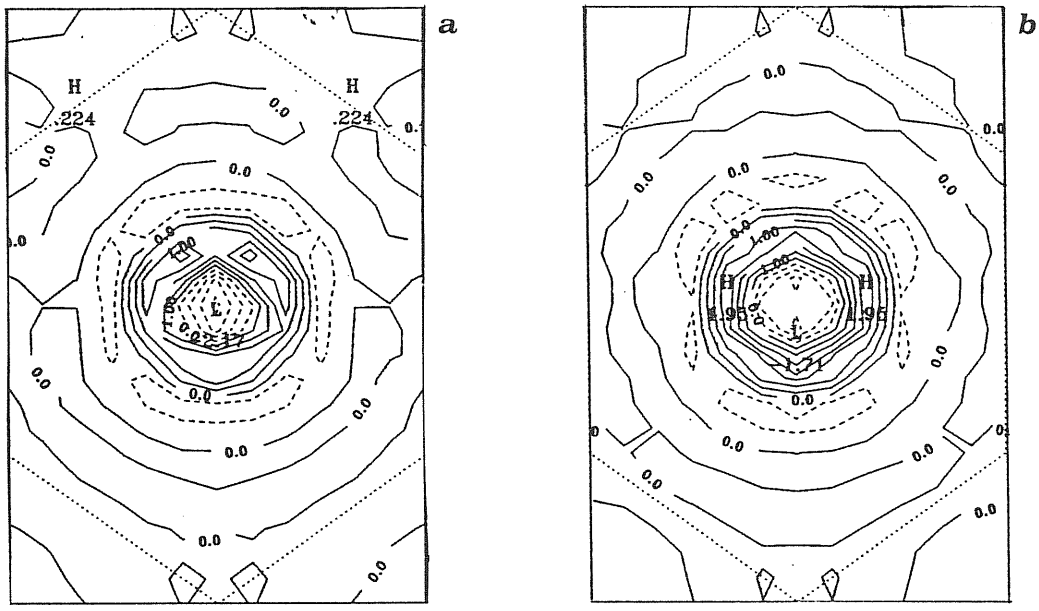

Figure 5.5. Differential plots of the electron density distribution calculated at 12 Ryd and at 20 Ryd for bulk InP (a) and $Ga_{0.47}In_{0.53}As$ (b); the plots are on planes containing the bond chains, and the anionic site is at the center of the figure.

We see that the differences are mainly concentrated in a small region around the anionic sites; since the electrostatic potential lineup is the interesting quantity, I estimate the correction to it from the second moment of these bulk charge distributions (†) and I find from 12 to 16 Ryd a difference of -0.03 eV for $Ga_{0.47}In_{0.53}As$ and of -0.01 eV for InP, on an absolute value of about 10 eV for each one, and from 16 to 20 Ryd a further correction, smaller than -0.01 eV, for $Ga_{0.47}In_{0.53}As$ only. Consequently, the electrostatic potential lineup should be corrected of a quantity equal to the difference of the corrections for the two bulks, i.e. of about $+0.03$ eV. The validity of such estimate has been checked by performing one supercell calculation at 16 Ryd: Fig. 5.6 shows the result compared with the one at 12 Ryd, and the relative difference in a magnified scale in the right panel. The difference of the electrostatic potential lineup obtained from these two calculations is just of $+0.02$ eV ($\Delta V = -0.25$ at 12 Ryd, and $\Delta V = -0.23$ at 16 Ryd), as predicted by the previous estimate.

Since ΔE_v from 12 to 16 Ryd has the same variation of ΔV, the total VBO is unchanged; I have thus decided to perform all the calculations for this system at 12 Ryd. being less than 0.01 eV the estimated uncertainty due to the not fully convergence with respect to the kinetic energy cutoff.

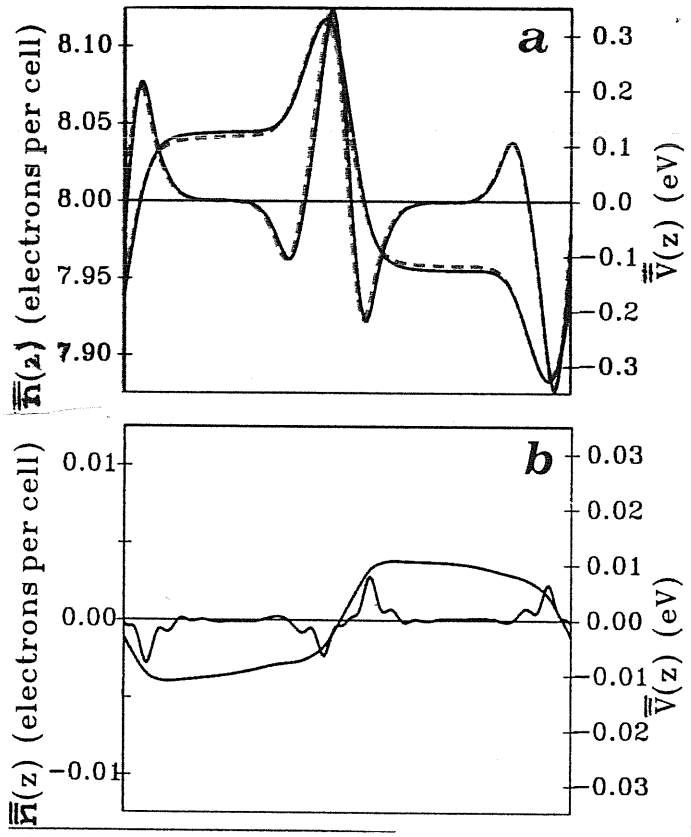

Figure 5.6. (a): Macroscopic averages of the electron density and electrostatic potential in $InP/Ga_{0.47}In_{0.53}As$ (001) 3+3 supercell calculated at 12 Ryd (solid line) and at 16 Ryd (dashed line); (b): macroscopic averages of the relative difference.

(†) I explain in details in Ch.5.1 the relationship between the average electrostatic potential and the second moment of the electron density.

Special points

As last but not less important technical detail I discuss the application of the special–point technique to perform k–space integration over the BZ.

As far as the bulk band structure calculations are concerned, in case of elemental or binary semiconductors —for instance— which have a zincblende structure I exclude the set of 2 Chadi–Cohen[60] points (see Tab. 5.2), and I usually use the set of 6 special points corresponding to the (333) Monkhorst–Pack[61] grid; results obtained with the 10 Chadi–Cohen special points are the same within 0.01 eV, so I conclude that in any case sets of special points somehow equivalent to the 6 ones for FCC BZ are enough to guarantee the desired accuracy of 10–20 meV on the final result.

basis vectors: $\frac{a}{2}(1,1,0)$, $\frac{a}{2}(1,0,1)$, $\frac{a}{2}(0,1,1)$

		x	y	z	weight
grid (2,2,2)	1	0.250000000	0.250000000	0.250000000	1/4
	2	0.750000000	0.250000000	0.250000000	3/4
grid (3,3,3)	1	0.166666667	-0.166666667	0.833333333	2/9
	2	0.166666667	0.166666667	0.166666667	2/27
	3	0.500000000	-0.166666667	-0.500000000	2/9
	4	0.500000000	0.166666667	-0.166666667	2/9
	5	0.500000000	0.500000000	0.500000000	1/27
	6	0.833333333	0.500000000	0.166666667	2/9
grid (4,4,4)	1	0.125000000	0.125000000	0.125000000	1/32
	2	0.375000000	0.125000000	0.125000000	3/32
	3	0.375000000	0.375000000	0.125000000	3/32
	4	0.375000000	0.375000000	0.375000000	1/32
	5	0.625000000	0.125000000	0.125000000	3/32
	6	0.625000000	0.375000000	0.125000000	3/16
	7	0.625000000	0.375000000	0.375000000	3/32
	8	0.625000000	0.625000000	0.125000000	3/32
	9	0.875000000	0.125000000	0.125000000	3/32
	10	0.875000000	0.375000000	0.125000000	3/16

Table 5.2. Set of 2, 6, and 10 special k points used for the Brillouin zone of the FCC structure, corresponding respectively to the (222), (333), and (444) Monkhorst–Pack grids.

The case of supercells is not a textbook one, and I spend here few words about it. One possible choice is to work independently for each supercell: to generate different sets of special points, to do convergence tests and to select a proper one; an alternative way is trying to select sets of special points commensurate for the various structures (see Tab. 5.3): if it is possible, the physical quantities obtained are directly comparable(*).

(*) This is fundamental in case one wants for instance differential charge density plots: the use of non commensurate k point grids can produce spurious differences which have no physical meaning and mask the real effects.

	FCC_2	$(001)_{m+m}$				$(110)_{m+m}$				$(111)_{m+m}$		
		1	2	3	4	1	2	3	4	1	2	3
(222) *grid*	*y*	*y*	*y*	*n*	*n*	*y*	*y*	*n*	*n*	*y*	*n*	*n*
(333) *grid*	*n*	*y*	*n*	*y*	*n*	*y*	*n*	*y*	*n*	*n*	*n*	*n*
(444) *grid*	*y*	*y*	*y*	*n*	*y*	*y*	*y*	*n*	*y*	*y*	*y*	*n*

Table 5.3. Equivalence of special point sets for the FCC unit cell with those for the other structures used in the present work.

			x	y	z	weight
FCC	grid (2,2,2)	1	0.500000000	0.500000000	0.500000000	1
SC	grid (3,3,3)	1	0.166666667	0.166666667	0.166666667	8/27
		2	0.500000000	0.166666667	0.166666667	4/9
		3	0.500000000	0.500000000	0.166666667	2/9
		4	0.500000000	0.500000000	0.500000000	1/27
(001) 3+3	grid (3,3,3)	1	0.000000000	0.166666667	0.117851130	2/9
		2	0.333333333	0.166666667	0.117851130	2/9
		3	0.166666667	0.333333333	0.117851130	2/9
		4	0.000000000	0.500000000	0.117851130	1/9
		5	0.333333333	0.500000000	0.117851130	2/9
(001) 4+4	grid (4,4,4)	1	0.000000000	0.125000000	0.088388348	1/8
		2	0.250000000	0.125000000	0.088388348	1/8
		3	0.125000000	0.250000000	0.088388348	1/8
		4	0.000000000	0.375000000	0.088388348	1/8
		5	0.375000000	0.250000000	0.088388348	1/8
		6	0.250000000	0.375000000	0.088388348	1/8
		7	0.125000000	0.500000000	0.088388348	1/8
		8	0.375000000	0.500000000	0.088388348	1/8
(110) 3+3	grid (3,3,3)	1	0.166666667	0.000000000	0.117851130	2/9
		2	0.166666667	0.000000000	0.353553391	1/9
		3	0.000000000	0.166666667	0.117851130	1/9
		4	0.000000000	0.166666667	0.353553391	1/18
		5	0.500000000	0.000000000	0.117851130	1/9
		6	0.500000000	0.000000000	0.353553391	1/18
		7	0.333333333	0.166666667	0.117851130	2/9
		8	0.333333333	0.166666667	0.353553391	1/9
	N(Janack)=2	1	0.125000000	0.041700000	0.088400000	1/8
		2	0.125000000	0.041700000	0.265200000	1/8
		3	0.125000000	0.125000000	0.088400000	1/8
		4	0.125000000	0.125000000	0.265200000	1/8
		5	0.375000000	0.041700000	0.088400000	1/8
		6	0.375000000	0.041700000	0.265200000	1/8
		7	0.375000000	0.125000000	0.088400000	1/8
		8	0.375000000	0.125000000	0.265200000	1/8
(110) 4+4	grid (4,4,4)	1	0.125000000	0.000000000	0.088388348	1/8
		2	0.125000000	0.000000000	0.265165043	1/8
		3	0.000000000	0.125000000	0.088388348	1/16
		4	0.000000000	0.125000000	0.265165043	1/16
		5	0.375000000	0.000000000	0.088388348	1/8
		6	0.375000000	0.000000000	0.265165043	1/8
		7	0.250000000	0.125000000	0.088388348	1/8
		8	0.250000000	0.125000000	0.265165043	1/8
		9	0.500000000	0.125000000	0.088388348	1/16
		10	0.500000000	0.125000000	0.265165043	1/16
(111) 3+3	N(Janack)=2	1	0.088388348	0.144300000	0.025500000	1/8
		2	0.088388348	0.144300000	0.076500000	1/8
		3	0.250000000	0.433000000	0.025500000	1/8
		4	0.250000000	0.433000000	0.076500000	1/8
		5	0.333333333	0.288700000	0.025500000	1/8
		6	0.333333333	0.288700000	0.076500000	1/8
		7	0.088388348	0.433000000	0.025500000	1/8
		8	0.088388348	0.433000000	0.076500000	1/8

Table 5.4. Special point sets (coordinates and relative weights) used in the present work.

Let us consider, for instance, the (001) 3+3 supercell: the 6 special points for the FCC BZ can be exactly folded in this geometry, being the (333) Monkhorst–Pack grid exactly commensurate in the two structures. Also for (110) 1+1 (or (001) 1+1 which is equivalent) and for (110) 3+3 supercells a set of special points equivalent to the 6 ones for FCC BZ does exist, whereas this is not the case for any (111) supercell. The (001) 4+4 supercell has a special point set equivalent to the 10 Chadi–Cohen points for FCC BZ, i.e. the two structures are commensurate with a (444) Monkorst–Pack grid.

I have also used alternative methods of generating special points for some test calculations of in (110) supercells and in all cases for (111) supercells. In particular the use of an $n = 2$ grid with Janack's routines provides results of accuracy comparable to that obtained with a (333) Monkorst–Pack grid. In Tab. 5.4 are reported all the different special point sets used in the present work, including those generated with Janack's routines.

5.2 Results

5.2.1 Isovalent common–anion heterojunctions: GaAs/AlAs

Previous first-principles calculations performed for this system [37,68] indicate that (001), (110), and (111) interfaces have similar offsets; experiments indicate at most a weak dependence on orientation and growth sequence [21]. The present experimental accuracy however is not high: the most reliable data for the valence-band offset in GaAs/AlAs range from -0.40 to -0.55 eV (see Tab. 2.2).

SCF 3+3 supercell calculations have been performed for (111), (110) and (001) interface orientations, [71] using the experimental equilibrium lattice parameter 10.5 a.u. at standard conditions of temperature and pressure. Fig. 5.7 shows both the planar and macroscopic averages of the electron density and the electrostatic potential. An inspection of the figures shows that the supercell is thick enough to satisfactorily reproduce bulk features midway the two interfaces and that computational noise is small even on the magnified scale of the macroscopic averages. Note that $\bar{\bar{n}}(z)$ has a typical dipolar shape around n_o across the interface. The average electrostatic potential shift produced by this dipolar distribution is *independent on the interface orientation*; in fact, despite large differences of the three–dimensional electron density $n(\mathbf{r})$ and of its planar average $\bar{n}(z)$ for the three interfaces, similar shapes are obtained for $\bar{\bar{n}}(z)$ and $\bar{\bar{V}}(z)$ and the identical value ΔV=-0.41 eV for the lineup.

When referring the band structure of each material to the average of its own electrostatic potential, the top of the valence band in GaAs is 0.04 eV higher than in AlAs. Adding to this quantity experimental spin-orbit data (0.03 eV) to the potential lineups calculated above, one obtains for all the three interfaces the same VBO=-0.48 eV, in substantial agreement with the previous first–principles calculations for (001) and (110) interfaces and well within the present experimental error bar (see Fig. 5.8).

The fact that the macroscopic dipole is the same for the interfaces oriented along the three main crystallographic directions cannot be *a priori* extended to any interface orientation; however it supports the idea that for GaAs/AlAs the lineup is basically a bulk effect, confirming the validity of the WSC model when applied to such system. Although up to now the independence on the interface orientation is not completely justified, the present first–principles calculations shed some light on the long debate about the importance of the interface details for the electrostatic potential lineup: the important thing are not the microscopic interface details, but their *macroscopic effects*; and the macroscopic effects may be the same even if the electronic distribution at the interface is strongly dependent on the interface details, as for instance on the interface orientation.

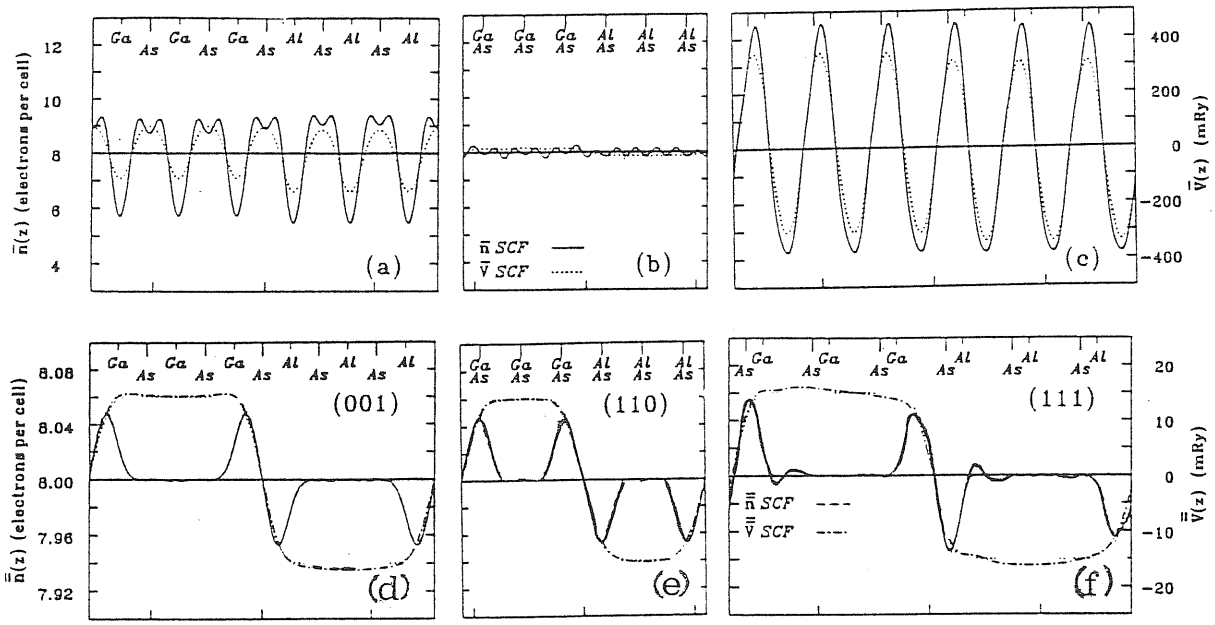

Figure 5.7. Planar averages of the electron density (solid line) and potential (dashed line) of $(GaAs)_3(AlAs)_3$ superlattices. a: (001); b: (110); c: (111); d,e,f: corresponding macroscopic averages.

5.2.2 Isovalent no–common–anion heterojunctions: $InP/Ga_{1-x}In_xAs$

The ternary alloy $Ga_xIn_{1-x}As$ grown on InP substrates forms at the composition x=0.47 a high–quality lattice–matched heterostructure —with the equilibrium lattice parameter of 11 a.u.— which is presently the subject of intensive study given its applications in optoelectronic devices[5,72]. Band offsets at this interface have been measured using different techniques and contradictory results have been obtained (see Tab. 2.2 and quoted references).

Contrary to the large amount of experimental data, no first–principles theoretical investigation has been done for this system, not even at $x = 0.47$; the only theoretical prediction of band offsets is due to Van de Walle[73] who found VBO=0.35 eV for this composition within his model–solid scheme.[37]

I use here the virtual-crystal approximation (VCA) for $Ga_{0.47}In_{0.53}As$[74] considering the anions and the virtual cations $X=\langle Ga_{0.47}In_{0.53}\rangle$ at their ideal lattice sites. I study first the (001) interface. I show in the left panels of Fig. 5.9 the planar

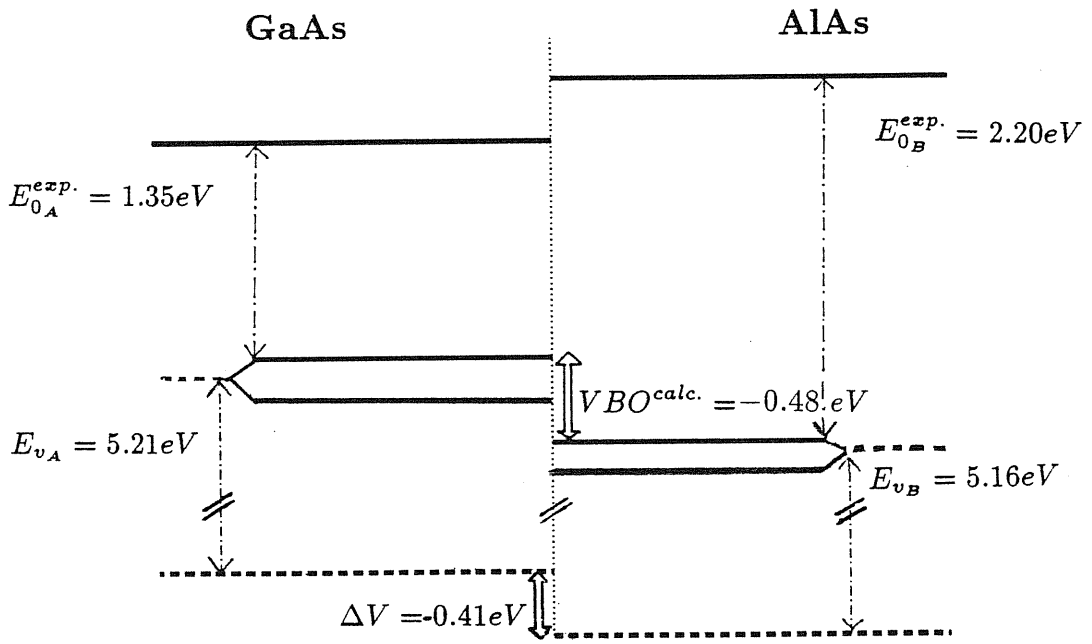

Figure 5.8. Schematic band structure at GaAs/AlAs interfaces.

and macroscopic[38] averages of the SCF supercell electron density and electrostatic potential; the corresponding value of the potential lineup is $\Delta V = -0.25$ eV. Adding to ΔV the quantity ΔE_v=+0.56 eV obtained from the bulk calculations corrected with experimental spin-orbit data (+0.09 eV)[75], one obtains VBO=+0.31 eV. I have also performed SCF calculations for 3+3 supercells oriented along (110) and (111). In the mid and right panels of Fig. 5.9 I report the corresponding electron densities and electrostatic potentials. The different amplitude of the oscillations in the planar averages $\overline{n}(z)$ and $\overline{V}(z)$ reveals the different geometry of the three interfaces. The simple typical dipolar shape already found in other materials[71] can be recognized in the macroscopic average $\overline{\overline{n}}(z)$ for the (110) oriented heterojunction, whereas the charge distribution at (001) and (111) heterojunctions gives rise to more complicated interface shapes. I stress that, despite the strong orientation dependence of the density profiles, their integrals give the same electrostatic potential lineup within our numerical accuracy, and therefore the same valence band offset, for all three orientations (see Fig. 5.10).

The charge rearrangement at a general isovalent heterojunction with no common ions comes from both anionic and cationic contributions which interfere differently for different orientations. In order to isolate the anionic and cationic contributions, I study the two ideal heterojunctions CP/CAs and InA/XA, where 'C' is a virtual cation whose pseudopotential is the average between the cationic pseudopotentials of the two bulks, i.e. $C=\langle In_{0.5}X_{0.5}\rangle = \langle In_{0.715}Ga_{0.285}\rangle$, and analogously for the virtual anion $A=\langle P_{0.5}As_{0.5}\rangle$.

