Chapter II
THEORETICAL CONSIDERATIONS

2.1. Wafer bonding

2.1.1. Introduction

When two clean and smooth solid surfaces are joined together at room temperature (RT) so that the distance between them is of the order of the interatomic spacing, they adhere spontaneously to each other without external forces or glue. Once the bonding is initiated, it propagates by itself across the whole interface. This phenomenon, although not necessarily involving wafer-shaped materials, is often referred to as wafer bonding, direct wafer bonding, fusion bonding or surface activated bonding [1].

The attraction originates either in weak interactions (van der Waals forces, hydrogen bonds) or strong ones (ionic, covalent or metallic bonds) depending on the chemistry of the surfaces involved and on the ambient conditions. The bonding achieved by weak interactions is reversible and necessitates subsequent thermal treatment to increase the bonding energy.

Wafer bonding can be implemented to virtually any material, provided that the requirements regarding surface cleanliness, smoothness, and flatness are met. Introduced first to semiconductor industry during the 80's for silicon-silicon bonding [2], [3], wafer bonding has received increasing attention ever since.

The following discussion focuses on theoretical and practical aspects of semiconductor wafer bonding.

2.1.2. Surface preparation

Wafer bonding is very sensitive to cleanliness and smoothness, because the initial interactions that mediate the attraction are weak and short-ranged in nature. For this reason, the first step is always to ensure that the surfaces meet the mechanical and chemical requirements.

When exposed to air, silicon readily develops a thin SiO_2 layer (typically 2-5 nm thick), which passivates the unsatisfied dangling bonds (DB) present in a large number on the surface. Thus, as-delivered silicon wafers are always covered with a native oxide on top of which typically undesirable contaminants are present, reflecting the processing history of the sample: water, ionic compounds, polar and non-polar organics from processing liquids or from the air. The latter render the wafer surface hydrophobic (water repellant) and should be first removed during the cleaning process, since the cleaning mixtures are polar and would be repelled from the surface as well. It is recommended to perform the cleaning in a controlled dust-free atmosphere, preferably in a

cleanroom. Additionally, ultrapure DI (deionised) water and semiconductor grade chemicals are to be used in order to avoid contamination during the chemical treatment itself. The chemicals used in the semiconductor industry are able to remove dust particles and contaminants without degrading the surface quality of the wafers too much.

First, the RCA1 solution (also termed ‘basic piranha’) is used for the removal of organics. This solution consists of a mixture of NH_4OH:H_2O_2:H_2O=1:1:5 and becomes effective when heated above 60°C. To increase its efficiency, it is usually combined with immersion in an ultrasonic bath. If gross contamination is suspected, an ‘acid piranha’ (H_2SO_4:H_2O_2=3:1) cleaning step can be performed before RCA1 in order to support further removal of organic contaminants [2].

In a second step, the RCA2 solution (HCl:H_2O_2:H_2O=1:1:6) is used to remove most metallic contamination, except for Au and Pt [10]. Again, the solution must be heated above 60°C in order to initiate the reaction. Between these steps the wafers are thoroughly rinsed with DI water. At this stage, the wafers are contamination-free and protected by a thin layer of SiO_2.

Depending on the desired properties of the bonded structures (that is, yielding electrically insulating or conducting interfaces), the bonding can be done with SiO_2 at the interface or after oxide removal.

2.1.3. Hydrophilic bonding

The SiO_2 layer that covers the wafer surface after cleaning is terminated by OH groups, forming the so-called silanol groups (Si-OH) which render the silicon surface hydrophilic [11]. As a consequence, water molecules will be chemisorbed on the surface via hydrogen bonds. The water excess is removed by spin-drying in a turbulence-free microcleanroom [12] or by pure nitrogen blowing [13]. After this step the wafers are joined together and small pressure is applied in order to initiate the bonding, which is caused by hydrogen bonds between remaining water molecules (Fig. 2.1). Not only these water molecules mediate the initial bonding, but they fill in the gap between non-perfectly mating surfaces, bridging distances as large as 1 nm [14].