The results of SCF (001) supercell calculations for these two heterojunctions are

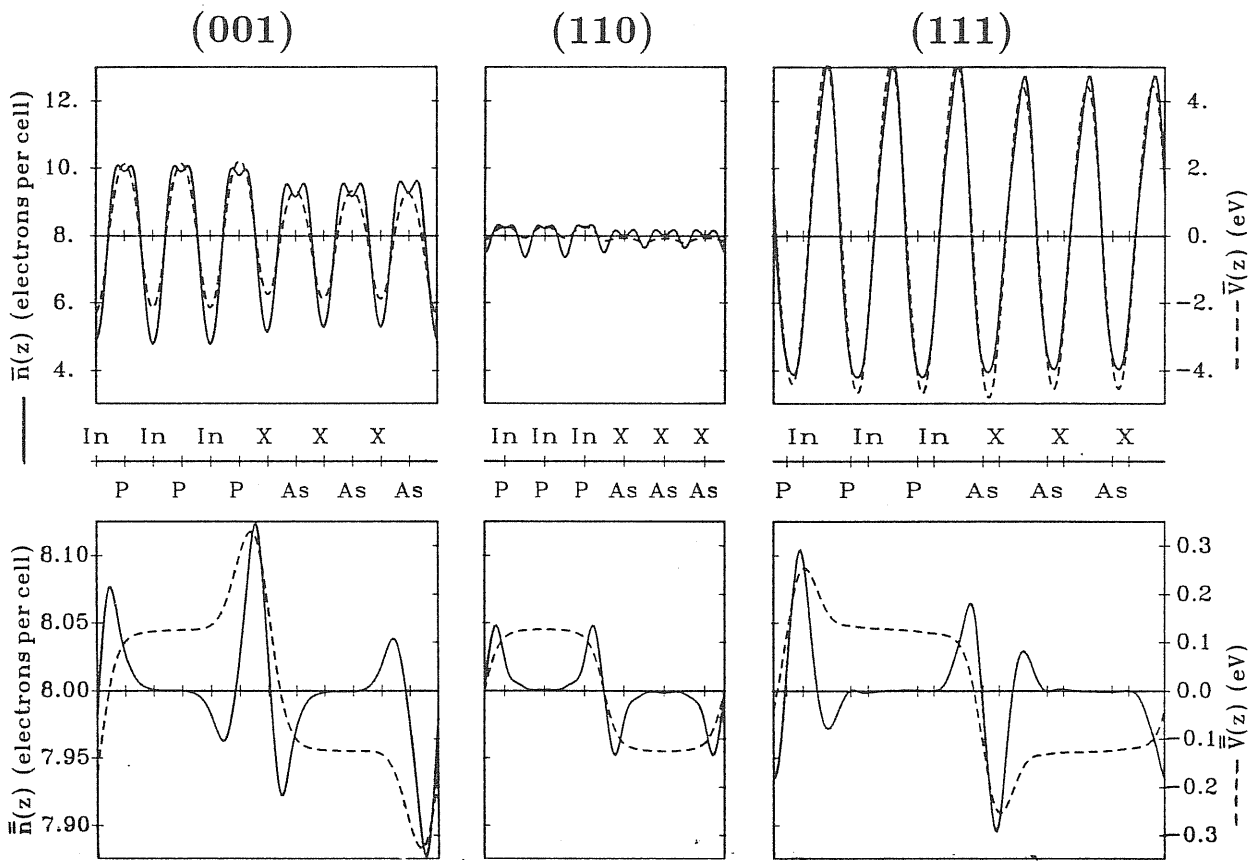

Figure 5.9. Planar averages of the electron density (solid line) and potential (dashed line) of $InP/Ga_{0.47}In_{0.53}As$ heterojunctions. a: (001); b: (110); c: (111); d,e,f: corresponding macroscopic averages.

reported in Fig. 5.11. The profiles of the anionic and cationic interface dipoles in CP/CAs and InA/XA heterojunctions have the typical simple dipolar shape, and the contributions to the electrostatic potential lineup are respectively $\Delta V_A = -0.61$ and $\Delta V_C = +0.36$ eV. Calculations performed for the common-anion and common-cation (110) and (111) heterojunctions show similar macroscopic averages as in (001) direction, and a complete orientation independence for the respective potential lineup, within our numerical accuracy.

Not only is the total lineup the sum of the anionic and cationic contributions, but even the fine details of the total charge-density profile are closely reproduced by superimposing the density profiles of CP/CAs and InA/XAs interfaces centered at the appropriate atomic planes (see Fig. 5.12). This explains the different shapes of the macroscopically averaged charge densities obtained for different orientations: the total density profile is the sum of dipoles opposite in sign and shifted one respect to the other for (001) and (111) oriented heterojunctions, where anions and cations stay on different planes perpendicular to the growth axis; opposite dipoles but centered on the same atomic planes characterize instead the (110) orientation, where all the planes contain both ionic species.

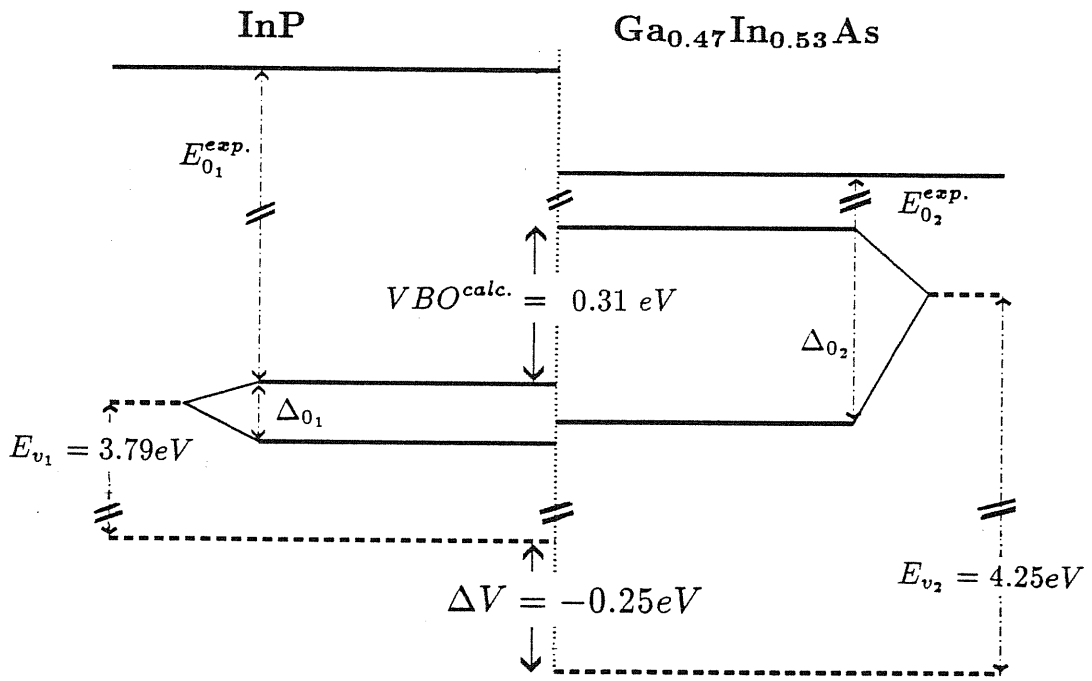

Figure 5.10. Schematic band structure at $InP/Ga_{0.47}In_{0.53}As$ interfaces.

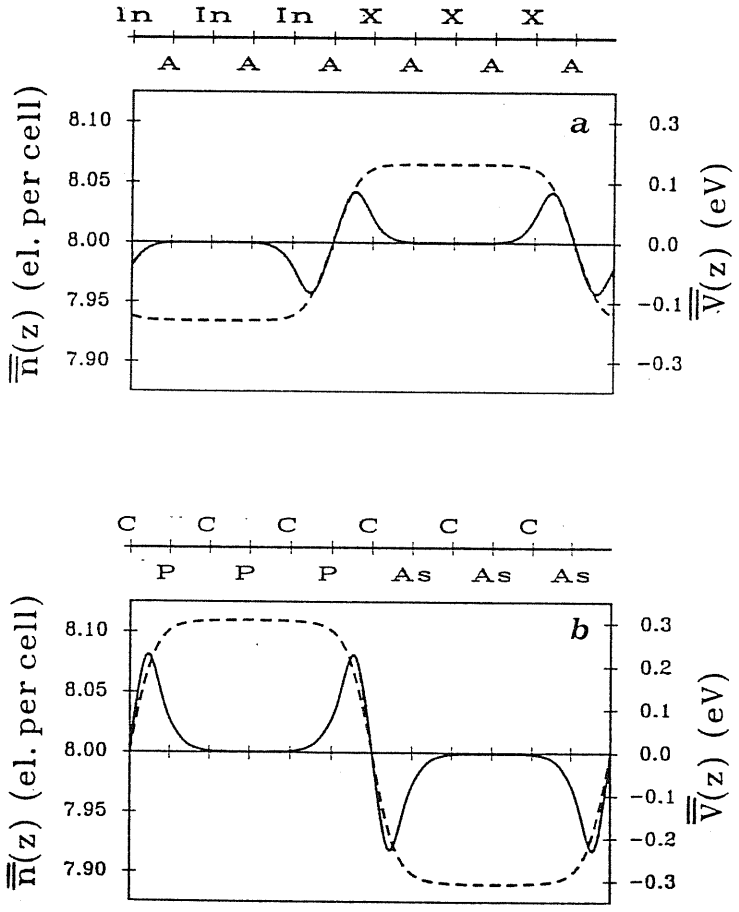

Figure 5.11. Macroscopic average of the supercell electron density (solid line) and electrostatic potential (dashed line) of (001) oriented InA/XA (**a**) and CP/CAs (**b**) ideal heterojunctions.

One may note that the contributions of anions and cations to the total potential lineup are competing —the anionic one being larger and opposite in sign—, and this explains the relative small ΔV in the present system where both cations and anions differ on the two sides of the junction. A first hint of this competition comes just from looking at the electronegativity scale where it appears that P is more electronegative with respect to As and Ga is with respect to In, so that one expects the electrons in the interface regions flowing from P to As and from the cations in the alloy to In.

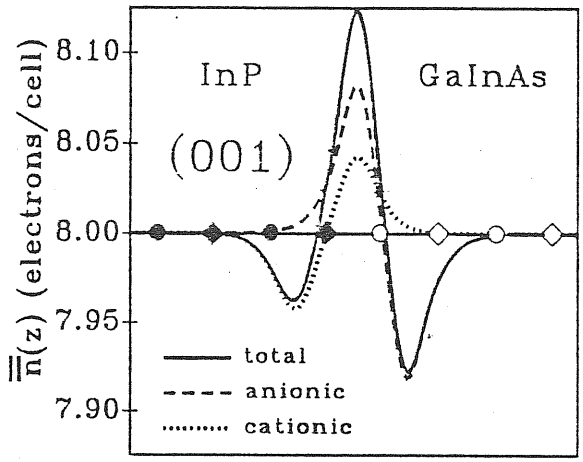

Figure 5.12. Macroscopic interface dipole in $InP/Ga_{0.47}In_{0.53}As$ (001), and its decomposition into anionic and cationic contributions; circles (squares) indicate the position of the cation (anion) planes.

5.2.3 Heterovalent heterojunctions: Ge/GaAs

I consider now the interface between two heterovalent semiconductors. I will discuss in details the case of Ge/GaAs interfaces,[76] oriented along (001) and (110) directions. In the (001) direction, there exist two non–equivalent interfaces, according to whether the GaAs region is Ga– or As–terminated: inspection of the corresponding macroscopic averages, as indicated in Fig. 5.13, show that these interfaces are *charged* and therefore thermodynamically unstable. In fact, a net interface charge would give rise to macroscopic electric fields which would alter the *bulk* energy of the system. Martin[77] has shown that the energy is lower when mixed Ga-Ge and/or As-Ge layers are present, and in particular the lowest values correspond to completely "compensated" mixture of atoms in the interface, when no free carriers are present.

The simplest stoichiometric configurations in (001) direction which give rise to a neutral interface are those where only *one* layer is mixed, and in particular it contains 50% Ga and 50% Ge atoms or 50% As and 50% Ge atoms; in any case, the equal number of Ga-Ge and As-Ge bonds in the interface region implies absence of free carriers. In the (110) direction atomic planes with two ions (one cation and one anion) per unit surface cell are present, and no polar interface occurs.

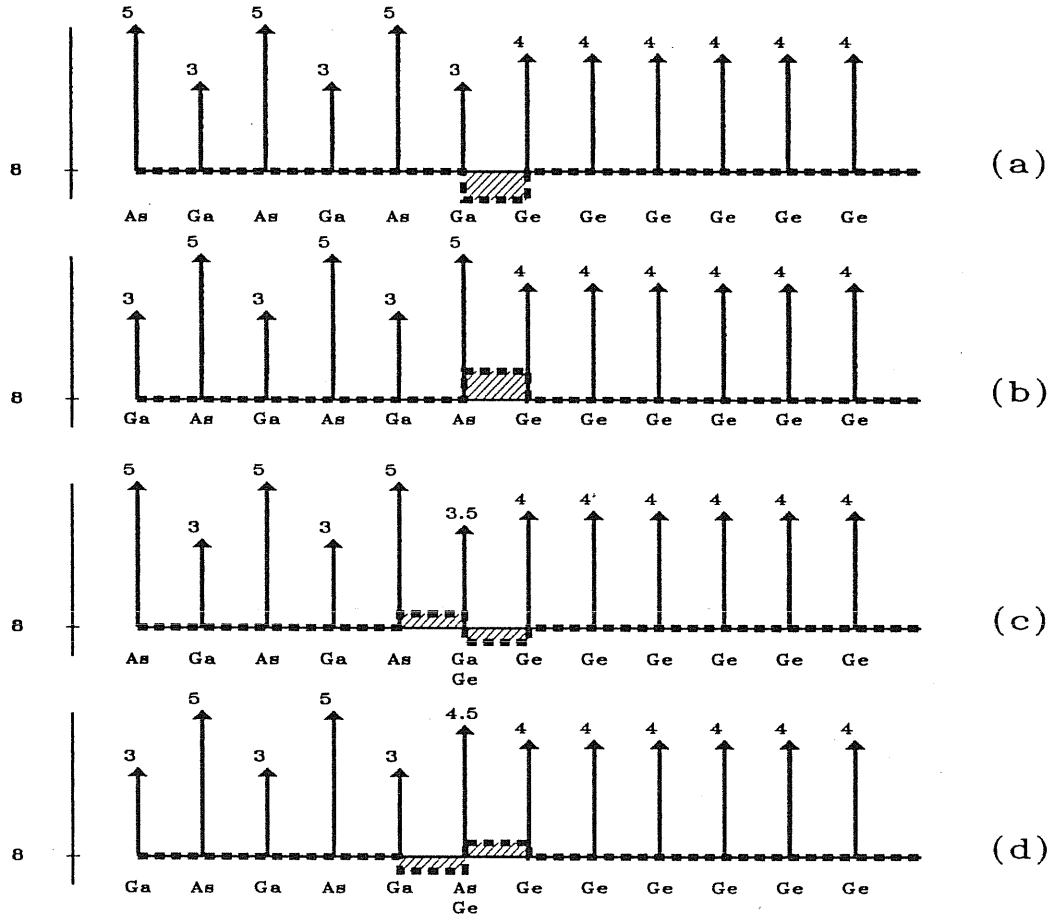

Figure 5.13. Distribution of planar charges for the succession of the atomic planes of the Ge/GaAs heterojunction in (001) direction in case whether the GaAs region is terminated with atomic planes containing only Ga (a), or As (b), or mixed planes of Ga–Ge (c) or As–Ge (d). Arrows indicate the delta functions corresponding to the nuclear charges; dashed line the macroscopic average of such point–like charge distribution with the corresponding scale (electrons per cell) displayed on the left y axis.

In Fig. 5.14 the macroscopic average of the electron charge density and of the electrostatic potential of (001) and (110) interfaces are reported. One may notice that the absolute value of the potential lineup is much larger in the (001) cases that in the (110) one. In fact, only the *electronic* contribution is represented in the figure. To this one must add the contribution of the bare point–charge (sketched in Fig.

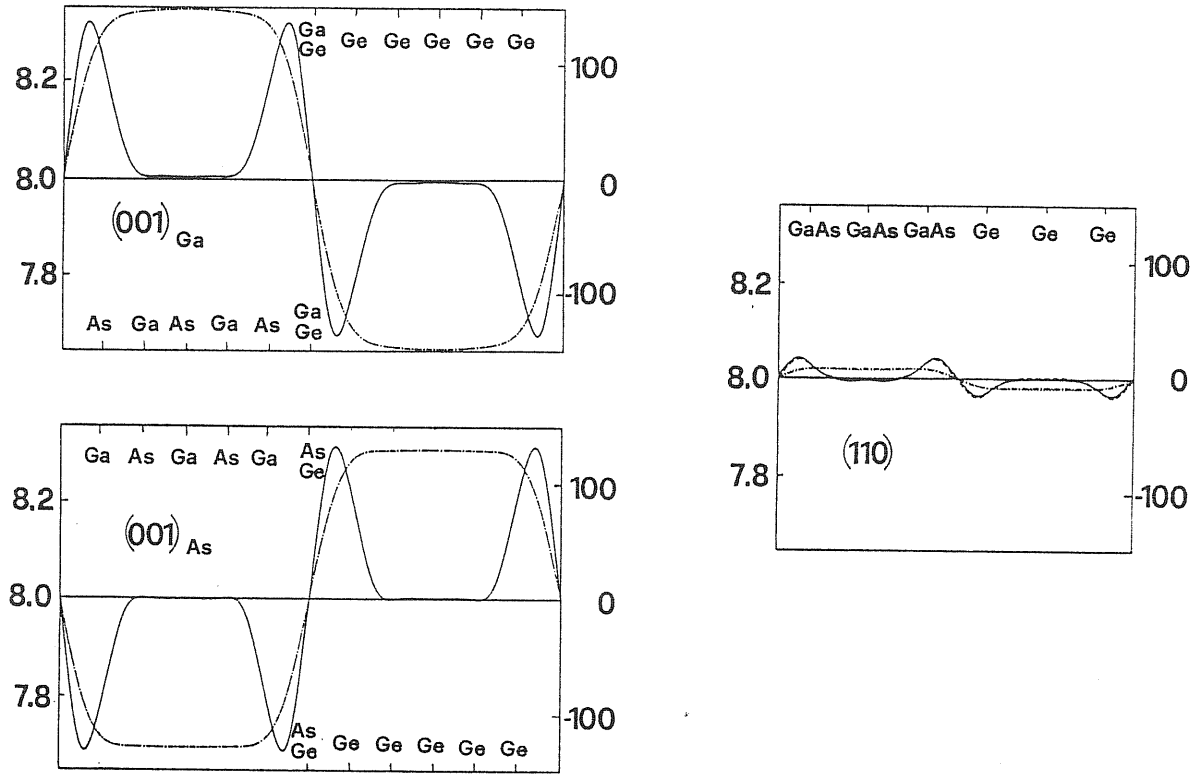

Figure 5.14. Macroscopic averages of the SCF charge densities (solid lines) and Hartree potentials (dotted–dashed lines) of Ga-Ge–terminated (001), As-Ge–terminated (001), and (110) Ge/GaAs $3+3$ superlattices; left scale (charge) is electrons per zincblende cell; right scale (potential) is mRyd. Dashed lines, barely distinguishable from the others, refer to LRT calculations.

5.13) $\Delta V_{ion} = \pm \frac{\pi e^2}{2a_0} = \pm 4.08$ eV, as it can be easily found by simple considerations of classical macroscopic electrostatics (see Eq. 3.4), which is opposite in sign and comparable in magnitude.

The total lineup is -0.22, $+0.06$, and -0.54 eV for the (110), $(001)_{Ga}$, and $(001)_{As}$ interfaces. Note that, as already empirically observed by Kunc and Martin a few years ago[10], the (110) lineup coincides with the average value of the two non–equivalent (001) interfaces. To obtain the band offsets, one has to add to the potential lineups the difference in the bulk band edges, $\Delta E_v = +0.76$ eV: this gives a VBO of $+0.54$, $+0.82$, and $+0.22$ eV respectively (see Fig. 5.15). The corresponding experimental figures are $+0.56$, $+0.55$, and $+0.60$ eV[26].

The agreement between theory and experiment is very good for the (110) case and poor for the (001) geometries. This may be a direct consequence of atomic intermixing occurring at the interface: we have in fact seen how sensitive is the lineup to the actual atomic arrangement at polar interfaces: it is a simple exercise

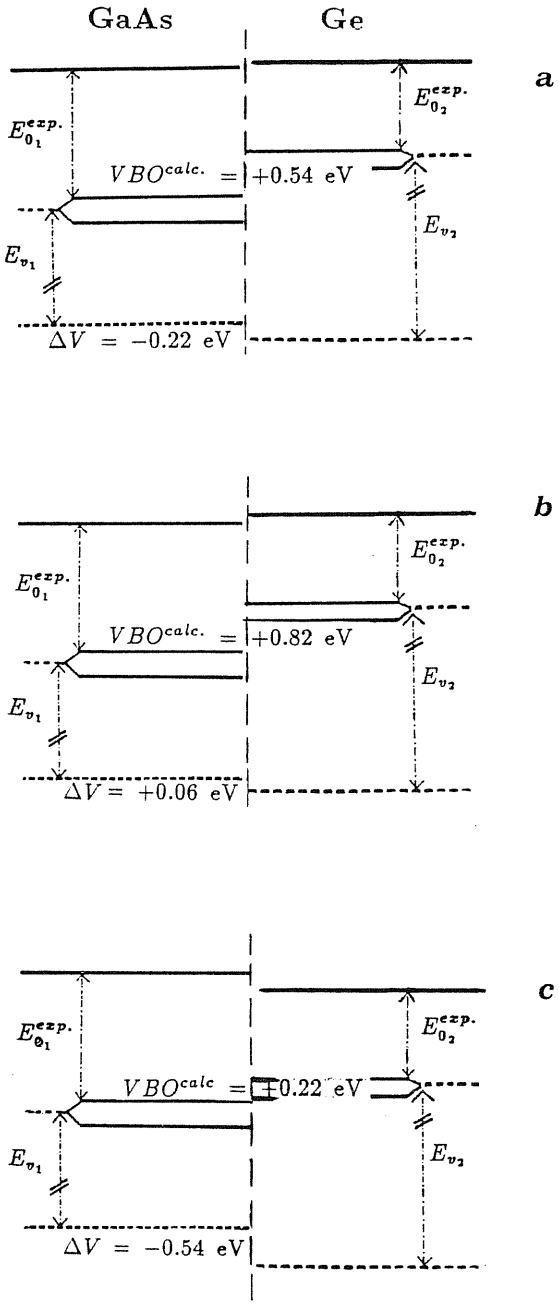

Figure 5.15. Schematic band structure at Ge/GaAs interfaces in case of (110) non polar interface (a), (001) Ge-As-terminated (b), (001) Ge-Ga-terminated.

to find different geometries having very different lineups. It is still an open question to determine the mechanisms responsible for the interface atomic arrangement in epitaxial structures; whatever such mechanisms may be, our results show that—for polar interfaces—the band offset is a very sensitive quantity to characterize the quality of the interface.

5.3 Analysis of first–principles results and open questions

Inaccuracies of the present SCF calculations

The main sources of inaccuracies of the calculations presented here are i) the neglect of many–body effects in the bulk–band contributions to the offset, ΔE_v, ii) the use of the VCA for describing the gallium–indium arsenide alloy in InP/$Ga_{0.47}In_{0.53}As$ heterojunctions, and iii) the neglect of possible lattice distortions at the interface and in the alloy region. The first question is common to all the systems studied here, whereas the lattice distortions can be actually neglected in GaAs/AlAs and Ge/GaAs, where all the bond lengths are the same even at the interface. The calculations performed for the InP/$Ga_{0.47}In_{0.53}As$ heterojunction are affected by the greatest number of approximations: I will discuss in detail for this system also the corrections to the VCA.

Many–body effects

Recent studies [78,79,80] have shown that the many–body corrections to the LDA valence–band maximum are negative and greater in more ionic materials. Few results are available, namely for GaAs, AlAs, Si, diamond, and Ge; but even among these, significant discrepancies are found by comparing the results obtained by different authors for the same material [78,79]. I should stress that the calculated individual absolute values of the self–energy corrections are not very accurate; one cannot therefore employ, for instance, results obtained by two different groups for the two semiconductors constituting an heterojunction. the important point is that fortunately differences required for the VBO problem can be considerably more accurate than the individual absolute magnitudes, provided that the two bulk calculations are comparably converged.

The only system for which informations are available in this sense is GaAs/AlAs. In particular, S.B. Zhang and cooworkers [78] have shown that the correct treatment of many–body effects increases the VBO in GaAs/AlAs by $+0.12\pm0.02$ eV, thus bringing to -0.60 eV the VBO here obtained; a similar value (+0.09 eV) is given by Godby and cooworkers [80].

It is important to estimate the many–body corrections to the VBO for InP/$Ga_{0.47}In_{0.53}As$; no calculations are available for InP, nor for GaInAs. Thanks to the similarities between GaAs/AlAs and InP/$Ga_{0.47}In_{0.53}As$ heterojunctions(*),

(*) The valence–band maximum in GaAs is higher than in AlAs, and in $Ga_{0.47}In_{0.53}As$ higher than in InP; GaAs and $Ga_{0.47}In_{0.53}As$ are also the more ionic semiconductors in the respective heterojunctions.

by scaling the corrections of Ref. 78 with the difference of the reciprocal dielectric constants of the bulk materials[81] one can estimate a correction of $+0.08 \pm 0.02$ eV[(†)] to the VBO for the system under study.

As for Ge/GaAs, studies for the two semiconductors separately are available (see Ref. 79 for Ge), but the numerics are not comparable, and no estimate can be extracted for the correction to the calculated VBO. However we have seen that for heterovalent heterojunctions the interface details can influence so strongly the VBO, that the LDA is not the most serious approximation actually made.

Beyond VCA: effects of structural order

InP/$(InAs)_1(GaAs)_1$ (001) (undistorted monolayer superlattice)

In order to discuss the validity of the VCA to describe the gallium–indium arsenide region in InP/$Ga_{0.47}In_{0.53}As$, let us calculate the VBO between InP and an ordered phase of the GaInAs alloy. I consider here the situation of extreme order for the cation sublattice, i.e. the $(InAs)_1(GaAs)_1$ monolayer superlattice. Progress in epitaxial growth techniques has made possible now to grow monolayer superlattices on (001) InP substrates[82], opening up possibilities for new devices[83]; but also in standard conditions some partial ordering is observed[84].

I have performed SCF calculations using 4+4 (001) tetragonal supercells (see Fig. 5.16) and treating the gallium–indium arsenide region as $(InAs)_1(GaAs)_1$ monolayer unrelaxed superlattice, i.e. ordering the cationic sublattice but taking all the bond lengths and angles frozen and equal to the VCA case; $Ga_{0.5}In_{0.5}As$ random alloy phase simulated using VCA, as already described, is also studied for comparison.[(*)]

The results of the SCF supercell calculations are shown in Fig. 5.17 (left and mid panels), and a summary is given in Tab. 5.5. The total electrostatic potential lineup is -0.25 eV and -0.19 eV for the random alloy and for the SL case respectively, i.e. the use of "real" atoms instead of the VCA increases the calculated potential lineup by $+0.06$ eV. Of course, corrections beyond the VCA not only affect the potential lineup ΔV, but also the bulk-band contribution to the offset, ΔE_v. A precise estimate of the latter would require the study of disorder-induced self-energy corrections in the bulk alloy, and is beyond the present work. A rough estimate of these self-energy corrections is obtained comparing the top of the valence band in the virtual crystal

(†) The dielectric constants used are 10.9, 8.16, 12, and 9.61 for GaAs, AlAs, $Ga_{0.47}In_{0.53}As$ and InP respectively.

(*) One reasonably expects that this case practically coincides with the (001) 3+3 supercell calculations already performed for the perfectly lattice–matched InP/XAs interface, because of the very small difference in the alloy composition; however, new calculations are necessary in order to compare the results in systems having exactly the same stoichiometry.

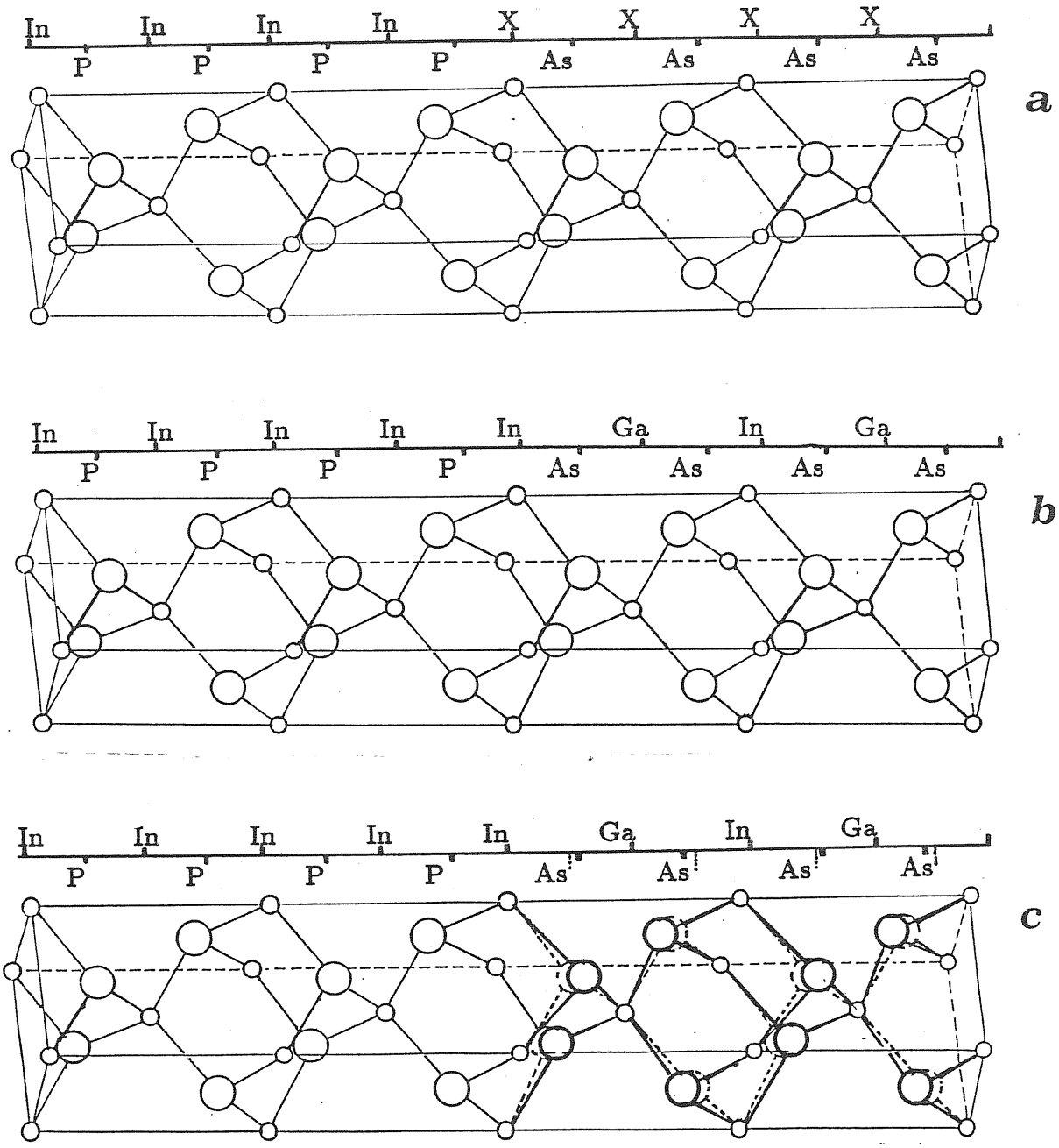

Figure 5.16. Tetragonal 4+4 supercells used to describe InP/GaInAs (001) heterojunctions when the GaInAs region is constituted by an alloy treated in VCA (a), $(InAs)_1(GaAs)_1$ monolayer unrelaxed superlattice (b), $(InAs)_1(GaAs)_1$ monolayer relaxed superlattice (c).

XAs and in the $(GaAs)_1(InAs)_1$ monolayer superlattice. The resulting correction amount to ≈ -0.04 eV. This partially cancels the correction to the potential lineup, giving a total positive correction of $\approx +0.02$ eV to the VBO, which therefore goes from $+0.31$ eV to $+0.33$ eV.

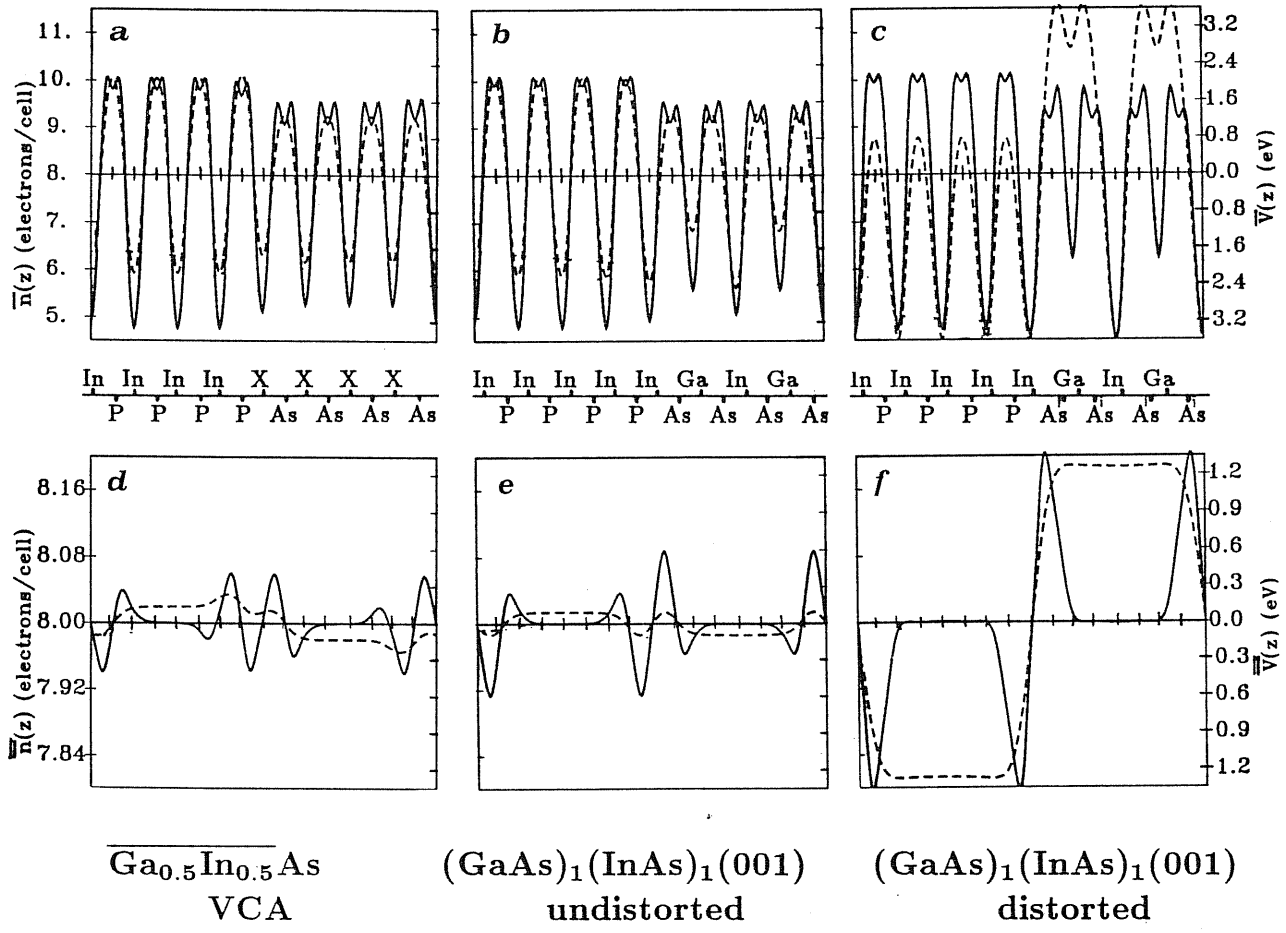

Figure 5.17. Planar (top panels) and macroscopic (down panels) averages of the electron densities (solid lines) and the electrostatic potentials (dashed lines) for (001) InP/GaInAs heterojunctions when the GaInAs region is constituted by an alloy treated in VCA (a,d), $(InAs)_1(GaAs)_1$ monolayer unrelaxed superlattice (b,e), $(InAs)_1(GaAs)_1$ monolayer relaxed superlattice (c,f).

Effects of lattice relaxation:
InP/$(InAs)_1(GaAs)_1$ (001) (distorted monolayer superlattice)

A further correction to VCA predictions comes from internal distortions present in the GaInAs alloy in order to allow the individual bonds to maintain the length they have at equilibrium in the bulk materials[70]. To linear order in the displacements, these effects cancel on the average in the bulk random alloy, but their balance is not zero in case of ordered alloy and even more significant in case of the $(InAs)_1(GaAs)_1$ (001) monolayer SL. In any case, lattice distortions could give some nonnegligible effects at the interface, due to incomplete cancellation.

Let me first discuss here the effects due to the internal lattice distortions; according to previous studies[85,86] regarding the stability of $(InAs)_1(GaAs)_1$ monolayer superlattice, in the equilibrium configuration the As–planes are rigidly shifted by a certain amount ξa_0 from In–planes towards Ga–planes, according to the fact that the equilibrium lattice parameter in GaAs is smaller than in InAs

($\approx$10.5 a.u. versus $\approx$11.5 a.u.), whereas cations remain at the ideal fcc sites (see Fig. 5.16 (c)). The structural parameters are taken from the cited existing works: more precisely the average lattice parameter is again 11 a.u., as considered for the previous calculations; the equilibrium structural distortion parameter is $\xi = 0.017$ (8% of the unrelaxed interplanar distance), whereas the tetragonal distortion of the superlattice is here completely neglected being less than 1%. I note that this is not the perfect equilibrium configuration for the 4+4 supercell here used: a rough estimate, according to the forces acting on the atoms, indicates that the As–planes should be actually displaced by a smaller quantity ($\xi \approx 0.012$), but this will not change dramatically the present analysis.

The SCF supercell results for the $InP/(InAs)_1(GaAs)_1$ (001) relaxed superlattice are reported in the right panels of Fig. 5.17 (electronic contribution only): the total electrostatic potential lineup results from two competing contributions, one purely electronic ($\Delta V_{electr.} = +2.54$ eV), and the other due to the dipole caused by the displacement of bare As ions which is $\Delta V_{ion} = -\frac{4\pi e^2 Z}{a_0}\xi = -2.64$ eV, exactly calculated in the framework of the classical electrostatics (see Eq. 3.4 and in Fig. 5.18(a) and (b)).

The change of the total electrostatic potential lineup due to the internal lattice relaxation is thus +0.09 eV (ΔV from -0.19 eV to -0.10 eV). The magnitude of the total effect (bare ions displacements screened by electronic rearrangement) can be easily predicted using the effective charges: the total effect is that obtained displacing of the same amount an effective charge $\frac{Z^*}{\varepsilon_\infty}$. This is briefly justified in the following. I compare the derivative of the internal energy of a system with respect to the applied electric field as described i) in the framework of the macroscopic electrostatics:

$$-\frac{\partial W}{\partial \mathbf{E}} = \frac{\Omega}{4\pi}\mathbf{D} = \frac{\Omega}{4\pi}(\mathbf{E} + 4\pi\mathbf{P}) \tag{5.1}$$

where the polarization $\mathbf{P}$ is defined as $\mathbf{P} = \frac{1}{\Omega}\sum_s^{rel.} \mathbf{d}_s$, and ii) according to a microscopic picture:

$$-\frac{\partial W}{\partial \mathbf{E}} = \frac{\Omega}{4\pi}\varepsilon_\infty \mathbf{E} + \sum_s^{rel.} eZ_s^* \mathbf{u}_s \tag{5.2}$$

where s labels the atomic sites, $\mathbf{u}_s$ the atomic displacements from equilibrium, and Z_s^* is defined as the proportionality constant through which the electric field acts on the atoms (nucleus plus electronic cloud), i.e. by the relationship $\mathbf{F}_s = eZ_s^*\mathbf{E}$. By comparing now Eqs. (5.1) and (5.2) one obtains

$$\frac{\Omega}{4\pi}(\mathbf{E} + 4\pi\mathbf{P}) = \frac{\Omega}{4\pi}\varepsilon_\infty \mathbf{E} + \sum_s^{rel.} eZ_s^* \mathbf{u}_s \tag{5.3}$$

and imposing **D**=0 (no external charges, which is the case for of periodically repeated supercells) we obtain:

$$\mathbf{P} = \frac{1}{\varepsilon_\infty}\frac{1}{\Omega}\sum_s^{rel.} eZ_s^*\mathbf{u}_s \tag{5.4}$$

which eventually indicates that the total dipole (nuclear and electronic) associated to each atomic displacement is nothing but $\mathbf{d}_s = \frac{Z_s^*}{\varepsilon_\infty}\mathbf{u}_s$. In this particular case $Z_{As}^* \approx -2.36$ [87] and $\varepsilon_\infty \approx 12$ [81], so that the variation in the electrostatic potential lineup due to the lattice internal distortion is $\Delta V_{distortion} = -\frac{4\pi e^2 Z^*}{a_0 \varepsilon_\infty}\xi \approx +0.10$ eV, reproducing the value obtained from the SCF supercell calculations ($\approx +0.09$ eV). Also the bulk term ΔE_v changes, passing from +0.52 eV for the unrelaxed superlattice — including the spin–orbit splitting [75]— to +0.56 eV in case of relaxation. Note that the corrections to δV and ΔE_v are of the same sign in this case, and do not cancel each other; I thus conclude that the VBO is slightly affected by the order in GaInAs region passing from +0.31 eV to +0.33 eV, and a larger effect due to the internal distortion of $(InAs)_1(GaAs)_1$ superlattice further shifts the VBO to +0.46 eV.

	$\Delta V_{electr.}$	ΔV_{ion}	$\Delta V_{tot.}$	$\Delta E_v^{non\ rel.}$	$VBO^{non\ rel.}$	$\Delta\Delta_0$	$\Delta E_v^{rel.}$	$VBO^{rel.}$
$\langle Ga_{0.5}In_{0.5}\rangle As$ *VCA*	−0.25	0	−0.25	+0.47	+0.22	+0.09	+0.56	+0.31
$(GaAs)_1(InAs)_1(001)$ *undistorted*	−0.19	0	−0.19	+0.43	+0.24	+0.09	+0.52	+0.33
effects of alloy internal order	+0.06	0	+0.06	−0.04	+0.02		−0.04	+0.02
$(GaAs)_1(InAs)_1(001)$ *distorted*	+2.54	−2.64	−0.10	+0.47	+0.37	+0.09	+0.56	+0.46
effects of superlattice internal distortion			+0.09	+0.04	+0.13		+0.04	+0.13

Table 5.5. Contributions to the VBO in InP/GaInAs heterojunctions in the case of a random $Ga_{0.5}In_{0.5}As$ alloy treated in VCA, of $(GaAs)_1(InAs)_1$ (001) unrelaxed superlattice, and of $(GaAs)_1(InAs)_1$ (001) relaxed superlattice. The last row gives for comparison the effects of the internal lattice distortion. The columns from the left to the right show the electrostatic potential lineup due to the electrons, that due to the bare ions, and the total one; the non relativistic ΔE_v and VBO, the spin–orbit contribution [75], and finally the relativistic ΔE_v and VBO values. Units are eV.

Structural distortions will also occur at the interface. For instance there is some very recent experimental evidence of intrinsic strain at $InP/Ga_{0.47}In_{0.53}As$ (001) interfaces [88], in particular a negative strain at the ...InPInP/XAsXAs... interface

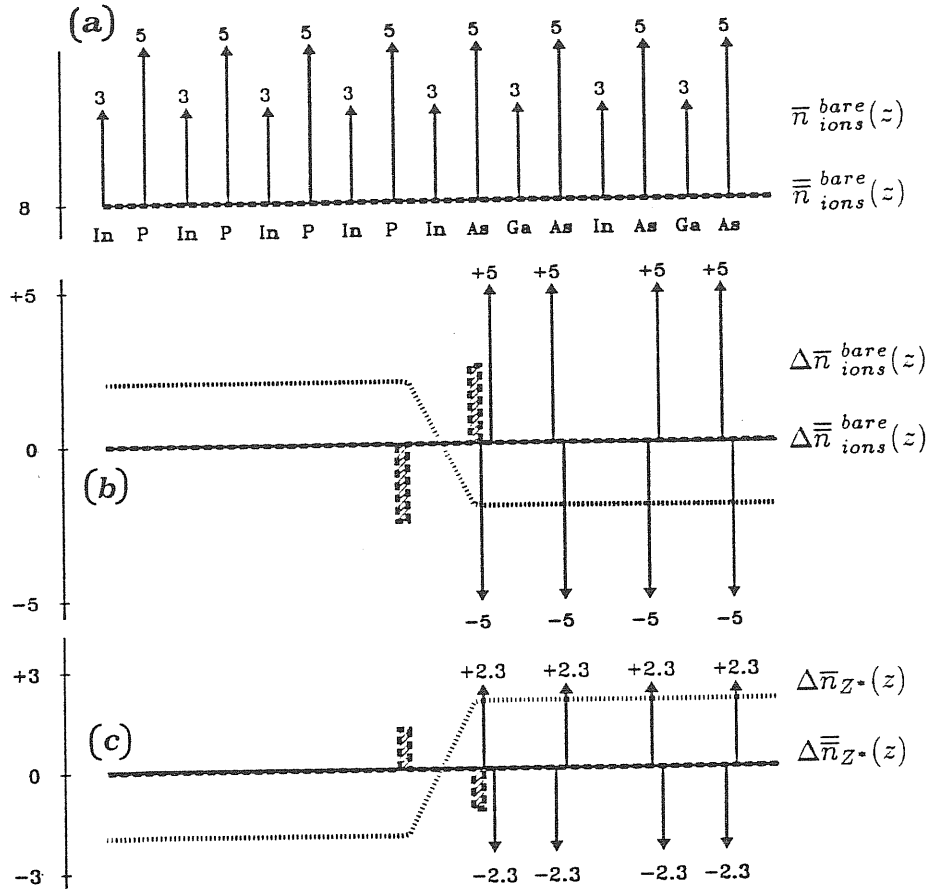

Figure 5.18. Distribution of planar charges for the succession of the atomic planes of the InP/$(GaAs)_1(InAs)_1$ (001) heterojunction in case of undistorted superlattice (a); the bare ionic perturbation (shift of As planes) leading to the $(GaAs)_1(InAs)_1$ (001) distorted superlattice is represented in (b), and analogously the total perturbation described in terms of effective charges is in (c). Arrows indicate the delta functions corresponding to the nuclear charges in (a) and (b) and to the atomic effective charges in (c); dashed line the macroscopic average of such point-like charge distribution with the corresponding scale (electrons per cell) displayed on the left y axis; dotted line the corresponding electrostatic potential.

and a positive complementary strain at the ...XAsXAs/InPInP... interface. This configuration is easily reproduced by a simple model in which the distances between the two adjacent atomic planes —which is $\frac{a_0}{4}$ for the undistorted configuration— are calculated from a weighted average of the relevant bulk bond lengths: whereas the distance between adjacent planes in the two bulk regions is $\frac{11}{4}$ a.u., the one between In and As planes is $\frac{11.5}{4}$ a.u. and the one between P and X planes is $\frac{10.5}{4}$ a.u.. An estimate of the effects of such interface strain to the electrostatic potential lineup and hence to the VBO can be obtained again trough the effective charges (see Fig. 5.19): $\Delta V_{strain} = \frac{4\pi e^2 Z^*}{a_0 \varepsilon_\infty}\xi \approx +0.07$ eV.(*) The interface strain is likely overestimated, and

(*) In Fig. 5.19 the substrate InP is considered fixed, and the XAs region is shifted, but one could also think of moving the InP region towards the XAs region. To take into account both the situations, I consider the average of the effective charges and of the dielectric constant of XAs and InP: $Z^* \approx 2.46$ and $\frac{1}{\varepsilon_\infty} = \frac{1}{2}\left[\frac{1}{\varepsilon_\infty(InP)} + \frac{1}{\varepsilon_\infty(XAs)}\right] \approx \frac{1}{10.67}$ [87].