The fracture surface energy (bonding energy) between the wafers is in the range of 100-150 mJ/m^2 directly after RT bonding, and increases above 200 mJ/m^2 after long-term storage in air [15], partly due to additional silanol groups formation (as a result of water reaction with the surfaces), and partly due to the slow diffusion of water molecules out of the interface. Slightly above RT, a small fraction of silanol groups will be close enough to bond directly with each other. Further, silanol groups condensate leading to covalent Si-O-Si bonds formation (siloxane) [16], thus increasing the overall bonding energy:

$$\text{Si-OH} + \text{OH-Si} \rightarrow \text{Si-O-Si} + H_2O \qquad (2.1)$$

This is a very slow process at RT, and leads to the formation of additional water molecules. Moreover, (2.1) is reversible, which means that the interfacial water causes further silanol group formation, weakening the bonding energy. For this reason, the bonding energy tends to saturate at relatively low values. In

order to accelerate the process and to increase the bonding strength, thermal annealing needs to be performed.

Below 110°C the bonding energy is similar to RT. Between 110°C and 200°C more and more hydrogen bonds between adjacent silanol groups are converted into covalent siloxane bonds depicted by the red lines in Fig. 2.1, and the interface energy increases rapidly up to 1200 mJ/m^2.

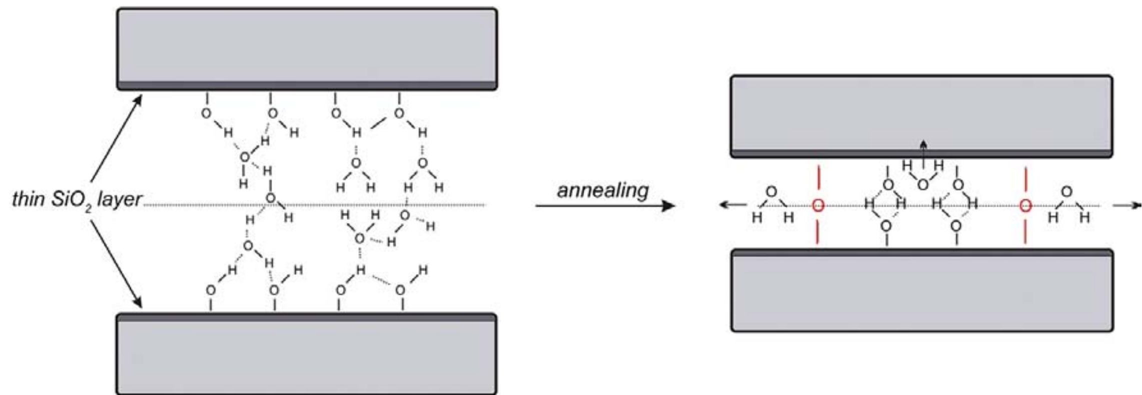

Figure 2.1: Interface morphology upon RT hydrophilic bonding (left) and after annealing (right).

Further temperature increase up to 700-800°C yields no change in the bonding energy [17], due to the saturation behaviour of the above described mechanism. That is, no further bonding energy increase is possible after all available silanol bonds are converted to siloxane bonds, the limitation being given by the actual contact area between mating surfaces. Since these surfaces are not atomically flat, microgaps are expected to form at the unbonded areas.

At 800°C, the SiO_2 becomes viscous and starts filling the microgaps. As a result, the bonding energy increases to 2000-2500 mJ/m^2 [18]. If the annealing temperature exceeds 1000°C, the native oxide layer at the interface disintegrates into islands leading to Si-Si covalent bonding and the interface loses its insulating properties, provided the SiO_2 layer is reasonably thin and the oxygen concentration in bulk silicon is low [2].