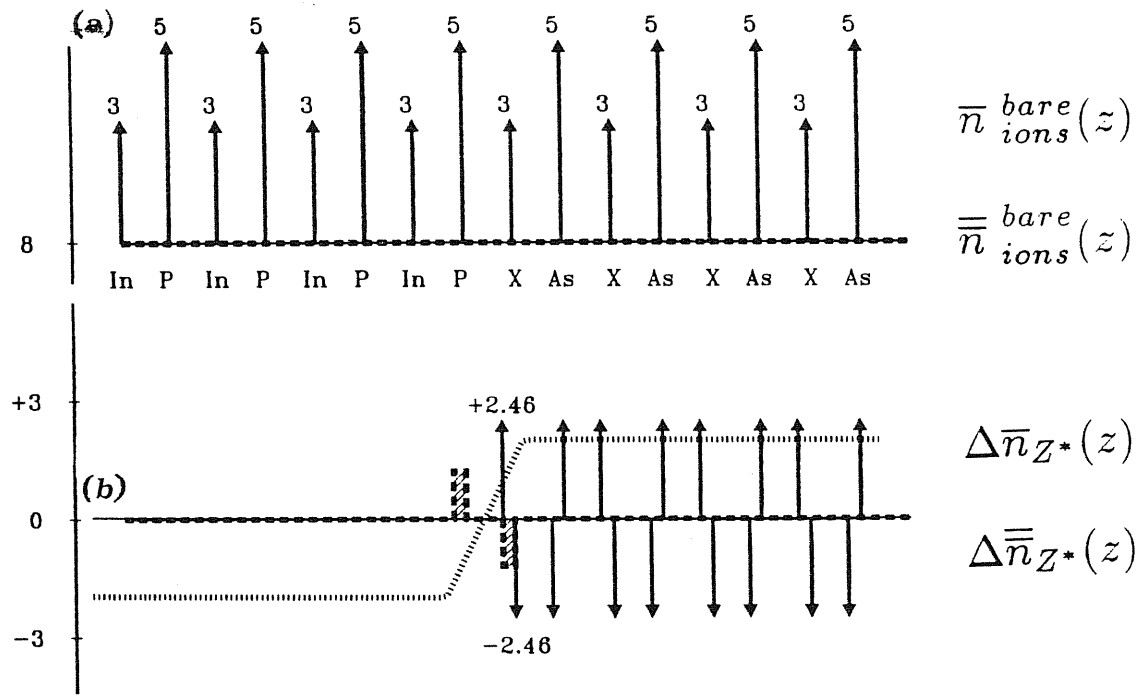

Figure 5.19. Distribution of planar charges for the succession of the atomic planes of the InP/XAs (001) heterojunction in case of absence of interface strain (a); the perturbation (shift of X and As planes) leading to the interface strain is represented in (b) in terms of effective charges. Arrows indicate the delta functions corresponding to the nuclear charges in (a) and to the atomic effective charges in (b); dashed line the macroscopic average of such point-like charge distribution with the corresponding scale (electrons per cell) displayed on the left y axis; dotted line indicates the corresponding electrostatic potential.

consequently ΔV_{strain} may be smaller than the value here obtained.

Conclusions

In summary, the supercell SCF calculations here performed for some different prototype systems have clearly show that:

i) the equal value of the VBO found for the three main crystallographic orientations is a general characteristic of the isovalent interfaces, both common-anion (as GaAs/AlAs) and no-common-ion (e.g. InP/GaInAs);

ii) in the general isovalent heterojunction with no common ions the anionic and cationic effects which determine the charge arrangement at the interface are additive: not only the total VBO can be exactly decomposed into anionic and cationic contributions, but even the fine details of the charge density can be explained in terms of linear superposition of anionic and cationic dipolar distributions;

iii) in case of heterovalent interfaces, the electrostatic potential lineup (and hence the VBO) is strongly dependent on the details of the interface (atomic arrangement

and even orientation), but the interface–dependent contribution can be isolated and easily calculated, for a given structural configuration, in terms of bulk macroscopic quantities; moreover, the (110) lineup coincides with the average value of the two non–equivalent (001) interfaces;

iv) the effects on the VBO due to atomic order in case of presence of alloys seem to be small ($\approx$0.01–0.02 eV): the estimate here extracted from the paradigmatic case of extreme order, i.e. monolayer superlattice, may be further reduced for real alloy;

v) more serious effects on the VBO ($\approx$ 0.10 eV) are produced by lattice internal distortions, possible both at the interface and in the bulk regions in case of compound materials; once the atomic displacements are known (but this question is not trivial, and often beyond our capabilities), their effects on the potential lineup and hence on the VBO can be roughly estimated by using the atomic effective charges.

Despite these interesting conclusions, the question whether the VBO is really a bulk quantity for the large class of isovalent interfaces is not yet completely ascertained and still requires some further investigation: in particular I should prove whether the VBO is the same in *any interface orientation* —and not only in (001), (110), and (111)—, and whether it is affected or not by structural disorder at the interface (i.e. deviation from abruptness). Also the effects of long–range structural order in the bulk regions (e.g. in case of partially ordered alloy) have to be more deeply investigated.

In general, a justification of these properties has to be found: why for the isovalent heterojunctions the VBO is equal for different orientations? Why anionic and cationic effects are exactly additive? Why ordering effects in the bulk regions are so small? Why for heterovalent heterojunctions the lineup at the (110) non polar interface is the average between the two non equivalent (001) interfaces?

The last part of the work will try to clarify these crucial questions.

Chapter 6

LINEAR RESPONSE APPROACH

The smallness of the chemical differences between different semiconductors suggests a new approach to the interface problem, based on low-order perturbation theory. The heterojunction is described as a perturbation with respect to a reference periodic system (virtual crystal); the bare potential describing such a perturbation is the sum of localized potentials which transform the ions of the virtual crystal into the true ions of the two semiconductors.

The low-order electronic response to a general perturbation is studied, starting from the response to an isolated perturbation, and the most suitable choice of the reference unperturbed system is discussed. The analysis shows that linear-response theory (LRT), if properly applied, gives results in excellent agreement with the SCF results. Moreover, LRT is a powerful tool to explain the general qualitative trends emerging from SCF calculations for the different classes of heterojunctions, and for extending their scope.

6.1 Formulation of the basic theory and methods

General considerations: infinite solid and electrostatic potential

As we have discussed in Ch. 3, the problem of determining the energy-band discontinuities is related to the difficulty of defining the average electrostatic potential in an infinite solid. From a more formal point of view, we observe that the average potential in any *finite* system can be defined as:

$$\langle V \rangle = 4\pi e^2 \lim_{\mathbf{q}\to 0} \frac{\widetilde{n}_{cryst}(\mathbf{q})}{q^2}. \tag{6.1}$$

In a periodic solid, Eq. (6.1) is meaningless because the lattice structure factor makes the Fourier coefficients of the crystal density to be defined only at *discrete* wavevectors. However, any *prescription* to interpolate between the discrete physical values of $\widetilde{n}_{cryst}(\mathbf{q})$ with a continuous function allows to calculate the limit, provided it exists. A recipe to decompose the crystal density into a sum of localized contributions is equivalent to a prescription to interpolate between its Fourier coefficients, and allows to give a precise meaning to Eq. (6.1). If we think of infinite systems as made up of *rigid* atomic-like charge distributions:

$$n_{cryst}(\mathbf{r}) = \sum_{\mathbf{R}} n_{loc}(\mathbf{r} - \mathbf{R}), \tag{6.2}$$

Eq. (6.1) can be rewritten as:

$$\langle V \rangle = \frac{4\pi e^2}{\Omega} \lim_{\mathbf{q}\to 0} \frac{\widetilde{n}_{loc}(\mathbf{q})}{q^2} \tag{6.3}$$

where Ω is the volume of the unit cell. The limit exists —and therefore the average potential inside the crystal is well defined— provided the long-wavelength behavior of the localized charge distribution is $\widetilde{n}_{loc}(\mathbf{q}) \sim \alpha q^2 + \mathcal{O}(q^3)$. The absence of constant, linear, and anisotropic quadratic terms in the long-wavelength behavior of n_{loc} is summarized by the property that it is neutral, bearing no dipole nor quadrupole. The possibility of decomposing the solid into such a sum of elementary bricks, implies that the potential drop across the surface of any semiinfinite sample would not depend on the details of the surface structure. Intuitively, this is so because the prescription on the localized charge distribution implicitly fixes all the relevant details of the surface density profile (i.e. of the "surface dipole").

Of course, the use of Eqs. (6.2-3) would also solve the lineup problem: the potential lineup across a semiconductor interface would simply be the difference between the average potentials of the two infinite solids, as calculated from Eq. (6.3). As a consequence of that, the potential lineup and the band offset would

be independent of the interface orientation and abruptness, and would also display transitivity.(*)

Heterojunction as perturbation with respect to a reference crystal

As for the band–offset problem, one is interested in the difference $\Delta V = \langle V \rangle_2 - \langle V \rangle_1$ between the electrostatic potential of the two semiconductors constituting the heterojunction. It is not therefore necessary to decompose the *full* bulk densities of the two semiconductors, but it is instead sufficient to decompose only their *difference*.

Let us consider the general heterojunction C_1A_1/C_2A_2. The ionic bare pseudopotentials of the two bulks are:

$$V^o_{cryst_{1,2}}(\mathbf{r}) = \sum_{\mathbf{R}} \left(v_{C_{1,2}}(\mathbf{r} - \mathbf{R}) + v_{A_{1,2}}(\mathbf{r} - \mathbf{R} - \boldsymbol{\delta}) \right), \tag{6.4}$$

where $v_{A_{1,2}}$ and $v_{C_{1,2}}$ are respectively the anionic and cationic potentials, and $\boldsymbol{\delta}$ is the distance between the anionic and cationic sublattices.

It is convenient to consider the heterojunction as a perturbation with respect to a suitable reference crystal, as for instance the *virtual crystal* whose ionic pseudopotentials are the averages between the ionic pseudopotentials of the two bulks:

$$V^o_{virt}(\mathbf{r}) = \sum_{\mathbf{R}} \left(\langle v_C(\mathbf{r} - \mathbf{R}) \rangle + \langle v_A(\mathbf{r} - \mathbf{R} - \boldsymbol{\delta}) \rangle \right), \tag{6.5}$$

where $\langle v_A \rangle = \frac{1}{2} \left(v_{A_1} + v_{A_2} \right)$ is the average of the anionic potentials, and analogously $\langle v_C \rangle$ is the average cationic potential.

Both the two bulks and the heterojunction can be described in terms of a bare perturbation applied to the virtual crystal:

$$V^o(\mathbf{r}) = V^o_{virt}(\mathbf{r}) + \Delta V^o(\mathbf{r}), \tag{6.6}$$

where:

$$\Delta V^o(\mathbf{r}) = \sum_{\mathbf{R}} \left(\sigma^C_{\mathbf{R}} \Delta v_C(\mathbf{r} - \mathbf{R}) + \sigma^A_{\mathbf{R}} \Delta v_A(\mathbf{r} - \mathbf{R} - \boldsymbol{\delta}) \right), \tag{6.7}$$

$\Delta v_{A,C} = \frac{1}{2} \left(v_{A_1,C_1} - v_{A_2,C_2} \right)$, and $\sigma^{A,C}_{\mathbf{R}}$ is an Ising–like variable whose value is $+1$ when an ion of crystal C_1A_1 sits at site $\mathbf{R}$, and -1 otherwise.

(*) The models of Van de Walle and Martin [37] and of Baldereschi, Baroni, and Resta [38], previously discussed, exploit the idea of Eqs. (6.2-3): in the first one, n_{loc} are assumed to be neutral atomic charge distributions, whereas in the second one they correspond to bulk crystal charge densities limited to a single cation-centered (or anion-centered) Wigner-Seitz unit cell.

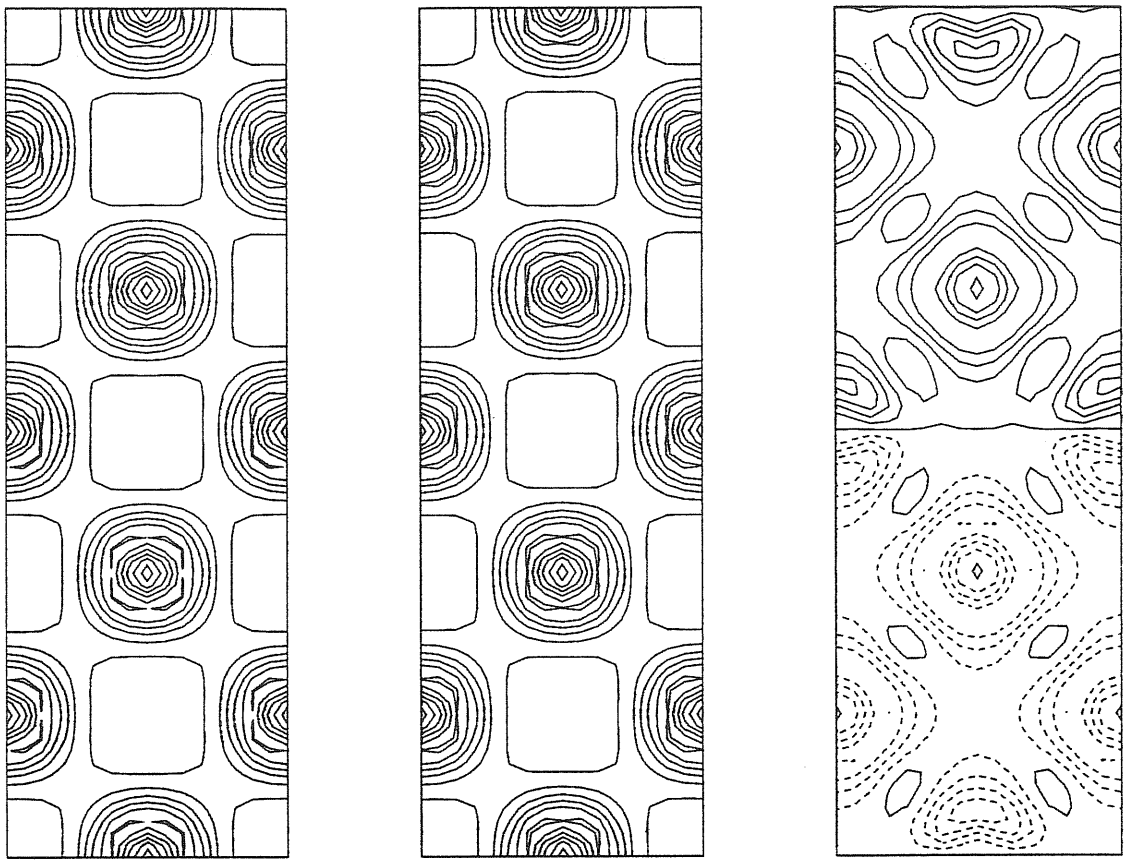

Figure 6.1. Electron-density maps of GaAs/AlAs (001) heterojunction projected onto the (010) plane (left panel), together with the corresponding virtual-crystal density (mid panel), and their difference (right panel). The latter has been plotted with a magnified level spacing ($\times 40$). GaAs down; AlAs up.

Fig. 6.1 shows a comparison between the charge densities of the GaAs/AlAs heterojunction and of the virtual crystal in the (001) supercell. The smallness of the difference between the charge density of the physical system (the heterojunction is this case) and that of the virtual crystal suggests that this difference can be accurately described using *low-order perturbation theory* starting from the virtual-crystal. The charge density response to the perturbation of Eq. (6.6) is:

$$\Delta n(\mathbf{r}) = \int \chi_{virt}(\mathbf{r},\mathbf{r}')\Delta V^o(\mathbf{r}')d\mathbf{r}' + \mathcal{O}\left[(\Delta V^0)^2\right], \tag{6.8}$$

where χ_{virt} is the density-response function of the virtual crystal. According to Eq. (6.7), the difference between the charge density of the physical system and that of the virtual crystal is indeed—to *linear order* in the perturbation—the sum of localized distributions:

$$\Delta n(\mathbf{r}) \approx \sum_{\mathbf{R}} \left(\sigma^C_{\mathbf{R}} \Delta n_C(\mathbf{r}-\mathbf{R}) + \sigma^A_{\mathbf{R}} \Delta n_A(\mathbf{r}-\mathbf{R}-\boldsymbol{\delta})\right) \tag{6.9(a)}$$

$$\Delta n_{A,C}(\mathbf{r}) = \int \chi_{virt}(\mathbf{r},\mathbf{r}')\Delta v_{A,C}(\mathbf{r}')d\mathbf{r}'. \tag{6.9(b)}$$

The Fourier transform of the Coulomb potential generated by the charge distribution of Eq. (6.9(a)) is:

$$\Delta\widetilde{V}(\mathbf{q}) = \frac{1}{N_C\Omega}\sum_{\mathbf{R}} \sigma^C_{\mathbf{R}}\, e^{i\mathbf{q}\cdot\mathbf{R}}\, \frac{4\pi e^2}{q^2}\, \Delta\widetilde{n}_C(\mathbf{q}) + \frac{1}{N_A\Omega}\sum_{\mathbf{R}} \sigma^A_{\mathbf{R}}\, e^{i\mathbf{q}\cdot(\mathbf{R}+\boldsymbol{\delta})}\, \frac{4\pi e^2}{q^2}\, \Delta\widetilde{n}_A(\mathbf{q}) \qquad (6.10)$$

where N_C=N_A=N is the number of cations (anions) in the crystal. From Eq. (6.10) one can calculate (considering the limit $\lim_{\mathbf{q}\to 0}$) the difference of the average electrostatic potential of the physical system with respect to the virtual crystal and, in case of the heterojunction, the difference between the two bulk regions.

Potential lineups from single localized perturbations

According to Eq. (6.3), the main ingredient to calculate the potential lineup from Eq. (6.10) is the limit $\lim_{\mathbf{q}\to 0} \frac{\Delta\widetilde{n}_{loc}(\mathbf{q})}{q^2}$. The localized electron density $\Delta n_{loc}(\mathbf{r})$ (Eq. 6.9(b)) induced by the isolated perturbation has the full point symmetry of the site (T_d for the elemental or binary semiconductors considered in the present work), being the perturbing bare potential spherically symmetric. The long-wavelength behavior of $\Delta\widetilde{n}_{loc}(\mathbf{q})$ is:

$$\Delta\widetilde{n}_{loc}(\mathbf{q}) = Q - i\mathbf{d}\cdot\mathbf{q} - \frac{1}{2}\mathbf{q}\cdot\overleftrightarrow{\mathbf{D}}\cdot\mathbf{q} + Aq^2 + \mathcal{O}(q^3), \qquad (6.11)$$

where Q is the net displaced charge, $\mathbf{d}$ is the induced dipole, $D_{\alpha\beta}$ is the corresponding quadrupole moment, and A is the second spherical moment of the induced charge:

$$Q = \int \Delta n_{loc}(\mathbf{r})d\mathbf{r} \qquad (6.12(a))$$

$$\mathbf{d} = \int \mathbf{r}\Delta n_{loc}(\mathbf{r})d\mathbf{r} \qquad (6.12(b))$$

$$D_{\alpha\beta} = \int (x_\alpha x_\beta - \frac{1}{3}r^2\delta_{\alpha\beta})\Delta n_{loc}(\mathbf{r})d\mathbf{r} \qquad (6.12(c))$$

$$A = \frac{1}{6}\int r^2\Delta n_{loc}(\mathbf{r})d\mathbf{r} \qquad (6.12(d))$$

Because of symmetry, Δn_{loc} does not carry dipole ($\mathbf{d}=0$) nor quadrupole ($\overleftrightarrow{\mathbf{D}}=0$) moments, and therefore Eq. (6.11) reduces to:

$$\Delta\widetilde{n}_{loc}(\mathbf{q}) = Q + Aq^2 + \mathcal{O}(q^3) \qquad (6.13)$$

The total (bare + screening) charge carried by the perturbation vanishes for isovalent substitutions, whereas it is different from zero for heterovalent ones. This

is obvious within LRT, because the total charge is ε_∞^{-1} times that of the bare perturbation, where ε_∞ is the dielectric constant of the reference crystal. It is convenient at this point to distinguish the isovalent from the heterovalent case, the latter showing qualitatively different features.

i) Isovalent interfaces: Q=0

In isovalent interfaces Δn_{loc} does not carry any net charge, and from Eq. (6.13) one obtains for the isolated perturbation:

$$\lim_{\mathbf{q}\to 0} \frac{\Delta\widetilde{n}_{loc}(\mathbf{q})}{q^2} = A. \tag{6.14}$$

Therefore, from Eq. (6.10), one has:

$$\lim_{\mathbf{q}\to 0} \Delta\widetilde{V}(\mathbf{q}) = \frac{4\pi e^2}{N_C\Omega} A_C \sum_{\mathbf{R}} \sigma_{\mathbf{R}}^C + \frac{4\pi e^2}{N_A\Omega} A_A \sum_{\mathbf{R}} \sigma_{\mathbf{R}}^A. \tag{6.15}$$

The potential drop across the interface C_1A_1/C_2A_2 is therefore:

$$\Delta V = \frac{8\pi e^2}{\Omega}(A_C + A_A) \tag{6.16}$$

independent on the interface orientation and even abruptness. In general, in the case of an interface between two alloys characterized for instance by different concentrations x and y of the cations C_1 and C_2, say $C_{1_x}C_{2_{1-x}}A/C_{1_y}C_{2_{1-y}}A$ (e.g. $Ga_xAl_{1-x}As/Ga_yAl_{1-y}As$), the potential lineup is given by :

$$\Delta V = \frac{8\pi e^2}{\Omega} A_C(x-y). \tag{6.17}$$

ii) Heterovalent interfaces: Q≠0

The case of heterovalent interfaces (e.g. between semiconductors of IV and III-V group) requires some care in handling net charges, since in this case $Q \neq 0$.