2.1.4. Hydrophobic bonding

Hydrophobic (i.e. ‘water repellent’) surfaces result when the native oxide is etched away using hydrogen fluoride [19] or buffered ammonium fluoride [20] solutions, for Si(100) and Si(111) surfaces, respectively. The dissolution of SiO_2 by HF is depicted in the following reaction:

$$SiO_2 + 4HF \rightarrow SiF_4 + 2H_2O \qquad (2.2)$$

After oxide removal, the dangling bonds of the surface atoms are passivated by hydrogen and only to a small extent by fluorine atoms. This occurs because of the strong polar character of the Si-F bonds which in turn causes bond polarization of the Si-Si back-bonds. HF molecules, having a highly polar character as well, attack these back-bonds resulting in the release of stable SiF_x species [21]. The Si-H bonds are quite stable in air (up to a few

hours). However, hydrophobic surfaces are very prone to hydrocarbon contamination; therefore bonding should be performed very quickly without water rinse which would cause partial re-hydrophilisation due to the conversion of Si-F bonds to Si-OH bonds.

Once the wafers are joined, the adhesion is caused by van der Waals forces between hydrogen atoms, as depicted in Fig. 2.2. The initial fracture strength in the case of hydrophobic bonding takes values between 20 and 30 mJ/m^2 (lower than those corresponding to hydrophilic bonding), and remains low up to 300-400°C annealing temperature when the hydrogen and HF molecules start to dissociate from the silicon surface and to diffuse along the interface, allowing covalent bond formation depicted by the red lines in Fig. 2.2 [1]. Thus, energies as high as 2000 mJ/m^2 can be reached after annealing at 700°C.

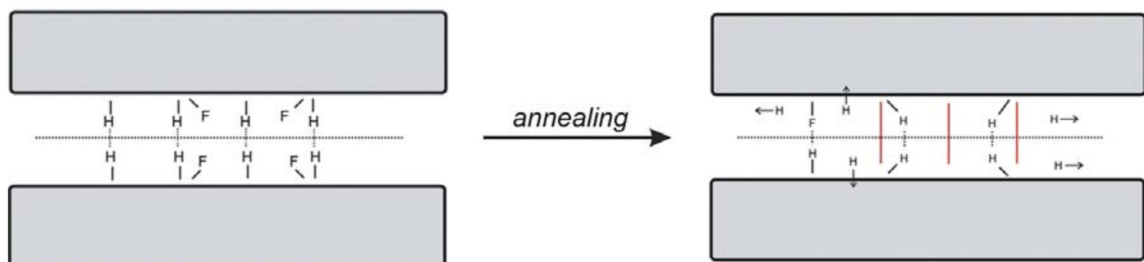

Figure 2.2: Interface morphology upon RT hydrophobic bonding (left) and after annealing (right).

As in the case of hydrophilic bonding, the bonding energy is limited by the actual contact area between mating surfaces. At temperatures higher than 700°C, silicon atoms start to flow along the interface filling the microcavities originating in the initial wafer roughness. A number of workers identified bubbles and voids at the interface [22], [23], formed by impurity agglomerates and gas trapping (hydrogen, fluorine, and nitrogen).

2.1.5. UHV bonding

Both hydrophilic and hydrophobic bonding exhibit low initial fracture energy and high temperature annealing steps are required to obtain strong bonding, which might not be sensible in many situations, especially when dealing with dissimilar materials. For this reason, new ways of achieving strong bonding upon joining at RT have been thoroughly investigated.

One possible solution is ultrahigh-vacuum (UHV) bonding [4], which can be regarded as a modified hydrophobic bonding performed in an ultrahigh-vacuum environment (pressure less than 10^{-9} mbar). The procedure of producing hydrophobic surfaces has already been outlined, so we can easily imagine a bonding experiment in which two such surfaces are joined together.