To better understand this case, I refer to a specific example, i.e. Ge/GaAs. In this case, the appropriate virtual crystal is an artificial zincblende–structure crystal $\langle Ge_{0.5}Ga_{0.5}\rangle\langle Ge_{0.5}As_{0.5}\rangle$, whose cation has valence charge $z_C = 3.5$, and the anion has valence charge $z_A = 4.5$. The relevant (bare) localized perturbations which transform a virtual ion in a real one are:

$$\Delta v_C = \pm\frac{1}{2}(v_{Ga} - v_{Ge}); \quad \Delta v_A = \pm\frac{1}{2}(v_{As} - v_{Ge}), \tag{6.18}$$

where the upper sign transforms a virtual cation (anion) into a gallium (arsenic) ion, whereas the lower one transforms both of them into germanium ions. The bare perturbations of Eq. (6.18) carry a net charge $\Delta z_{A,C} = \pm 0.5$, therefore $Q = \pm\frac{0.5}{\varepsilon_\infty}$.

According to the previous discussions, the Fourier transform of the total charge induced by one localized perturbation is:

$$\Delta\tilde{n}_{A,C}(\mathbf{q}) = \pm\left(\frac{\Delta z_{A,C}}{\varepsilon_\infty} + A_{A,C}q^2\right) + \mathcal{O}(q^3). \tag{6.19}$$

The potential lineup is the sum of two different contributions:

$$\Delta V = \Delta V_Q + \Delta V_{iso}. \tag{6.20}$$

The latter contribution, ΔV_{iso}, is due to neutral localized perturbations, and it is independent on the orientation and abruptness of the interface. The former contribution is due to the part of the perturbation related to the net difference between the ionic charges, and comes from the constant term of Eq. (6.19). It is the same one would obtain for a distribution of point charges of absolute value $\frac{\Delta z_{A,C}}{\varepsilon_\infty}$, centered on the atomic sites of the perturbations $\delta v_{A,C}$. It depends therefore on the orientation and sharpness of the interface, but it can be calculated, once the distribution of the localized perturbations is known, from elementary electrostatics.

Localized response to an isovalent perturbation

The estimate of the isovalent part of the potential lineup requires the knowledge of the charge–density response to a localized perturbation and, in particular, the calculation of its second spherical moment A (see Eq. (6.12(d)). The linear response to any given perturbation can be calculated along the lines indicated in Ref. 89, but I follow here a direct approach, by calculating the response from the difference between two independent calculations for the perturbed and unperturbed reference systems. In practice I calculate the SCF(*) charge density of the systems in a suitably chosen unit cell, and the isolated perturbation consists in replacing only one of the reference atoms in the cell with one of the real atoms: the response $\Delta n_{loc}(\mathbf{r})$ is then simply obtained by difference.

For the sake of definiteness I refer to the case of GaAs/AlAs, but the analysis is valid for any kind of isovalent heterojunction. Following the general procedure sketched before, let us consider the virtual crystal $\langle Ga_{0.5}Al_{0.5}\rangle As$ as reference crystal (unperturbed system). This is an artificial zincblende-structure crystal whose anion is As and whose cation is represented by a pseudopotential which is the arithmetic average of the Ga and Al potentials. I then consider the linear response of the reference crystal to a *single* isovalent substitution, where a mixed cation $\langle Ga_{0.5}Al_{0.5}\rangle$ is replaced by a Ga (or Al) ion. The perturbations leading from the virtual cations to the true ones are:

$$\Delta v_{Ga} = -\Delta v_{Al} = \frac{1}{2}(v_{Ga} - v_{Al}). \tag{6.21}$$

(*) For the parameters used see Sect. 5.1.

It is convenient to introduce a factor λ to measure the strength of the applied perturbation in terms of Δv_{Ga}, for instance: $\lambda = \pm 1$ corresponds to Δv_{Ga} and Δv_{Al} respectively, and $\lambda = 0$ corresponds to the unperturbed case. By developing the charge density response in terms of λ, one formally has:

$$n(\lambda) = n(0) + \lambda n'(0) + \frac{\lambda^2}{2} n''(0) + \frac{\lambda^3}{6} n'''(0) + \ldots \tag{6.22}$$

The charge density responses to Δv_{Ga} and Δv_{Al} are Δn_{Ga} and Δn_{Al}; they are equal in magnitude and opposite in sign only at the linear order, and the differences are due to higher–order terms:

$$\Delta n_{Ga} = \Delta n_{Ga}^{(1)} + \Delta n_{Ga}^{(2)} + \Delta n_{Ga}^{(3)} + \ldots \tag{6.23(a)}$$

$$\Delta n_{Al} = \Delta n_{Al}^{(1)} + \Delta n_{Al}^{(2)} + \Delta n_{Al}^{(3)} + \ldots \tag{6.23(b)}$$

From the two responses one can approximately extract the linear and the quadratic terms, corresponding respectively to n' and $\frac{1}{2}n''$:

$$\Delta n_C^{(1)} = +\Delta n_{Ga}^{(1)} = -\Delta n_{Al}^{(1)} \approx \frac{\Delta n_{Ga} - \Delta n_{Al}}{2} \tag{6.24(a)}$$

$$\Delta n_C^{(2)} = +\Delta n_{Ga}^{(2)} = +\Delta n_{Al}^{(2)} \approx \frac{\Delta n_{Ga} + \Delta n_{Al}}{2} \tag{6.24(b)}$$

Higher–order terms can be obtained, if necessary, by computing the response to the perturbation with values of λ other than ± 1.

Linear and quadratic terms are displayed in Fig. 6.2. The charge density response was obtained from SCF calculations by substituting the central virtual cation with Ga and Al in turn, in a FCC cell with basic vectors of double length (FCC_2) with respect to the unit zincblende cell, and hence containing 16 atoms. From Figs. 6.2 and 6.3 we can see that the rearrangement of the electron density occurs almost entirely within a bulk FCC Wigner-Seitz cell centered at the substitutional site, but there are also some nonnegligible tails spilling out, more clearly visible in the spherical-average plots. For InP/XAs the extension of the responses is slightly larger.

The accuracy of the calculated value of the second spherical moment A, Eq. (6.12(d)), is mainly determined by two factors: i) the size of the supercell $\mathcal{C}$ where the single substitution is actually made, which has to be large enough to describe the isolated perturbation; ii) the size of the region $\mathcal{R}$ contained within $\mathcal{C}$ where the integral is performed.

If $\mathcal{C}$ (and hence $\mathcal{R}$) is not large enough, the tails of the response are necessarily refolded within the supercell itself, and their inclusion in the integral is a source of

error; on the other hand, if a too large integration region $\mathcal{R}$ is used, possible numerical errors, even small, in $\Delta n_{loc}(\mathbf{r})$ significantly affect the value of A for $\mathbf{r}$ far from the center, because of the r^2 weight. It is also important to check the charge neutrality (i.e. the integrated charge response is zero), which is automatically guaranteed by construction only if the integral is performed over all the supercell $\mathcal{C}$. A compromise for the choice of suitable supercells for $\mathcal{C}$ and $\mathcal{R}$ regions is therefore necessary to minimize the inaccuracies. To this purpose, I have quantitatively analyzed the extension of the charge density response by different tests: I have used as $\mathcal{C}$ both a 8–atom simple–cubic (SC) cell and a FCC_2 cell; in both cases I have then performed the integrals 6.12(a) and (d) over regions $\mathcal{R}$ of increasing volume contained within $\mathcal{C}$. The results are reported for GaAs/AlAs in Tab. 6.1; the same test for InP/XAs gives analogous results for the convergence.

$\mathcal{C}$	$\mathcal{R}$	$\Delta n^{(1)}$	$\Delta V^{(1)}$
SC	FCC_1	$+0.01$	-0.30
	SC	0.00	-0.39
FCC_2	FCC_1	$+0.01$	-0.29
	SC	-0.00	-0.40
	FCC_2	0.00	-0.38

Table 6.1. Linear terms of the integrated localized electron density response (electrons/unit bulk cell) and of the corresponding induced electrostatic potential lineup (eV) calculated from the single isolated perturbations $\Delta v_{Ga}, \Delta v_{Al}$ over $\langle \mathrm{Ga}_{0.5}\mathrm{Al}_{0.5} \rangle \mathrm{As}$ in a SC and in a FCC_2 supercell; integrations are performed in different regions $\mathcal{R}$.

In summary, I have checked that using the FCC_2 supercells the perturbations are sufficiently far apart to allow an estimate of the potential lineup with the accuracy of few ten meV ($\approx$ 30 meV in the worst case, referring to the anionic perturbation in InP/XAs); using such a cell, the second spherical moment of the charge density response does not change (within $\approx$ 10 meV) if calculated within a SC cell or over the whole FCC_2 cell. All the results presented in the following refer to this geometry.

6.2 Results

6.2.1 Isovalent common–anion heterojunctions: GaAs/AlAs

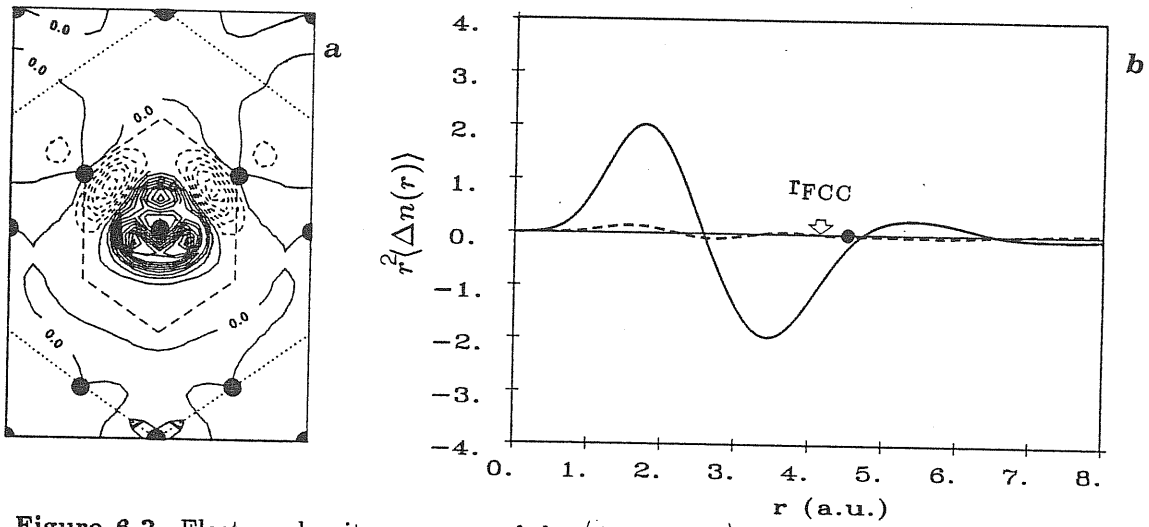

Figure 6.2. Electron density response of the $\langle Ga_{0.5}Al_{0.5}\rangle As$ virtual crystal to a single $\langle Ga_{0.5}Al_{0.5}\rangle \rightarrow Ga$ cationic substitution. **a:** Contour plots of the linear term $\Delta n^{(1)}(\mathbf{r})$ in the (110) plane (dashed lines indicate a negative density response); the intersection of the plane with cation-centered 2-atoms and 16-atoms FCC Wigner-Seitz cells are indicated with dashed and dotted lines. **b:** Radial spherical averages of the linear (solid line) and quadratic (dashed line) density responses. The average radii of the Wigner-Seitz cell and neighbor distances are also indicated.

I display in Fig. 6.2 the linear response to a single cationic substitution in the virtual crystal $\langle Ga_{0.5}Al_{0.5}\rangle As$ as obtained by self-consistent (SCF) calculations for a FCC_2 supercell where a virtual cation is replaced by Ga and Al in turn; the quadratic term is also shown for comparison. The value of the LRT potential lineup obtained according to Eq. (6.16) is +0.39 eV, independent on interface orientation, in agreement with the value +0.41 eV obtained from SCF superlattice calculations.

It is interesting to check whether the LRT reproduces not only the value of the SCF potential lineups, but also the electron–density and potential profiles. To this purpose, it is convenient to study the response to the substitution of a whole atomic plane in the (001), (110), and (111) directions rather than an isolated atom. I start therefore from the virtual crystal XAs and substitute all the X atoms in a cationic plane perpendicular to the growth direction with Ga atoms: the planar average of the induced electron density is shown in Fig. 6.3. The computations have been performed with a 1+5 superlattice geometry.

The electron density induced by a X→Al planar substitution is, to linear order, equal and opposite to the one shown in Fig. 6.3; as for the isolated substitution, the linear term can be evaluated as one half the difference between the responses to X→Ga and X→Al substitutions. The planar average of the charge density response

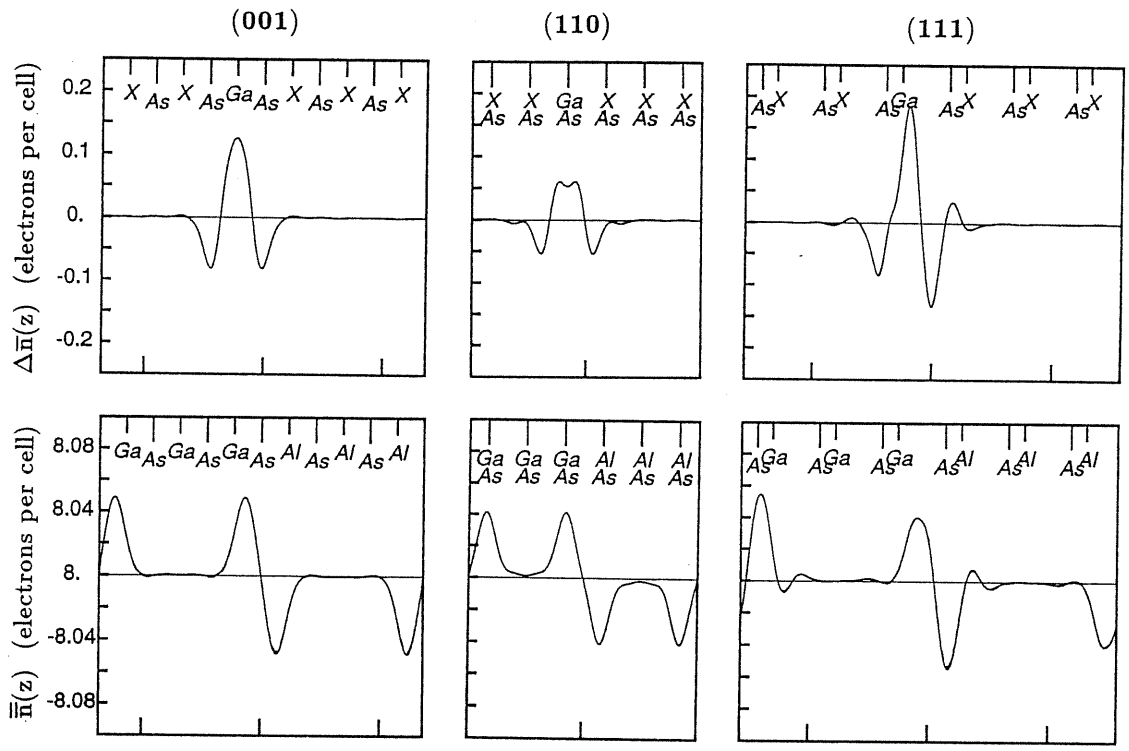

Figure 6.3. Top panels: charge–density response of the virtual crystal $\langle Ga_{0.5}Al_{0.5}\rangle As$ to a single planar cationic substitution. Bottom panel: macroscopic averages of the electron charge densities of GaAs/AlAs. Continuous lines indicate results from full self-consistent calculations, whereas dashed lines indicate results from LRT. Left panels: (001); center: (110); right: (111).

is the only ingredient necessary to build up the full supercell charge within LRT: one starts from the virtual–crystal charge distribution and one simply superposes these responses, centered at the appropriate planes and with the proper sign.

The charge density profiles obtained from SCF and LRT calculations are hardly distinguishable. Note that even the very small oscillations of the charge density in the (111) direction are well reproduced by LRT.(*)

6.2.2 Isovalent no–common–ion heterojunctions: $InP/Ga_{0.47}In_{0.53}As$

For the linear-response treatment, I use the average cation $C=\langle In_{0.5}X_{0.5}\rangle \equiv \langle In_{0.765}Ga_{0.235}\rangle$ and the average anion $A=\langle P_{0.5}As_{0.5}\rangle$ already defined for the study of CP/CAs and InA/XA interfaces, and I consider the CA zincblende-structure virtual crystal as the reference crystal. The actual interface is then recovered substituting C

(*) The WSC model reproduces pretty well for this system the charge density profiles, but not these small oscillations, which are related, as we now can understand in the spirit of LRT, to the outermost tails of the charge density response spilling out from the bulk Wigner–Seitz FCC unit cell.

with In on the left and X on the right of the interface, and analogously the anion A with P and As; the corresponding bare atomic perturbations are respectively:

$$\Delta v_{In} = -\Delta v_X = 0.235(v_{In} - v_{Ga}) \qquad (6.25(a))$$

$$\Delta v_P = -\Delta v_{As} = 0.500(v_P - v_{As}) \qquad (6.25(b))$$

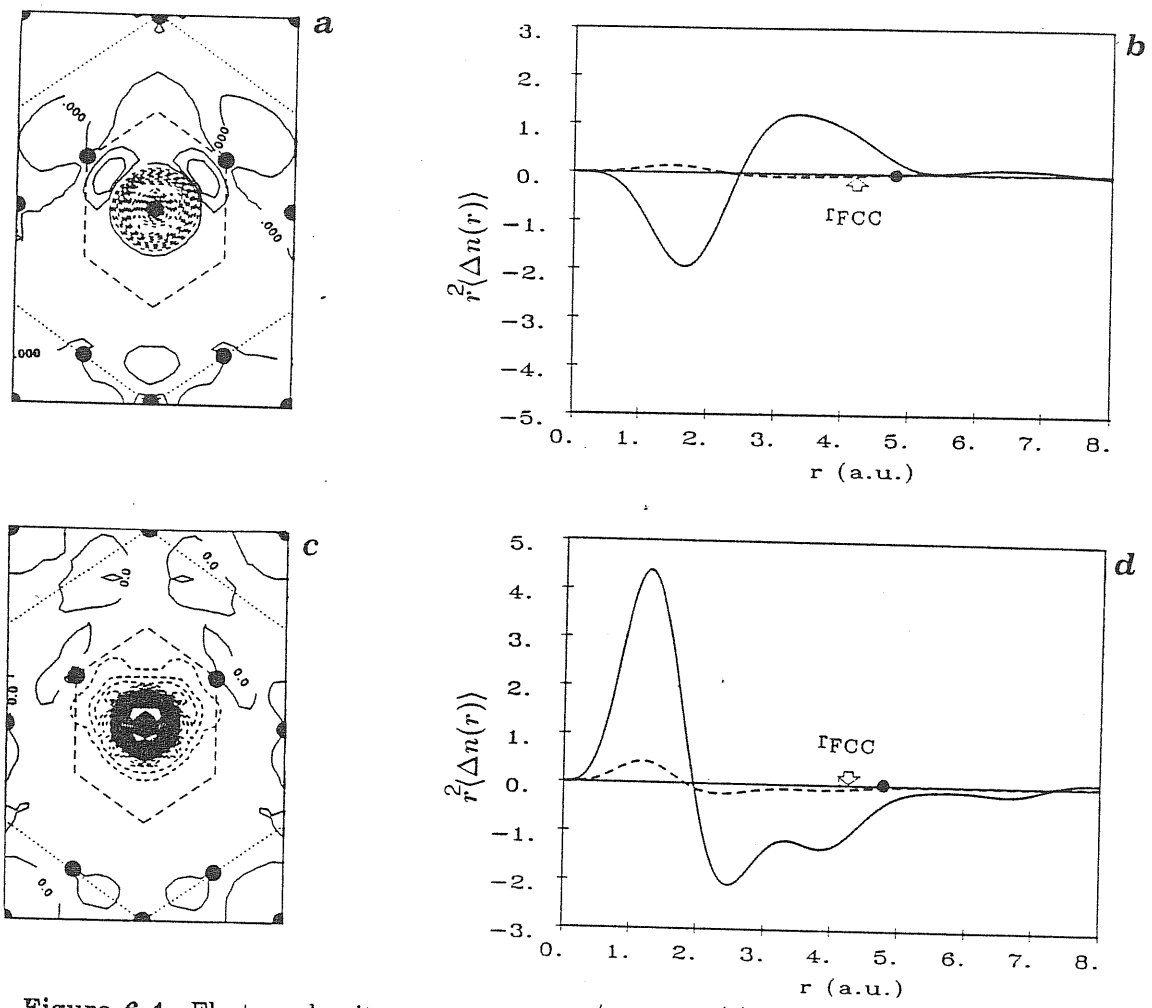

Figure 6.4. Electron density response of the $\langle In_{0.5}X_{0.5}\rangle\langle As_{0.5}P_{0.5}\rangle$ virtual crystal to a single cation (a,b: $\langle In_{0.5}X_{0.5}\rangle \rightarrow$In) or anion (c,d: $\langle As_{0.5}P_{0.5}\rangle \rightarrow$P substitution. a,c: Contour plots of the linear term $\Delta n^{(1)}(\mathbf{r})$ in the (110) plane (dashed lines indicate a negative density response); the intersections of the plane with cation-centered 2-atoms and 16-atoms FCC Wigner-Seitz cells are indicated with dashed and dotted lines. b,d: Radial spherical averages of the linear (solid line) and quadratic (dashed line) density responses. The average radii of the Wigner-Seitz cell and neighbor distances are also indicated.

In Fig. 6.4 I show a contour plot of the linear electron–density response of the virtual crystal to single C→In and A→P substitutions. The spherical average of the linear and quadratic density responses are also reported. From these charge responses one can calculate the anionic and cationic contributions to the potential lineup: the

results are $\Delta V_C = +0.34$ eV, $\Delta V_A = -0.58$ eV, $\Delta V_{tot} = -0.24$ eV, independent on orientation, and in excellent agreement with the values obtained by SCF calculations on supercells (+0.36 eV, −0.61 eV, −0.25 eV respectively). Also for this system, LRT is able to reproduce well even the tiny details of the electron–density profiles, which are undistinguishable from those obtained with SCF supercell calculations (see Fig. 5.11).

6.2.3 Heterovalent heterojunctions: Ge/GaAs

The presence of a constant term in the Fourier transform of the total charge induced by a localized heterovalent substitution (see Eq. (6.13)), originates an interface–dependent contribution ΔV_Q to the electrostatic potential lineup, and hence one can expect new interesting features of the band–offsets in this class of systems.

I examine in some details the VBO at (110) and (001) Ge/GaAs interfaces. In the (110) direction, the virtual crystal is made of atomic planes with two ions (one cation and one anion) per unit surface cell: the average bare charge density is 8 electrons per unit surface cell, for the virtual crystal plane as well as for a plane of Ge or of GaAs. The planar perturbations leading to the physical interface are therefore neutral, and ΔV_Q vanishes. The LRT–potential lineup VBO at (110) interface is therefore $\Delta V(LRT) = \Delta V_{iso} = -0.22$ eV, in good agreement with the SCF result, $\Delta V(SCF) = -0.22$ eV.

In the (001) direction, the virtual crystal is made of an alternating stack of cationic and anionic planes, carrying a surface ionic bare charge density of 3.5 $|e|$ and 4.5 $|e|$ per unit surface cell. This is schematically shown in Fig. 6.5 for the two non–equivalent Ga-Ge– or As-Ge–terminated interfaces. The heterovalent contribution to the potential lineup, ΔV_Q, is due to the planar charge distributions sketched in Fig. 6.5: it is calculated from their macroscopic average through Eq. (3.4), scaled by ε_∞^{-1} to account for the electronic response.(*) Its value is $\Delta V_Q = \pm\frac{\pi e^2}{2a_0\varepsilon_\infty} = \pm 0.31 eV$, and therefore the VBO at $(001)_{Ga}$ and $(001)_{As}$ interfaces is, within LRT and using the value ΔV_{iso} given above for the non–polar (110) interface, +0.85 eV and +0.23 eV respectively; the corresponding results from full SCF supercell calculations were +0.82 eV and +0.22 eV. Note that, according to the LRT formulation, the *average* between the potential lineups of the two non–equivalent (001) interfaces is exactly equal to the electrostatic potential lineup for the non–polar (110) interface. LRT provides a natural and simple explanation of this feature, which was empirically observed by Kunc and Martin several years ago[10].

The response to heterovalent substitutions corresponds to large charge densities (see for instance in Fig. 5.14 their macroscopic average in the (001) interfaces, to

(*) The dielectric constant ε_∞ is taken as the average of the two bulks: $\frac{1}{\varepsilon_\infty} = \frac{1}{2}\left[\frac{1}{\varepsilon_\infty(Ge)} + \frac{1}{\varepsilon_\infty(GaAs)}\right] \approx \frac{1}{13}$; from Ref. 81.

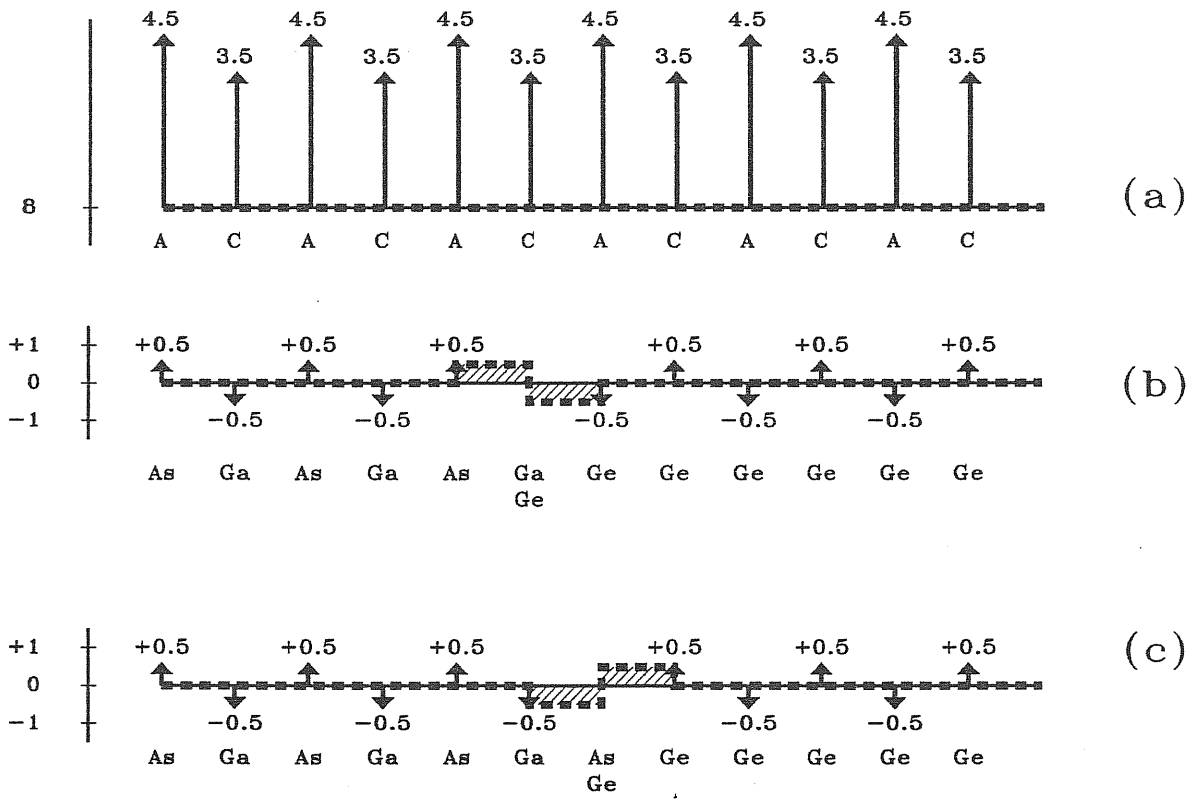

Figure 6.5. (a): Distribution of planar charges for the succession of cationic planes of the CA virtual crystal in the (001) direction. (b): Charge distribution for the perturbation leading to a Ga-Ge–terminated interface. (c): Charge distribution for the perturbation leading to an As-Ge–terminated interface. Arrows indicate delta functions. Dashed lines indicate the macroscopic average of the point-like charge distributions: the corresponding scale (electrons per cell) is displayed on the y axis.

be compared with the corresponding data in the isovalent heterojunctions), and it is more extended than in the isovalent cases studied before; tests performed in 1+5 (110) supercell geometry by substituting only one plane of the virtual crystal with one of Ge or GaAs, have shown that the response extends —with non negligible tails— over the entire supercell, whereas in the isovalent case it is practically limited to the nearest–neighbour planes (see Fig. 6.2 and Fig. 6.3, center panels). Because of this extension, in order to test the capability of LRT of reproducing the SCF electron–density and potential profiles, I had to consider the substitution of a whole atomic plane in a supercell geometry, rather than studying the isolated heterovalent perturbation. Also in this case —as already found for the isovalent heterojunctions— the LRT result is very similar to the full SCF calculation: the curves are practically coincident (see Fig. 5.14), and the resulting VBO is +0.54 eV, +0.82 eV, and +0.22 eV for (110), $(001)_{Ga}$ and $(001)_{As}$ interfaces respectively. The fact that the (110) lineup is not exactly the average between the two values for the (001) interfaces is a consequence of the finite thickness of the supercell used, an effect which is estimated to be $\sim \pm 0.02$ eV.

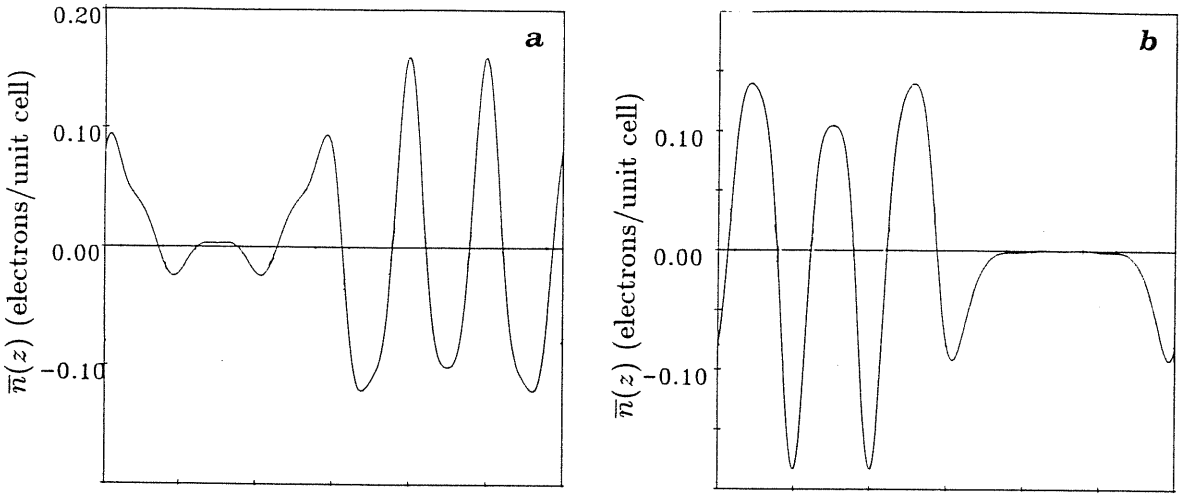

Figure 6.6. Planar average of the differences between the charge density of Ge/GaAs and of GaAs (**a**) or Ge (**b**) in turn.

The problem of supercell thickness therefore does not affect dramaticaly the convergence of the potential lineup in both LRT and SCF calculations, and it is only barely visible in Fig. 5.14 because of the small scale; but inspection of Fig. 6.6 reveals that in Ge/GaAs (110), GaAs bulk features are *not exactly* recovered in the region between two adjacent interfaces; the same has been found in $\langle Ga_{0.5}As_{0.5}\rangle$/GaAs (Fig. 6.7(c)), and for AlAs in Ge/AlAs.

In order to study the dependence of such deviations on the thickness of the supercell, I have performed further SCF supercell calculations; it appears that the deviations are still present even considering a larger supercell (5+3 instead of 3+3): this is illustrated in Figs. 6.7(a) and Fig. 6.7(b) for Ge/GaAs and Ge/AlAs respectively.

It should be noticed that these deviations are mostly evident in the region with the zincblende semiconductor, while bulk features are more easily recovered in the region of the heterojunction with diamond structure.

In order to obtain the response of each one of the two semiconductors to the substitution of one atomic layer of the other material, I have studied the following 7+1 (110) superlattices: seven Ge layers plus one layer of GaAs, and seven GaAs layers plus one layer of Ge. The bare perturbing potentials transforming ions into each other are: $\Delta v_C = v_{Ga} - v_{Ge}$, which transforms Ge atoms into Ga atoms, and $\Delta v_A = v_{As} - v_{Ge}$, which transforms Ge atoms into As atoms. The perturbation transforming an entire layer of Ge into one of GaAs is the superposition of perturbations Δv_C and Δv_A centered on the atomic sites of the layer, and analogously the perturbation transforming an entire layer of GaAs into one of Ge is the superposition of perturbations $-\Delta v_C$ and $-\Delta v_A$: the bare plane perturbations applied to Ge and GaAs are hence equal in magnitude and opposite in sign.

Figs. 6.8(a) and (d) show the macroscopic averages of electron–density response

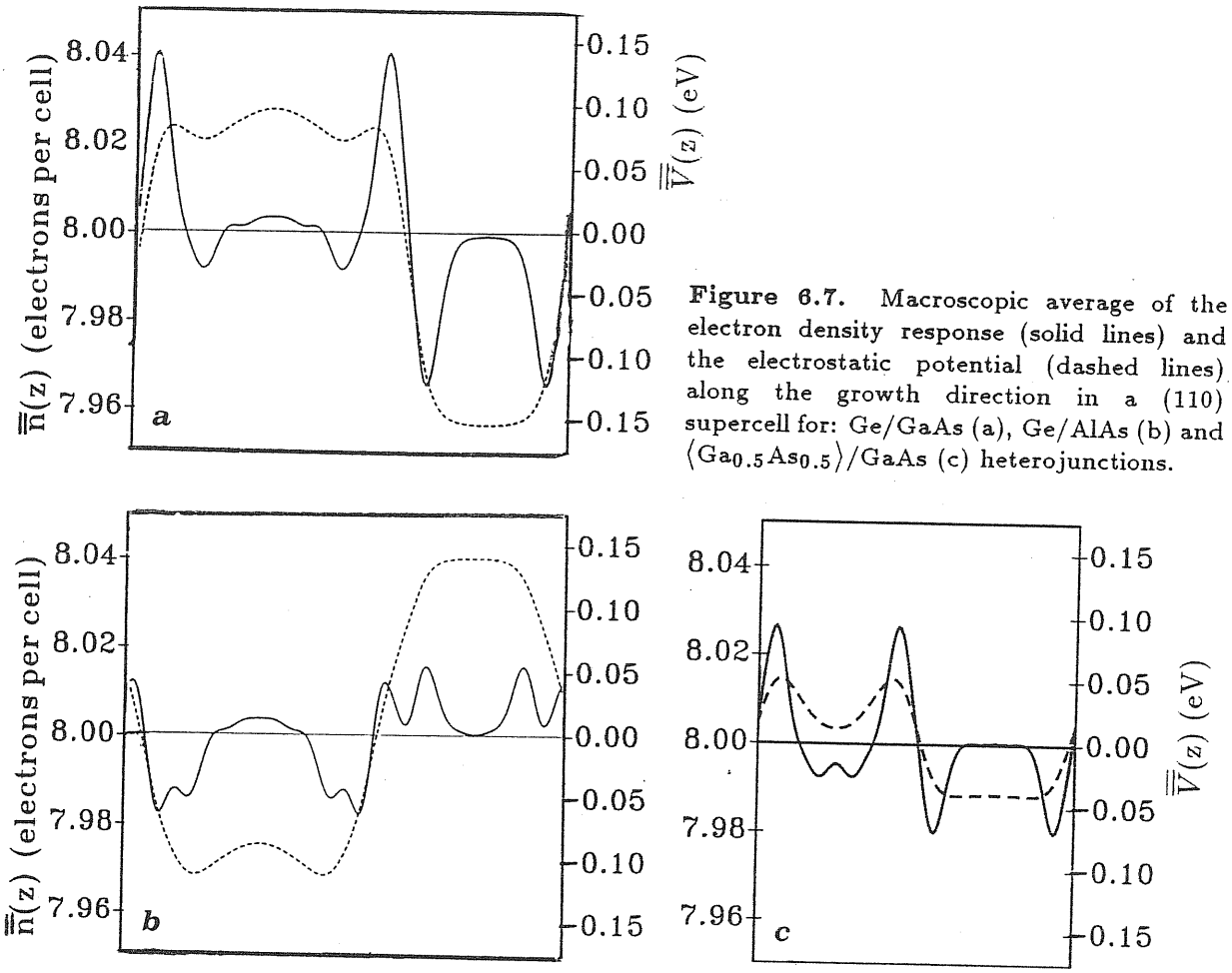

Figure 6.7. Macroscopic average of the electron density response (solid lines) and the electrostatic potential (dashed lines) along the growth direction in a (110) supercell for: Ge/GaAs (a), Ge/AlAs (b) and $\langle Ga_{0.5}As_{0.5}\rangle$/GaAs (c) heterojunctions.

for both situations. Inspection of such figures reveals that, even if the bare perturbing potentials are equal and opposite, the responses look different, reflecting the difference of the two unperturbed systems: the response of bulk Ge to one layer of GaAs is quite localized; conversely, in the case of bulk GaAs with one layer of Ge, the oscillations of the electron–density response are not negligible, and not vanishing even four layers away from the substituted one. One can easily check, e.g. by studying the response to opposite perturbations, that such different behaviours mostly originate from *linear* response effects. This result allows us to better understand our SCF data for the GaAs/Ge 3+3 superlattice. We can in fact imagine to construct the latter system starting from a Ge crystal and periodically transforming three consecutive Ge layers out of six into GaAs layers. Since the response to such perturbation is rather short–range, we obtain that our 3+3 superlattice has a sufficiently thick Ge slab to recover bulk Ge features in the middle of the slab itself. Constructing instead the superlattice from a GaAs crystal and replacing three GaAs layers with Ge ones, the response of GaAs has a quite long range and this explains why with both 3+3 and 5+3 superlattices have not been able to recover bulk features in the middle of the

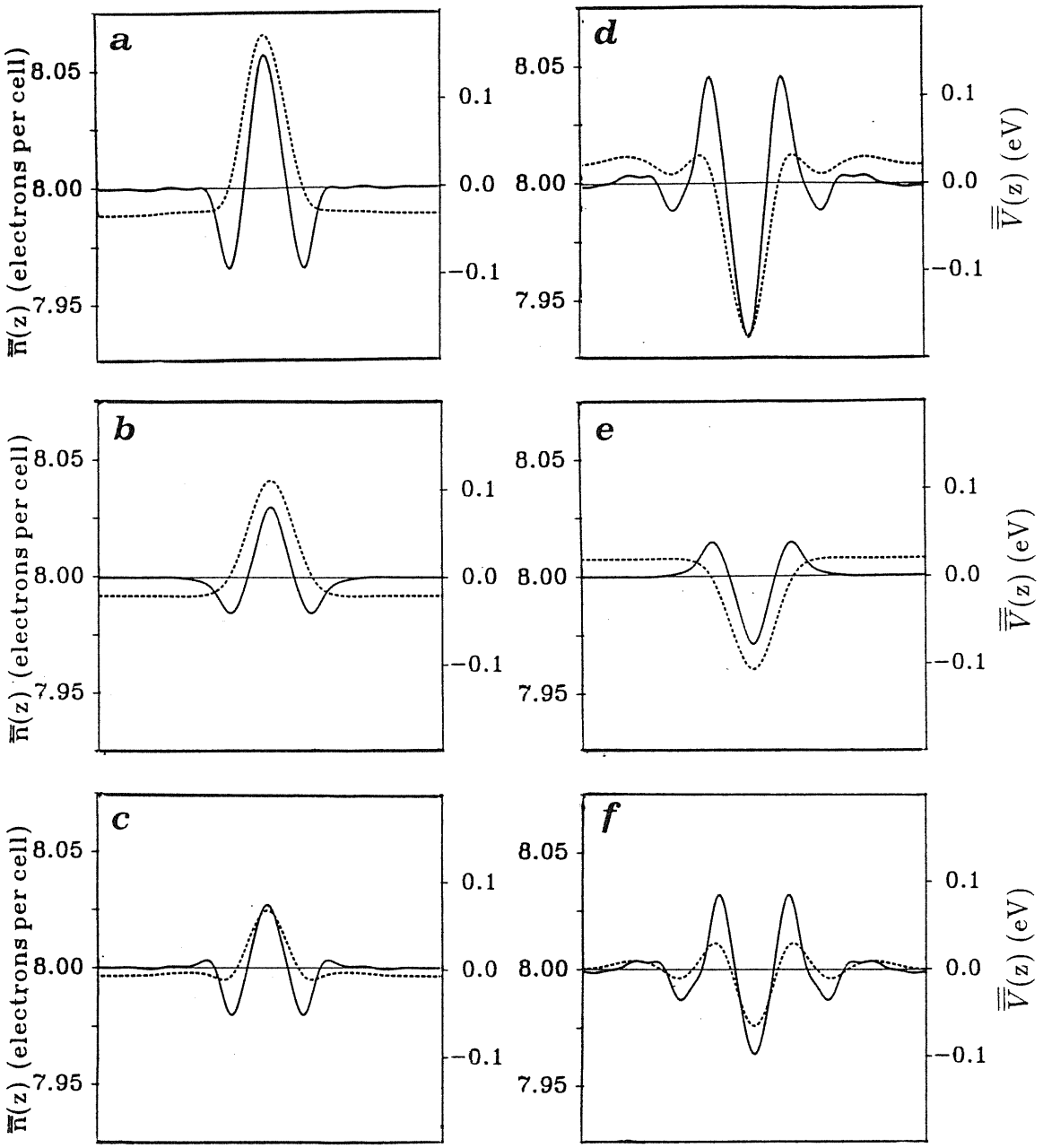

Figure 6.8. Macroscopic average of the electron density response (solid lines) and the electrostatic potential (dashed lines) along the growth direction in a (110) supercell with 8 atomic layers for bulk Ge with one layer of GaAs (**a**), and the decomposition (see text for details) into isovalent (**b**) and heterovalent (**c**) contributions; bulk GaAs with one layer of Ge (**d**), and the decomposition into isovalent (**e**) and heterovalent (**f**) contributions.

GaAs slab.

To better investigate the nature of the different responses, it is convenient to separate the perturbations Δv_A and Δv_C defined before into a *heterovalent* and an *isovalent* part: such procedure is justified since most of the effect under study is linear. To this purpose I introduce the average virtual atom $\langle Ga_{0.5}As_{0.5}\rangle$, and I define the heterovalent part of the perturbation as $\Delta v_{hetero} = v_{Ga} - v_{\langle Ga_{0.5}As_{0.5}\rangle} =$

$v_{\langle Ga_{0.5}As_{0.5}\rangle} - v_{As} = \frac{1}{2}(v_{As} - v_{Ga})$, and the isovalent part as $\Delta v_{iso} = v_{\langle Ga_{0.5}As_{0.5}\rangle} - v_{Ge}$. The perturbation transforming Ge into Ga is thus $\Delta v_C = \Delta v_{iso} + \Delta v_{hetero}$, and that one transforming Ge into As is $\Delta v_A = \Delta v_{iso} - \Delta v_{hetero}$. One can now investigate the response of the two semiconductors to the isovalent and to the heterovalent part of the perturbations separately: bulk Ge with one layer perturbed with $+\Delta v_{iso}$ (see Fig. 6.8(b)) or $\pm\Delta v_{hetero}$ (see Fig. 6.8(c)), and the complementary situations, i.e. bulk GaAs with one layer perturbed with $-\Delta v_{iso}$ (see Fig. 6.8(e)) or $\pm\Delta v_{hetero}$ (see Fig. 6.8(f)). The separation into isovalent and heterovalent perturbations turns out to be useful, since: i) it allows one to introduce symmetry considerations into the problem, since the short–range part of the perturbation Δv_{iso} is symmetric on cation and anion sites whereas the relevant, long–range part Δv_{hetero} is antisymmetric; ii) it shows that the symmetric part Δv_{iso} gives rise to a short–range response only, confirming all the results for isovalent perturbations; iii) it gives a hint about the physical origin of the deviations from bulk behaviour found in the GaAs region of the heterojunction, i.e. the deviations are due to the heterovalent part Δv_{hetero} of the perturbation, which induces long–range responses, which extend at least over four interplane distances away from the tranformed plane.

Considering now the heterovalent part of the perturbation, it is the non–equivalence (equivalence) of the cationic and anionic sites in a zincblende (diamond) semiconductor which determines the features of the response. In the (110) geometry, the plane perturbation is constituted by alternating equal and opposite perturbations $\pm\Delta v_{hetero}$ on cation and anion sites. Its planar average therefore vanishes. When the perturbation is applied to Ge, and in general to a semiconductor with diamond structure, the equal and opposite perturbations are centered on *equivalent* sites: the linear terms of the response are also opposite and their planar average vanishes, and hence the leading terms of the electron density response to the perturbation are the *quadratic* ones. Conversely, when the unperturbed system is GaAs or a semiconductor with zincblende structure, the cation and anion perturbations $\pm\Delta v_{hetero}$ are centered on *non–equivalent* sites, and hence the responses are completely different[90]: consequently also a *linear* term is present in the planar and macroscopic average of the response to the planar perturbation.