Without subsequent annealing, the bonding energy is low and the wafers can be separated at any time, e. g. in UHV. Assuming a flat Si(100) surface, the two DBs on each surface atom are hydrogen terminated, leading to a Si(100)-(1x1):H structure (neglecting the small amount of fluorine). Upon annealing to

about 500°C the hydrogen is desorbed from the separated surfaces in UHV, leaving two DB per atom, i. e. about 10^{15} DB/cm^2.

The energy of the free surfaces is very high in this case (ca. 2.25 J/m^2, ref. [24]), and a (2x1) reconstruction is adopted via formation of Si-Si dimers which give a more stable configuration with only one DB per surface atom. For Si(111) surfaces, a (7x7) reconstruction is favoured [25]. The energy of a reconstructed Si-Si bond is still higher than that of an undisturbed bond because of the strong bending. For sufficiently small separation between surfaces, the electronic surface states overlap and an ‘adhesive avalanche’ occurs via reconstruction breaking and covalent bonding as predicted by molecular dynamics simulations for Si(111)-(1x1) [26] and Si(100)-(2x1) [27]. The schematic drawing in Fig. 2.3 shows the sequence of heating and cooling before room temperature bonding in UHV. The red lines represent direct Si-Si bonds.

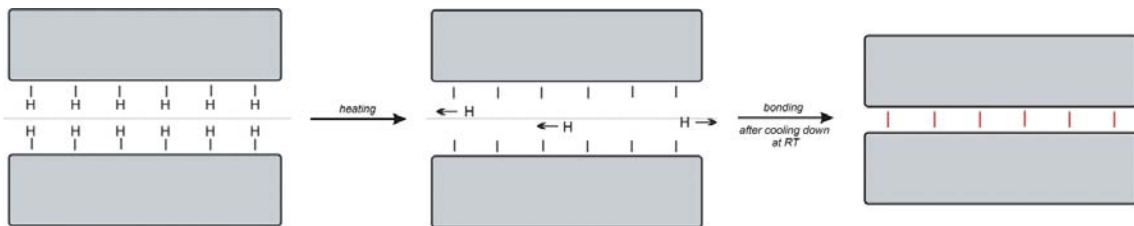

Figure 2.3: UHV bonding at RT after hydrogen desorption.

The energy released during the bonding process is spent to re-arrange surface atoms. Even if the two wafers are bonded in such a manner that their crystallographic orientations are perfectly aligned, the interface doesn’t resemble a perfect structure due to incomplete bonding arising from atomic steps and fast energy dissipation. In real cases there is always a tilt (ψ) related to the surface miscut and a twist (θ) between the bonded wafers, which makes the re-arrangement less perfect, resembling a grain boundary at the bonded interface. The atoms at the transition between the two crystals have different arrangements than in the bulk material, with important consequences on the electrical properties. In either case, the bonding energy is above 2 J/m^2 directly after RT bonding [8]. No further annealing treatment is required. However, at high temperatures one can expect a re-arrangement of the interface atoms into more relaxed configurations.

In summary, not only does UHV bonding ensure high fracture strength without annealing, but the voids observed after hydrophobic bonding caused by trapped gas atoms and surface steps can be avoided to a large degree. Based on the high driving force associated with UHV bonding, local imperfections can be bonded by mechanical distortion. The approach was extended to crystals with different surface orientation and also to dissimilar materials [28], [29] becoming a very attractive choice for materials integration.

2.2. Layer transfer by ion implantation and wafer bonding

The transfer of single crystalline layers onto various substrates plays a major role for materials integration. Hydrophilic wafer bonding followed by implantation

induced splitting is already established for producing silicon-on-insulator (SOI) wafers and offers many advantages over traditional methods such as polishing/etching techniques. This process has been termed “Smart-Cut” by Bruel [7] who also patented it. The patent was later on licensed exclusively to the commercial silicon-on-insulator supplier SOITEC in France. Its versatility has been demonstrated by extending the approach to a variety of other material combinations, such as silicon on poly-SiC [30], GaAs [31] and InP [32]. In the following, the general features of layer transfer will be discussed.