I conclude mentioning some potential applications of the peculiarities of polar interfaces to device design. The lineup between non–equivalent interfaces for a given crystallographic orientation could be exploited to induce a potential drop within different regions of a same material. Consider for instance a Ge crystal with an atomic intralayer of GaAs oriented along (001), as represented in Fig. 6.9. From the arguments given above, it is evident that an electrostatic potential lineup $\Delta V = \frac{2\pi e^2}{a_0 \varepsilon_\infty}$ is set up between the left and the right semiinfinite Ge crystals. The possibility of obtaining such an effect depends very much on the sharpness of the GaAs intralayer, and it is not evident whether the present epitaxial techniques allow one to an interface

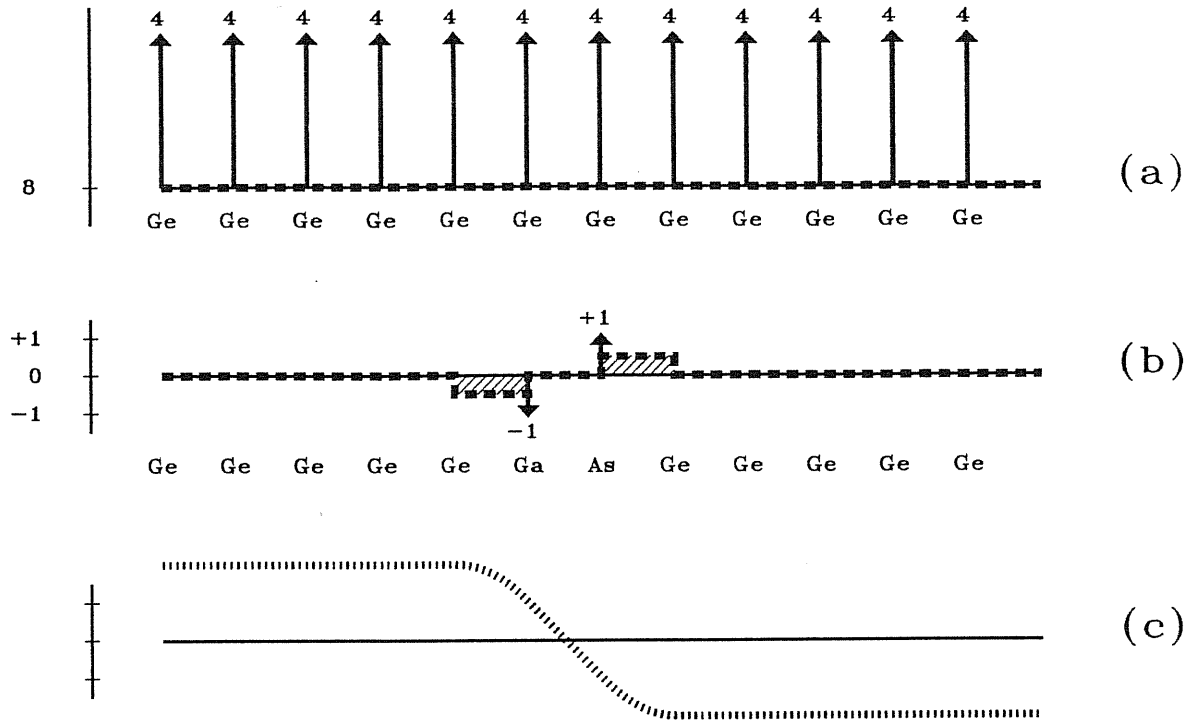

Figure 6.9. (a): Distribution of planar charges for Ge in the (001) direction. (b): Charge distribution for the perturbation corresponding to a GaAs double layer. (c): Electrostatic potential generated by the charge distribution (b).

of sufficient quality(*); however, if this were the case, this would open new perspectives for device design.

(*) Some attempts are under way at the University of Wisconsin (Madison), by G. Margaritondo and cooworkers (private communication).

6.3 Validity of the linear–response theory and its consequences

Higher–order response

A complete formulation of the perturbative approach to the interface problem requires also the knowledge of higher–order terms in the response, and a criterion to choose the most convenient reference unperturbed system. For the sake of clarity, I will refer in the following —if not otherwise specified— to GaAs/AlAs, and I will separate the discussion into two points: i) magnitude of the terms of order higher than linear in the charge density response, both for the atomic and the extended perturbations (i.e. sum of atomic perturbations leading from the virtual crystal to the heterojunction), and ii) effects of such terms, i.e. accuracy of LRT in predicting the electrostatic potential lineup.

Non-linear terms in the charge density response

In order to examine higher–order terms present in the isolated atomic substitution, I consider the perturbation $\Delta v = \Delta v_{Ga} = -\Delta v_{Al}$ transforming the $\langle Ga_{0.5}Al_{0.5}\rangle$ cation in $\langle Ga_{0.5}Al_{0.5}\rangle As$ into Al and Ga cations. From Fig. 6.2(b) (see also Fig. 6.4(b) for InP/XAs) and Tab. 6.2 one can recognize that the quadratic response to the perturbation is i) quite smaller than the linear term, ii) more localized (practically confined within a bulk unit cell within the numerical accuracy of the present calculations).

$\mathcal{C}$	$\mathcal{R}$	$\Delta n^{(2)}$	$\Delta V^{(2)}$
FCC_2	FCC_1	$+0.003$	-0.002
	SC	0.000	-0.038
FCC_2	FCC_1	$+0.003$	-0.001
	SC	0.000	-0.040
	FCC_2	0.000	-0.034

Table 6.2. Quadratic terms of the integrated localized electron density response (electrons/unit bulk cell) and of the corresponding induced electrostatic potential lineup (eV) calculated from the single isolated perturbations $\Delta v_{Ga}, \Delta v_{Al}$ over $\langle Ga_{0.5}Al_{0.5}\rangle As$ in a FCC_2 supercell; integrations are performed in different regions $\mathcal{R}$.

Terms of order higher than quadratic can also be extracted from the total localized response (see Eq. (6.23)): the results I have obtained indicate that cubic

terms are much smaller than quadratic ones, and difficult to estimate since they are of the same order of magnitude as the numerical noise.

How to recognize higher–order terms in the extended perturbation? The quadratic terms neglected in Eq. (6.8) are:

$$\Delta n^{(2)}_{cryst}(\mathbf{r}) = \sum_{\mathbf{R}',\mathbf{R}''} \sigma_{\mathbf{R}'}\sigma_{\mathbf{R}''} \int\int \chi^{(2)}_{virt}(\mathbf{r};\mathbf{r}',\mathbf{r}'')\Delta v(\mathbf{r}'-\mathbf{R}')\Delta v(\mathbf{r}''-\mathbf{R}'')d\mathbf{r}'d\mathbf{r}'' \quad (6.27)$$

The sum contains both terms of type $(\Delta v(\mathbf{r}' - \mathbf{R}'))^2$ and $\Delta v(\mathbf{r}' - \mathbf{R}')\Delta v(\mathbf{r}'' - \mathbf{R}'')$ with $\mathbf{R}' \neq \mathbf{R}''$. The former (I call them *diagonal* terms) are the quadratic responses to isolated perturbations: as observed before, they are not negligible, but have the advantage that can be extracted and transferred from the isolated perturbation. The latter terms (*off-diagonal*) cannot be extracted from the single atomic perturbations, since they are properly due to the interference between responses to adjacent perturbations.

In order to estimate the *off-diagonal* quadratic terms, I compare the full response to adjacent perturbations with the superposition of the localized responses to single perturbations. I consider e.g. in the (110) supercell geometry the GaAs/AlAs heterojunction obtained from full SCF calculations and that obtained with a superposition on the reference crystal $\langle Ga_{0.5}Al_{0.5}\rangle As$ of responses to the perturbation leading from a plane of $\langle Ga_{0.5}Al_{0.5}\rangle$ to a plane of Ga or Al (see Fig. 6.10). In the comparison, all the orders of the localized responses are included (in particular also the *diagonal* quadratic terms), and the differences are only due to *off-diagonal* terms.(*) Contrary to the *diagonal* quadratic terms, the *off-diagonal* terms are very small and cannot be estimated since they are masked by the numerical noise; their effects are negligible with respect to the numerical accuracy of the results presented here.

These conclusions strictly apply only to the case of GaAs/AlAs, since the magnitude of the *off-diagonal* terms depends on the extension of the localized responses which may vary from one system to another: tests performed for instance for InP/XAs indicate that in such system the *off-diagonal* terms are again very small, but are distinct from the numerical noise and are not completely negligible.

Effects of non-linear terms on the electrostatic potential lineup

Given that non–linear terms are present in the charge–density response, the question arises of how much they can affect the electrostatic potential (and hence of the VBO) obtained with the LRT.

(*) More precisely, the interference terms between responses to perturbations on the same anionic plane are included in the response to a whole planar substitution, and the *off-diagonal* terms we are looking at in this procedure are only interferences between planes.

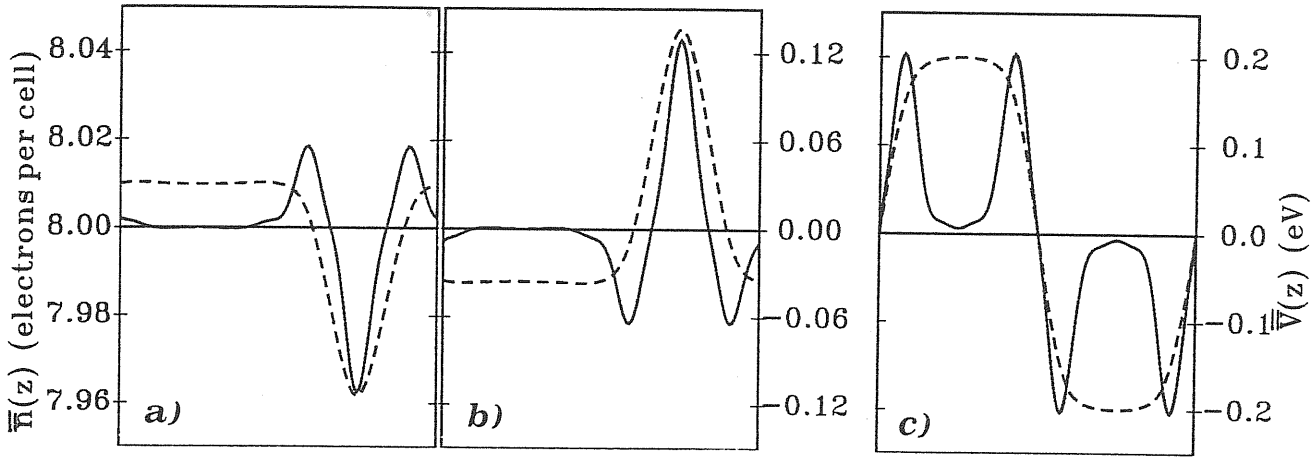

Figure 6.10. Macroscopic average of the electron density response (solid lines) to the perturbations leading from a (110) plane of $\langle Ga_{0.5}Al_{0.5}\rangle$ in $\langle Ga_{0.5}Al_{0.5}\rangle$ As to a plane of Al (a) or Ga (b), and induced electrostatic potential (dashed lines). (c): macroscopic average of the electron density and electrostatic potential in the GaAs/AlAs (110) heterojunction as obtained from full SCF supercell calculations and from the reference $\langle Ga_{0.5}Al_{0.5}\rangle$As crystal with superposition of the planar responses: on this scale the two cases are not distinguishable.

As for the *off–diagonal* terms, if they were important, no low–order perturbative approach starting from the isolated perturbation would provide meaningful results. Fortunately this is not the case: we have seen in fact that they are small. However, they can be different in magnitude according to the systems considered, and therefore also their influence on the electrostatic potential lineup can vary: in the case of GaAs/AlAs it is negligible and smaller than the accuracy limit, whereas in the case of CP/CAs (which correspond to the strongest isovalent perturbation that I have considered) it amounts to $\approx +0.03$ eV.

As for the *diagonal* quadratic terms, they are not negligible, and it would be desirable if one could define a reference system such that their contribution to the lineup rigorously vanishes. One can easily check, for instance, that such situation is indeed verified in GaAs/AlAs, considered as a perturbation with respect to the average virtual crystal: the *diagonal* quadratic responses to single cation perturbations are the same on every site —both in sign and in magnitude—, and their contribution to the charge density response in the whole heterojunction has the periodicity of the bulk and hence their effect on the electrostatic potential lineup vanishes.

The situation described for GaAs/AlAs can be realized for a general heterojunction: provided one is able to choose a reference unperturbed system such that the quadratic *diagonal* terms are identical on both sides of the interface, independently of the sign $\sigma_{\mathbf{R}}$ of the applied perturbation. This is always possible, and is guaranteed if the reference crystal is the average, atom by atom, between the two solids forming the interface. A way to construct properly the reference crystal for a $n+n$ superlat-

tice, is to consider the average, in terms of bare atomic potentials, of the supercell representing the heterojunction and another one obtained from the former with a shift of half the period along the growth axis. Such a reference crystal takes into account both chemical and structural differences between the two semiconductors constituting the heterojunction (see e.g. the case of Ge/GaAs).

I thus conclude that neglecting the *diagonal* quadratic terms may not give a very good description of the 3–D details of the charge–density distribution, but fortunately —provided the reference unperturbed system is properly chosen— this does not affect the estimate of the electrostatic potential lineup. To stress again the importance of the choice of the reference system, note that an effect which is *linear* with the use of one reference unperturbed system may manifest itself as *quadratic* by changing the reference; in this sense one cannot make the absolute claim that the quadratic terms are negligible and that the *linear order response* is sufficient for the interface problem, but, rather, that LRT works well with a suitable choice of the reference system.

To clarify these comments, I discuss the validity of the VCA in the description of the GaInAs alloy in the InP/$\langle Ga_{0.5}In_{0.5}\rangle$As heterojunction. SCF calculations have shown that the electrostatic potential lineup in InP/$\langle Ga_{0.5}In_{0.5}\rangle$As is -0.25 eV, while the one in InP/$(GaAs)_1(InAs)_1$ is -0.19 eV, i.e. they differ by 0.06 eV. Provided that the transitivity rule holds, as it has been demonstrated for the class of isovalent heterojunctions, the above difference of 0.06 eV should correspond to the electrostatic potential lineup at the $\langle Ga_{0.5}In_{0.5}\rangle$As/$(GaAs)_1(InAs)_1$ heterojunction. At a first sight and using LRT, one expects the latter lineup to be zero. What is therefore responsible for such an effect? Can it be predicted by a low–order perturbation approach?

i) To answer this question, I consider the $\langle Ga_{0.5}In_{0.5}\rangle$As/$(GaAs)_1(InAs)_1$ heterojunction(*), starting from the average virtual crystal C_2A, where $C_2=\langle\langle Ga_{0.5}In_{0.5}\rangle In_{0.5}\rangle$ $=\langle Ga_{0.25}In_{0.75}\rangle$. The perturbations transforming the reference cation C_2 into Ga, In, and $\langle Ga_{0.5}In_{0.5}\rangle$ are $3\Delta v$, $-\Delta v$, and Δv, respectively, where $\Delta v = \frac{1}{4}(v_{Ga} - v_{In})$; the extended cationic perturbation transforming C_2A into $\langle Ga_{0.5}In_{0.5}\rangle$As/$(GaAs)_1(InAs)_1$ heterojunction is shown in Fig. 6.11. The proper filter function for the macroscopic average extends over the tetragonal cell containing four atoms, which is the unit cell of the $(GaAs)_1(InAs)_1$ monolayer superlattice. In each one of such cells, the linear terms add up to the same total effect on the two sides of the heterojunction and do not contribute to the lineup, but the *diagonal* quadratic terms $\Delta n^{(2)}_{loc}$ are present with different weights: $+1$ on all the $\langle Ga_{0.5}In_{0.5}\rangle$ sites in the $\langle Ga_{0.5}In_{0.5}\rangle$As region, while in the $(GaAs)_1(InAs)_1$ region they are $+1$ on the In site and $+9$ on the Ga site. They

(*) To compare the alloy with the $(GaAs)_1(InAs)_1$ monolayer superlattice, I consider here the alloy with the composition 50% of both Ga and In, instead of 47% and 53% respectively.

consequently cause a difference in the average electrostatic potential on the two sides of the heterojunction $\Delta V^{(2)} = \frac{16\pi e^2}{\Omega} A = +0.05$ eV (with $A = \frac{1}{6}\int r^2 \Delta n^{(2)}_{loc}(\mathbf{r}) d\mathbf{r}$), which agrees, within the numerical accuracy, with the full SCF supercell result. The quadratic *diagonal* terms are thus responsible for the effect.

ii) Let us consider again the $\langle Ga_{0.5}In_{0.5}\rangle As/(GaAs)_1(InAs)_1$ heterojunction, and ask whether exists a reference system such that the effect of the *diagonal* terms on the electrostatic potential lineup vanish. Such reference average periodic crystal is a monolayer superlattice $(C_1As)_1(C_2As)_1$ with a tetragonal unit cell containing two anions and the cations $C_1 = \langle\langle Ga_{0.5}In_{0.5}\rangle Ga_{0.5}\rangle = \langle Ga_{0.75}In_{0.25}\rangle$ and C_2 as defined above. If I define as above $\Delta v = \frac{1}{4}(v_{Ga} - v_{In})$, for instance, the bare atomic perturbations transforming the virtual cations C_1 into Ga, In, and $\langle Ga_{0.5}In_{0.5}\rangle$ are Δv, $-3\Delta v$, and $-\Delta v$ respectively; the perturbations transforming the virtual cations C_2 into Ga, In, and $\langle Ga_{0.5}In_{0.5}\rangle$ are $3\Delta v$, $-\Delta v$, and Δv respectively. By considering carefully now the perturbation that transforms $(C_1As)_1(C_2As)_1$ into the heterojunction (see Fig. 6.12), one can realize that the lineup due to the electronic response to the perturbations on the C_1 atoms is not exactly cancelled by that on the C_2 atoms, since the perturbations are equal but they are centered on non-equivalent atomic sites: the responses to a given perturbation centered on the site C_1 and on the site C_2 will in general be different, and this is the reason of the non vanishing electrostatic potential lineup at this interface.

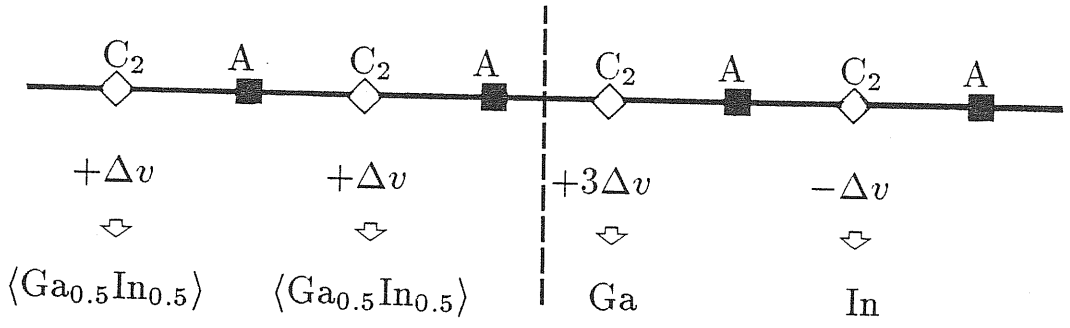

Figure 6.11. Extended cationic perturbation transforming the average reference crystal C_2A (see text for the definitions) into $\langle Ga_{0.5}In_{0.5}\rangle As/(GaAs)_1(InAs)_1$ heterojunction.

Note that the low-order perturbation theory not only reproduces the SCF result, but extends it to any alloy. In fact, the contribution of the *diagonal* quadratic terms to the electrostatic potential lineup does not depend on the particular distribution of the single atomic substitutions; the result is therefore the same for any other configuration describing the real alloy, it does not depend on the particular cationic

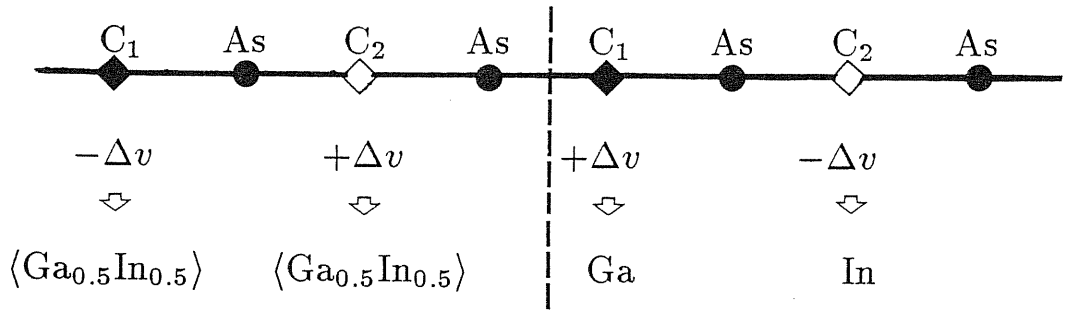

Figure 6.12. Extended cationic perturbation transforming the average reference crystal $(C_1As)_1(C_2As)_1$ (see text for the definitions) into $\langle Ga_{0.5}In_{0.5}\rangle As/(GaAs)_1(InAs)_1$ heterojunction.

ordering but only on the fact that *real* atoms and not *virtual* atoms have to be considered.

I stress that the above considerations are limited to the electrostatic potential lineup. I have shown –from SCF–LDA bulk calculations— that also the bulk band contribution ΔE_v to the VBO in InP/GaInAs heterojunctions is different in the two situations, i.e. in the case of a gallium–indium arsenide alloy treated with the VCA, and in the case of an ordered $(GaAs)_1(InAs)_1$ monolayer superlattice.

Summary

I conclude mentioning the main achievements of the LRT study. Besides reproducing quantitatively the SCF results, LRT allows to evidentiate a series of general features governing the physics of VBO's, namely:

i) independence of the VBO on interface orientation for isovalent heterojunctions;
ii) independence of the VBO on the abruptness of the interface region for the isovalent heterojunctions;
iii) additivity of anionic and cationic effects in the isovalent heterojunctions with no common ions;
iv) evidence of an interface–dependent contribution to the electrostatic potential lineup in the heterovalent heterojunctions, and relationship between the VBO's at polar and non–polar interfaces;
v) importance of the choice of the reference system for the perturbative approach, in order to neglect the quadratic terms in the response;
vi) effect on the electrostatic potential lineup of the VCA used instead of considering *real* atoms in describing alloys, and independence on their configuration (random or ordered).

consequently cause a difference in the average electrostatic potential on the two sides of the heterojunction $\Delta V^{(2)} = \frac{16\pi e^2}{\Omega} A = +0.05$ eV (with $A = \frac{1}{6}\int r^2 \Delta n^{(2)}_{loc}(\mathrm{r}) d\mathrm{r}$), which agrees, within the numerical accuracy, with the full SCF supercell result. The quadratic *diagonal* terms are thus responsible for the effect.

ii) Let us consider again the $\langle Ga_{0.5}In_{0.5}\rangle As/(GaAs)_1(InAs)_1$ heterojunction, and ask whether exists a reference system such that the effect of the *diagonal* terms on the electrostatic potential lineup vanish. Such reference average periodic crystal is a monolayer superlattice $(C_1As)_1(C_2As)_1$ with a tetragonal unit cell containing two anions and the cations $C_1 = \langle\langle Ga_{0.5}In_{0.5}\rangle Ga_{0.5}\rangle = \langle Ga_{0.75}In_{0.25}\rangle$ and C_2 as defined above. If I define as above $\Delta v = \frac{1}{4}(v_{Ga} - v_{In})$, for instance, the bare atomic perturbations transforming the virtual cations C_1 into Ga, In, and $\langle Ga_{0.5}In_{0.5}\rangle$ are Δv, $-3\Delta v$, and $-\Delta v$ respectively; the perturbations transforming the virtual cations C_2 into Ga, In, and $\langle Ga_{0.5}In_{0.5}\rangle$ are $3\Delta v$, $-\Delta v$, and Δv respectively. By considering carefully now the perturbation that transforms $(C_1As)_1(C_2As)_1$ into the heterojunction (see Fig. 6.12), one can realize that the lineup due to the electronic response to the perturbations on the C_1 atoms is not exactly cancelled by that on the C_2 atoms, since the perturbations are equal but they are centered on non–equivalent atomic sites: the responses to a given perturbation centered on the site C_1 and on the site C_2 will in general be different, and this is the reason of the non vanishing electrostatic potential lineup at this interface.

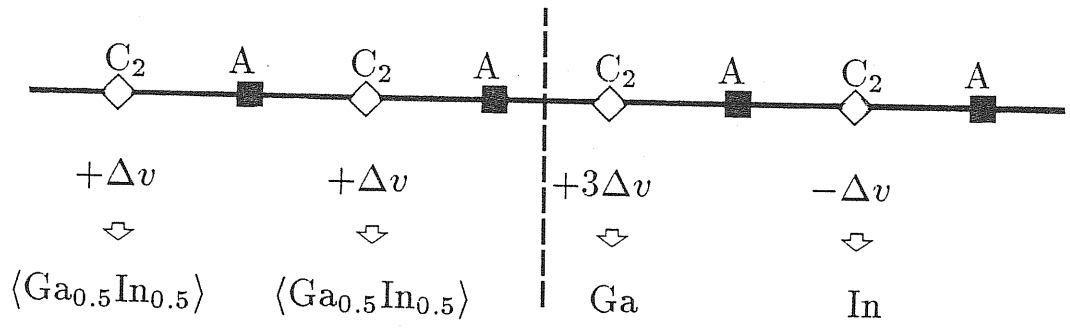

Figure 6.11. Extended cationic perturbation transforming the average reference crystal C_2A (see text for the definitions) into $\langle Ga_{0.5}In_{0.5}\rangle As/(GaAs)_1(InAs)_1$ heterojunction.