2.2.1. Basics of implantation

Ion implantation is a process by which energetic ions are introduced into a target in order to change its mechanical and electrical properties. Due to its unique characteristics such as accurate control of depth and dose, it has experienced rapid development for semiconductor industry as an alternative method for doping. Here, the particularities of ion implantation when applied to layer transfer will be outlined.

Upon entering a semiconductor material, the energetic ions lose energy by two stopping mechanisms [33]. The first is by energy transfer to the target nuclei, which causes deflection of the former and dislodging of the latter from their original sites. This process defines a *nuclear stopping* which results in structural damages of the semiconductor, such as point and line defects.

A second mechanism is by the inelastic collision of the ion with both bound and free electrons in the target. Energy is lost by means of transient generation of electron-hole pairs and ionization of the target atoms without inducing displacement. This process gives the *electronic stopping*. The energy loss per unit distance (dE/dx) is related to the corresponding stopping power S via the following equation:

$$S_{e,n} = -\frac{1}{N}\left(\frac{dE}{dx}\right)_{e,n} \tag{2.3}$$

where N is the density of target atoms in the semiconductor and the subscripts “e”, “n” denote electronic and nuclear stopping, respectively.

In order to establish which mechanism prevails a number of physical parameters have to be considered, including the masses and the atomic numbers of both the incident ion and the target material, the energy E of the incident ion, the density of the target, the crystal orientation with respect to the beam, and the lattice structure of the target. For moderate energies (50-500 keV) implants of H and He in Si and GaAs, most of the energy is lost through electronic stopping given by:

$$S_e = kE^{1/2} \tag{2.4}$$

where k is a function of ion-target combination [34].

Once S_n and S_e are known, one can calculate the range R given by the distance travelled by the ion before coming to rest. It is actually the projection of

this range along the direction of the incident ion (projected range) that bears more significance in semiconductor technology, together with its standard deviation along the same direction (straggle). The implantation profile is usually non-Gaussian and higher-order statistics are required to describe its shape [34].

If the incident beam is aligned to a major crystallographic axis the ions travel a considerable distance through the lattice with minor energy loss, by a process known as channelling. This process is reflected in an increased penetration depth without considerable lattice disorder [35], but the range depends critically on the degree of alignment and the ion dose. Such an effect is avoided by tilting the wafer by an angle of about 7-10° away from any low index direction.

2.2.2. Hydrogen in silicon and gallium arsenide

The role of hydrogen in semiconductors, particularly in silicon has been investigated over the last two decades due to its ability to passivate the electrical activity of defects [36].

Upon entering a semiconductor hydrogen generates displacement damage in the form of interstitials, vacancies and complexes. Given a sufficiently high dose (>1.8×10^{16} cm^{-2}), these point defects agglomerate along [100] and [111] planes, generating free inner surfaces, i.e. finite areas with a high number of broken bonds called platelets [37]. Already present in high concentrations, H is quickly trapped at these surfaces passivating the broken bonds [38]. It has been estimated that only about 1/8 of the implanted H is contained in the platelets, while most of H atoms are trapped by other defects like vacancy agglomerates and dislocation loops [37].

During annealing the trapped H dissociates from complexes and diffuses into the platelets forming H_2 molecules. Thus, the platelets expand into micro-cracks while the density of interstitials, vacancies and complexes decreases [39]. The pressure inside the micro-cracks increases to such values that expansion in lateral direction occurs via an Ostwald ripening process [40]. As soon as the micro-cracks join together in the same plane, material exfoliation occurs locally.

In order to understand the evolution of defects upon annealing, it is required to know the preferred lattice sites for hydrogen and possible migration paths in silicon. Van de Walle et al. [41] determined that the bond-centered site is favourable for the neutral and positively charged hydrogen, while the tetrahedral interstitial site is preferred by the negatively charged hydrogen in n-type silicon. In addition, Chang et al. [42] postulated the existence of two diatomic complexes, H_2 and H_2^* respectively. H_2 is formed by two hydrogen atoms occupying two adjacent tetrahedral interstitial sites oriented in the [100] direction and H_2^* contains one center-bonded hydrogen and one interstitial hydrogen. It is generally accepted that hydrogen exists in silicon in the form of H^+ (center-bonded), H^- (tetrahedral interstitial), H_2, H_2^*, B-H (boron doped silicon) and P-H (phosphorus doped silicon).

In GaAs, hydrogen produces very little damage up to a dose of 10^{16} cm^{-2} [43]. Few studies have been performed in order to understand the nature of defects in GaAs and their evolution during annealing [34], [44]. It has been

shown that the implantation temperature must fall within an interval specific to each selection of energy and dose in order to induce blistering [26]. Radu [45] found that for doses ranging from 3 to 5x10^{16} cm^{-2} and energies of 130 keV blistering occurs directly during implantation if the temperature exceeds 100°C.

Neethling et al. [46] performed TEM investigations and proposed a theoretical model which predicts that for low doses (5x10^{15} cm^{-2}) only 1.9% of the hydrogen is trapped in the platelets developing a pressure of about 1 GPa, which corresponds to a density of 4x10^{22} cm^{-3}, i.e. 1-2 H atoms per vacancy.

2.2.3. Helium in silicon and gallium arsenide

Helium in semiconductors was studied before the development of the Smart-Cut process. Gas bubbles and micro-cracks were observed upon implantation starting in early 1980's [47], [48]. Having a lower solubility than H in silicon, He forms vacancy-related complexes and bubbles, depending on implantation parameters [49]. During annealing the bubbles grow and He diffuses out leading to void formation, provided that the dose is larger than 1x10^{16} cm^{-2}. For doses lower than 5x10^{15} cm^{-2} He vacancy clusters are formed but they dissociate below 250°C preventing bubble formation [50]. At intermediate doses, some He atoms remain trapped after annealing forming a low density of voids, but the density is not high enough to cause blistering/exfoliation.

The mechanism of exfoliation in He implanted GaAs is less clear, partly because it received less attention compared to H implantation. However, data about defect formation upon implantation exist for both (100) [51] and (111) oriented GaAs [52]. In case of (100) GaAs, TEM investigations depict a combined dislocation-bubble microstructure localized near the end-region of the damaged layer which seems to be responsible for surface flaking. Implanted and annealed (111) GaAs feature extended He platelets lying in (111) planes which grow until splitting occurs. Large area exfoliation instead of blistering was reported for RT implantation [45], favoured by the high density of small micro-cracks (with diameter much smaller than the depth where they are located) present in the implanted material. On the other hand, blisters were observed in the as-implanted sample when the implantation was performed above 100°C.

2.2.4. Blistering and splitting

So far the effects of ion implantation followed by thermal annealing have been outlined. The microscopic defects created during implantation are essential for micro-crack formation and later blistering/exfoliation.

Experimentally it has been found that for a specific set of implantation parameters and for a specific ion-target combination the surface blistering occurs only with a proper combination of annealing temperature and time. More exactly, for a specific annealing temperature surface blisters form only if the time has reached a critical value, termed blistering time (t_b). Tong et al. [53] found that t_b has an exponential dependence on the annealing temperature:

$$t_b \propto e^{\frac{E_a}{kT}} \quad (2.5)$$

where k is Boltzmann's constant, T is the absolute temperature and E_a is the effective activation energy of the process. Fig. 2.4 features the dependence of the blistering time on the annealing temperature for H_2^+ implanted Si(100) according to [54].