Note that the low–order perturbation theory not only reproduces the SCF result, but extends it to any alloy. In fact, the contribution of the *diagonal* quadratic terms to the electrostatic potential lineup does not depend on the particular distribution of the single atomic substitutions; the result is therefore the same for any other configuration describing the real alloy, it does not depend on the particular cationic

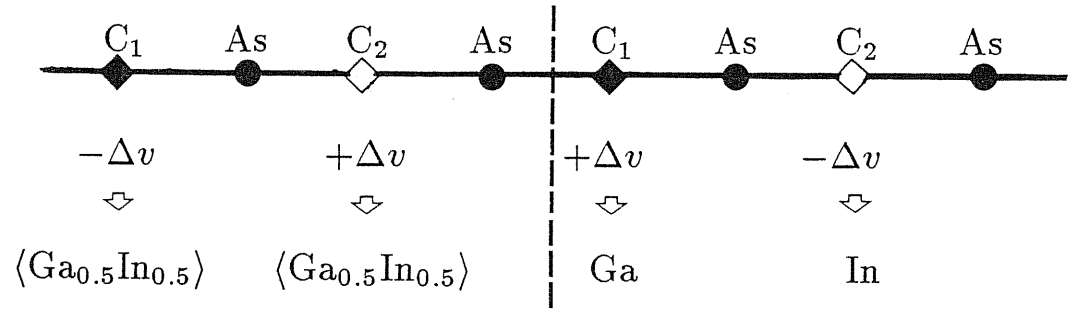

Figure 6.12. Extended cationic perturbation transforming the average reference crystal $(C_1As)_1(C_2As)_1$ (see text for the definitions) into $\langle Ga_{0.5}In_{0.5}\rangle As/(GaAs)_1(InAs)_1$ heterojunction.

ordering but only on the fact that *real* atoms and not *virtual* atoms have to be considered.

I stress that the above considerations are limited to the electrostatic potential lineup. I have shown –from SCF–LDA bulk calculations— that also the bulk band contribution ΔE_v to the VBO in InP/GaInAs heterojunctions is different in the two situations, i.e. in the case of a gallium–indium arsenide alloy treated with the VCA, and in the case of an ordered $(GaAs)_1(InAs)_1$ monolayer superlattice.

Summary

I conclude mentioning the main achievements of the LRT study. Besides reproducing quantitatively the SCF results, LRT allows to evidentiate a series of general features governing the physics of VBO's, namely:

i) independence of the VBO on interface orientation for isovalent heterojunctions;
ii) independence of the VBO on the abruptness of the interface region for the isovalent heterojunctions;
iii) additivity of anionic and cationic effects in the isovalent heterojunctions with no common ions;
iv) evidence of an interface–dependent contribution to the electrostatic potential lineup in the heterovalent heterojunctions, and relationship between the VBO's at polar and non–polar interfaces;
v) importance of the choice of the reference system for the perturbative approach, in order to neglect the quadratic terms in the response;
vi) effect on the electrostatic potential lineup of the VCA used instead of considering *real* atoms in describing alloys, and independence on their configuration (random or ordered).

CONCLUSIONS

The present work has allowed to classify lattice–matched semiconductor heterojunctions and to identify some general trends of the corresponding VBO's. The VBO has been proven to be independent on the interface orientation and abruptness for the isovalent heterojunctions, even in the general case of heterojunctions with no common ions; in the latter case, moreover, the interface dipole is the sum of anionic and cationic contributions, which separately do not depend on the details of the interface. On the other hand, the VBO at heterovalent heterojunctions is strongly dependent on the interface details, and in particular the composition of the interface layer strongly controls the stability of the interface itself and its energy–band discontinuities; fortunately, the interface–dependent contribution can be easily calculated, once the interface details are known, from simple macroscopic quantities (bulk lattice parameters and dielectric constants of the constituents). These properties, which are in agreement with experimental evidencies, have been obtained from first-principles supercell calculations performed for prototype systems, and have been justified and extended to entire classes of heterojunctions by the LRT approach, which has been shown to be a fundamental tool in this problem of semiconductor physics.

During this work some questions have been investigated only in a preliminar way, and they require further study, both with the techniques developed in this work and with new tools. For instance, the effects on the electrostatic potential lineup of internal lattice distortions and interfacial strain have been obtained with simple estimates based on effective–charge schemes; these results have to be accurately checked with SCF calculations. More specifically, a quantitative analysis will require the knowledge of the effective forces acting on the ions and the determination of their equilibrium positions. No efforts in this direction has yet been done. Internal lattice distortions are particularly important in the case of alloys. Internal distortions are present in the real alloy in order to keep the bond lengths close to their values in the respective bulks, and have to be taken into account in this problem. In particular one should determine how much they affect the VBO at semiconductor heterojunctions when one or both materials consist of a solid solution. Furthermore for alloys the treatment of the electronic term is trivial within LRT, using the VCA, but a realistic estimate requires also to take into account the localized quadratic responses.

In the present work, the case of mismatched heterojunctions and superlattices has not been treated. It is closely related to the problem of the behaviour of the energy bands under strain in a single semiconductor. The shifts of the bands under strain are described through *deformation potentials*, which describe the *relative* change in energy of different electronic states (e.g. band–gap changes under pressure), or also changes in the band energies with respect to an *absolute* reference. The latter

potentials are called *absolute deformation potentials*[73,91], and the problem of deriving them is actually related to that of calculating the band offset at the interfaces, and more generally to the existence of an absolute energy scale. It should be stressed that the LRT approach can be applied also to the problem of lattice–mismatched interfaces. One should start again from a reference periodic crystal, and then consider two independent perturbations: substitution of the reference virtual atoms with the real ones and inclusion of strain; in particular, the screened perturbation describing the deformation potential is the superposition of perturbations induced by the displacements of each ion separately. Some preliminar works in this direction[91] have indicated that for group-IV semiconductors, e.g. Si/Ge, when ionic effective charges vanish, the band offset remains essentially a bulk property. It is not evident at present if this result can be generalized to other systems.

Concerning lattice–mismatched interfaces, further studies are necessary for the heterojunction InP/$Ga_{1-x}In_xAs$, that I have studied here only for x =53%, i.e. in the case of lattice matching. In fact, it is nowadays possible to grow strained layers of $Ga_{1-x}In_xAs$ on InP substrates over the entire range $0 < x < 1$ of In composition; in particular a type–I to type–II superlattice transition is observed at $x \approx 0.2\%$.[92]

Finally, the possibility to introduce intralayers in the interface region has been considered recently[33,34,35]. How much these intralayers will affect the VBO has yet to be determined. According to the LRT approach, the substitution itself should not strongly affect the VBO in the case of isovalent materials, but a possible induced interfacial strain might be important. On the other hand, intralayers of heterovalent materials can dramaticaly change the electronic structure of the heterojunction. On this subject, I have given in this work an example showing the peculiarities of polar interfaces and of heterovalent intralayers, and their possible use to tune the VBO and to design devices with novel properties. This field, which is very promising, also requires further investigations.

References

[1] *Molecular Beam Epitaxy and Heterostructures*, NATO Advanced Science Institute Series, edited by L.L. Chang and K. Ploog (Nijhoff, Dordrecht, 1985).

[2] *Metal Organic Vapor Phase Epitaxy*, edited by J.B. Mullin, S.J.C. Irvine, R.H. Moss, P.N. Robson and D.R. Wight (North–Holland, Amsterdam, 1984).

[3] L. Esaki and R. Tsu, IBM J. Res. Develop. **14**, 61 (1970).

[4] R.S Bauer and G. Margaritondo, Phys. Today **40**, 27 (1987); see also the issues: J. Vac. Sci. and Technol. B **4** No. 4 (1986), and B **5**, No. 4 (1987).

[5] See, e.g., *Heterojunction Band Discontinuities: Physics and Device Application*, edited by F. Capasso and G. Margaritondo (North–Holland, Amsterdam, 1987).

[6] L. Esaki, Journal de Physique, Colloque C5, 1 (1987).

[7] V. Narayanamurti, in 18th Int. Conf. on The Physics of Semiconductors (Stokolm), edited by O. Engström, p. 3 (World Scientific, Singapore, 1987).

[8] See for instance the review of H.L. Stormer, *Festkorperprobleme* Vol. XXIV, edited by H.J. Queisser, p. 25 (1984).

[9] C.P. Kuo, S.K. Vong, R.M. Cohen, and G.B. Stringfellow, J. Appl. Phys. **57**, 5428 (1985).

[10] K. Kunc and R.M. Martin, Phys. Rev. B **35**, 8154 (1987).

[11] H. Nelson, RCA Rev., 603 (1963).

[12] M.B. Panish, S. Summski and I. Hayashi, Metall. Trans. **2**, 795 (1971).

[13] J. Szuber, Thin Solid Films **105**, 33 (1983); O. Tejayadi, Y.L. Sun, J. Klem, R. Fisher, M.V. Klein and H. Morkoc, Solid State Commun. **46**, 251 (1983).

[14] R.S. Bauer and H.W. Sang Jr, Surf. Sci. **132**, 479 (1983).

[15] (a) W.M. Stobbs, J. de Physique, Colloque C5, 33 (1987); (b) H. Oppolzer, J. de Physique, Colloque C5, 65 (1987).

[16] P. Perfetti, Surf. Sci. **189/190**, 362 (1987).

[17] R. Dingle, W. Weigmann and C.H. Henry, Phys. rev. Lett. **33**, 827 (1974).

[18] R.C. Miller, A.C. Gossard and D.A. Kleinman, Phys. Rev. B **32**, 5443 (1985).

[19] H. Kroemer, Surf. Sci. **132**, 543 (1983).

[20] A.G. Milnes and D.L. Feucht, *Heterojunctions and Metal-Semiconductor Junctions* (Academic Press, New York, 1972).

[21] W.I. Wang, T.S. Kuan, E.E. Mendez, and L. Esaki, Phys. Rev. B **31**, 6890 (1985); W.I. Wang and F. Stern, J. Vac. Sci. Technol. B **3**, 1280 (1985); D. Arnold, A. Ketterson, T. Henderson, J. Klem, and H. Morkoc, J. Appl. Phys. **57**, 2880 (1985); A.D. Katnani and R.S. Bauer, Phys. Rev. B **33**, 1106 (1986); J. Batey and S.L. Wright, J. Appl. Phys. **59**, 1200 (1986); J.R. Waldrop, R.W. Grant, and E.A. Kraut, J. Vac. Sci. Technol. B **5**, 1209 (1987); G. Danan, B. Etienne, F. Mollot, R. Planal, A.M. Jean-Luois, F. Alexandre, B. Jusserand,

G. Le Roux, J.Y. Marzin, H. Savary, and B. Sermage, Phys. Rev. B **35**, 6207 (1987).

[22] P.E. Brunemeier, D.G. Deppe, and N. Holonyak, Jr., Appl. Phys. Lett. **46**, 755 (1985); H. Temkin, M.B. Panish, P.M. Petroff, R.A. Hamm, J.M. Vandenberg, and S. Sumski, Appl. Phys. Lett. **47**, 394 (1985); M.S. Skolnick, P.R. Tapster, S.J. Bass, A.D. Pitt, N. Apsley and S.P. Aldred, Semic. Sci. Technol. **1**, 29 (1986); R. Sauer, T.D. Harris, and W.T. Tsang, Phys. Rev. B **34**, 9023 (1986); D.J. Westland, A.M. Fox, A.C. Maciel, and J.F. Ryan, Appl. Phys. Lett. **50**, 839 (1987).

[23] S.R. Forrest, P.H. Schmidt, R.B. Wilson, and M.L. Kaplan, Appl. Phys. Lett. **45**, 1199 (1984); K. Steiner, R. Schmitt, R. Zuleeg, L.M.F. Kaufmann, K. Heime, E. Kuphal, and J. Wolter, Surf. Sci. **174**, 331 (1986).

[24] J.R. Waldrop, R.W. Grant, E.A. Kraut, Appl. Phys. Lett. **54**, 1878 (1989).

[25] D. V. Lang, M. B. Panish, F. Capasso, J. Allam, R.A. Hamm, A.M. Sergent and W.T. Tsang, J. Sci. Vac. Technol. B **5**, 1215 (1987); Appl. Phys. Lett. **50**, 736 (1987).

[26] R.W. Grant, J.R. Waldrop, and E.A. Kraut, Phys. Rev. Lett. **40**, 656 (1978); J.R. Waldrop, E.A. Kraut, S.P. Kowalczyk, and R.W. Grant, Surf. Sci. **132**, 513 (1983).

[27] J.R. Waldrop, S.P. Kowalczyk, R.W. Grant, E.A. Kraut and D.L. Miller, J. Vac. Sci. Technol. **19**, 573 (1981).

[28] A.D. Katnani and R.S. Bauer, Phys. Rev. B **33**, 1106 (1986).

[29] J.R. Waldrop and R.W. Grant, Phys. Rev. Lett. **26**, 1686 (1978).

[30] J.R. Waldrop, R.W. Grant, S.P. Kowalczyk, and E.A. Kraut, J. Vac. Sci. Technol. A **3**, 835 (1985).

[31] J. Batey and S.L. Wright, Surf. Sci. **174**, 320 (1986).

[32] F. Capasso, A.Y. Cho, K. Mohammed and P.W. Foy, Appl. Phys. Lett. **46**, 664 (1985); F. Capasso, K. Mohammed and A.Y. Cho, J. Vac. Sci. Technol. B **3**, 1245 (1985).

[33] D.W. Niles, G. Margaritondo, P. Perfetti, C. Quaresima and M. Capozi, Appl. Phys. Lett. **47**, 1092 (1985); D.W. Niles, G. Margaritondo, E. Colavita, P. Perfetti, C. Quaresima and M. Capozi, J. Vac. Sci. Technol. A **4**, 962 (1986).

[34] P. Perfetti, C. Quaresima, C. Coluzza, C. Fortunato and G. Margaritondo, Phys. Rev. Lett. **57**, 2065 (1985).

[35] J.T. McKinley, Y. Hwu, D. Rioux, A. Terrasi, F. Zanini, and G. Margaritondo, private communication.

[36] L. Kleinman, Phys. Rev. B **24**, 7412 (1981).

[37] C.G. Van de Walle and R.M. Martin, J. Vac. Sci. Technol. B **4**, 1056 (1986), and Phys. Rev. B **35**, 8154 (1987).

[38] A. Baldereschi, S. Baroni, and R. Resta, Phys. Rev. Lett. **61**, 734 (1988).

[39] J.D. Jackson, *Classical Electrodymamics* (Wiley, New York, 1975).
[40] F.N.H. Robinson, *Macroscopic Electromagnetism* (Pergamon, Oxford, 1973).
[41] L.R. Anderson, Solid-State Electron. **5**, 341 (1962).
[42] J.L. Freeouf and J.M. Woodall, Surf. Sci. **168**, 518 (1986); Appl. Phys. Lett. **39**, 727 (1981).
[43] W.A. Harrison, J. Vac. Sci. Technol. **14** 1016 (1977).
[44] W.R. Frensley and H. Kroemer, Phys. Rev. B **16**, 2642 (1977).
[45] J. Tersoff, Phys. Rev. B **30**, 4874 (1984); Surf. Sci. **168**, 275 (1986); J. Vac. Sci. Technol. B **3**, 1157 (1985); Phys. Rev. Lett. **56**, 675 (1986).
[46] C. Tejedor and F. Flores, J. Phys. C **11**, L19 (1987); J. Phys. C **12**, 731 (1979); J.C. Duran, F. Flores, C. Tejedor, and A. Muñoz, Phys. Rev. B **36**, 5920 (1987).
[47] W.E. Pickett, S.G. Louie and M.L. Cohen, Phys. Rev. B **17**, 815 (1978); J. Jhm and M.L. Cohen, Phys. Rev. B **20**, 729 (1979); W.E. Pickett and M.L. Cohen, Phys. Rev. B **18**, 939 (1978);
[48] C. Priestyer, G. Allan, and M. Lannoo, Phys. Rev. B **33**, 7386 (1986).
[49] A.D. Katnani and G. Margaritondo, Phys. Rev. B **28**, 1944; G. Margaritondo, Surf. Sci. **168**, 439 (1986).
[50] P. Hoemberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[51] W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965).
[52] See for instance M.L. Cohen in *International School of Physics "Enrico Fermi" (couse 89, Varenna, 1983): Highlights of Condensed-Matter Theory*, edited by F. Bassani, F. Fumi and M.P. Tosi, p.16 (North–Holland, Amsterdam, 1985).
[53] See for instance O.H. Nielsen and R.M. Martin, Phys. Rev. B **32**, 3792 (1985).
[54] J.C. Phillips and L. Kleinman, Phys. Rev. **116**, 287 (1959); F. Bassani and V. Celli, J. Phys. Chem. Solids **20**, 64 (1961); further references can be found in F. Bassani and G. Pastori Parravicini, *Electronic States and Optical Transitions in Solids*, edited by R.A. Ballinger (Pergamon Press, Oxford, 1975).
[55] M.L. Cohen and T.K. Bergstresser, Phys. Rev **141**, 789 (1966); V. Heine and I. Abarenkov, Phil. Mag. **9**, 451 (1964); A.O.E. Animalu, Phil. Mag. **11**, 379 (1965); I. Abarenkov and V. Heine, Phil. Mag. **12**, 529 (1965).
[56] D.R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[57] G.B. Bachelet, D.R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).
[58] L. Kleinman and J.C. Phillips, Phys. Rev. **116**, 880 (1959);
[59] A. Baldereschi, Phys. Rev. B **12**, 5212 (1973).
[60] D.J. Chadi and M.L. Cohen, Phys. Rev. B **8**, 4547 (1973).
[61] H.J. Monkhorst and J.P. Pack, Phys. Rev. B **13**, 5188 (1976).
[62] S. Froyen, Phys. Rev. B **39**, 3168 (1989).
[63] Shang Yuan Ren and J.D. Dow, Phys. Rev. B **38**, 1999 (1988).
[64] M. Cardona and N.E. Christensen, Phys. Rev. B **35**, 6182 (1987); N.E. Christensen, Phys. Rev. B **37**, 4528 (1988); *ibid.* B **38**, 12687 (1988).

[65] Su–Huai Wei and A. Zunger, Phys. Rev. Lett. **59**, 144 (1987).

[66] S. Massidda, B.I. Min, and A.J. Freeman, in *Superlattice and Microstructures* **4**, No. 1, 15 (1988); Phys. Rev. B **38**, 1291 (1988).

[67] W.R.L. Lambrecht and B. Segall, Phys. Rev. Lett. **61**, 1764 (1988).

[68] D.M. Bylander and L. Kleinman, Phys. Rev. B **34**, 5280 (1986); *ibid.* **36**, 3229 (1987); *ibid.* **39**, 5116 (1989); Phys. Rev. Lett. **59**, 2091 (1987).

[69] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980); J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[70] J.C. Mikkelsen and J.B. Boyce, Phys. Rev. B **28**, 7130 (1983).

[71] S. Baroni, R. Resta, A. Baldereschi, and M. Peressi, in *Proc. of the NATO Adv. Res. Workshop on SPECTROSCOPY OF SEMICONDUCTOR MICROSTRUCTURES*, edited by G. Fasol, A. Fasolino, and P. Lugli (Plenum, New York, 1989).

[72] *GaInAsP Alloy Semiconductors*, edited by T.P. Persall (Wiley, New York, 1982).

[73] C.G. Van de Walle, Phys. Rev. B **39**, 1871 (1989); C.G. Van de Walle and R.M. Martin, Phys. Rev. Lett. **62**, 2028 (1989).

[74] M. Peressi, S. Baroni, A. Baldereschi, and R. Resta, to be submitted to Phys. Rev. B.

[75] O. Berolo, J.C. Wolley, *Proceedings of the 11th International Conference on THE PHYSICS OF SEMICONDUCTORS* (Pol. Sci. Publ., Warsaw, 1972), p. 1420; P.M. Laufer, F.H. Pollak, R.E. Nahory, and M.A. Pollack, Sol. State Commun. **36**, 419 (1980).

[76] S. Baroni, R. Resta, and A. Baldereschi, in *Proceedings of the 19th International Conference on THE PHYSICS OF SEMICONDUCTORS*, edited by W. Zawadazki (Institute of Physics, Polish Academy of Sciences, Wroclaw, 1988), p. 525.

[77] R.M. Martin, *Proceedings of the Conference on Physics of Compound Semiconductor Interfaces, Estas Park, Colorado, 1980,* J. Vac. Sci. Technol. **17**, 978 (1980).

[78] S. B. Zhang, D. Tománek, S. G. Louie, M. L. Cohen and M. S. Hybertsen, Sol. State Commun. **66**, 585 (1988).

[79] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **34**, 5390 (1986); *ibid.* **35**, 5585 (1987); *ibid.* p. 5602.

[80] R. W. Godby, M. Schlüter and L. J. Sham, Phys. Rev. B **36**, 6497 (1987); *ibid.* **37**, 10159 (1988).

[81] *Numerical Data and Functional Relationships in Science and Technology*, Group III, **17** a–b, edited by Landolt–Börnstein (Springer, New York, 1982); in particular the quoted references for $Ga_{0.5}In_{0.5}As$ and InP are Hass, Henvis, J. Phys. Chem. Sol. **23**, 1099 (1962), and Thomas, Woolley, Can. J. Phys. **49**, 2025 (1971).

[82] T. Fukui and H. Saito, Japan. J. appl. Phys. **23**, L521 (1984); Inst. Phys. Conf. Ser. **79**, 397 (1986).

[83] D. Gershoni, H. Temkin, M.B. Panish, and R.A. Hamm, Phys. Rev. B **39**, 5531 (1989).

[84] H. Nakayama and H. Fujita, in *Gallium Arsenide and Related Compounds — 1985*, edited by M. Fujimoto, IOP Conference Proceedings No. 79 (Hilger, Boston, 1986), p.289; T.S. Kuan, W.I. Wang, and L. Wilkie, Appl. Phys. Lett. **51**, 51 (1987); M.A. Shashid, S. Mahajan, and D.E. Laughlin, Phys. Rev. Lett. **48**, 2567 (1987).

[85] P. Bogusławski and A. Baldereschi, Phys. Rev. B **39**, 8055 (1989).

[86] R. Padjen and D. Paquet, private communication.

[87] *Numerical Data and Functional Relationships in Science and Technology*, Group III, **17** a–b, edited by Landolt–Börnstein (Springer, New York, 1982); the effective charges for XAs are considered as the average of the effective charges for InAs and GaAs.

[88] J.M. Vandenberg, M.B. Panish, H. Temkin, and R.A. Hamm, Appl. Phys. Lett. **53**, 1920 (1988).

[89] S. Baroni, P. Giannozzi, and A. Testa, Phys. Rev. Lett. **58**, 1861 (1987).

[90] R. Resta and A. Baldereschi, Phys. Rev. B **23**, 6615 (1981).

[91] R. Resta and S. Baroni, Bull. Am. Phys. Soc. **34**(3), 832 (1989).

[92] D. Gershoni, H. Temkin, J.M. Vandenberg, S.N.G. Chu, R.A. Hamm, and M.B. Panish, Phys. Rev. Lett. **60**, 448 (1988);