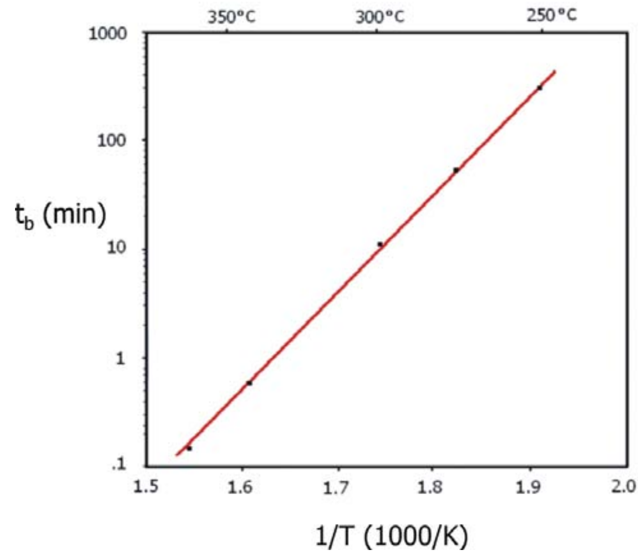

Figure 2.4: Blistering time vs. annealing temperature according to [54] (H_2^+ implantation, 135 keV, $5x10^{16}$ ions/cm^2).

In order to achieve layer transfer instead of blistering, the implanted wafer is stiffened by means of bonding to a handle wafer. Upon annealing, layer transfer occurs via large area splitting, provided that the bonding energy (at the required splitting temperature) in higher than the fracture energy of the implanted layer. Experimentally it was determined that the time required for splitting is about ten times larger than the blistering time [55], the latter being often used for estimate purposes. The layer transfer approach is summarized in Fig. 2.5.

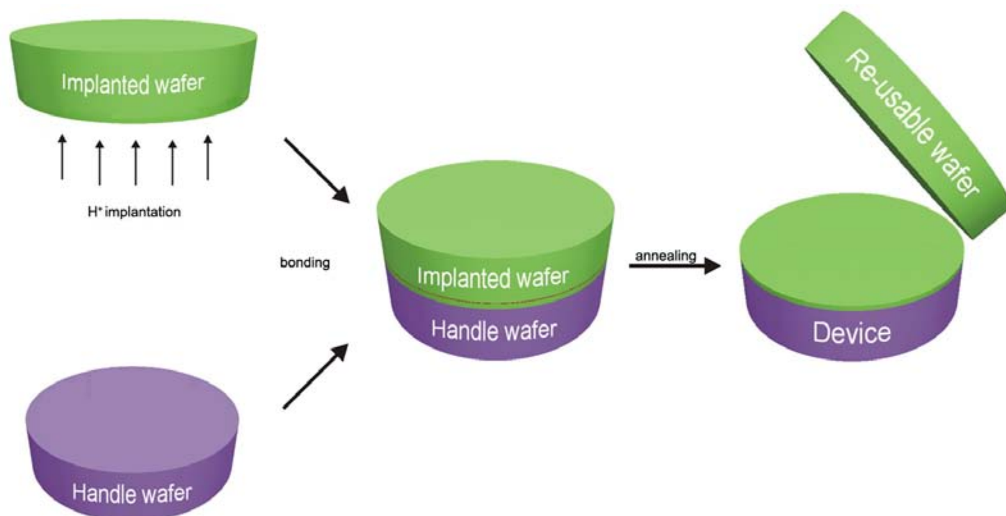

Figure 2.5: Layer transfer by ion implantation, wafer bonding and anneal induced splitting.

2.3. Defects in crystalline semiconductors

Any crystalline imperfection where the atoms do not have a perfectly periodic arrangement on the lattice sites represents a defect. Defects always exist in semiconductor materials. They can be unintentionally introduced at any stage of the process flow from crystal growth to metal contacting. Undesired in most situations, defects can degrade the performance of semiconductor devices in a number of ways. They can be electrically or optically active and might become more important than dopants in determining the material properties.

According to their dimensionality the defects are usually classified in the following way:

- 0-dimensional defects, also called point defects. They include intrinsic point defects (vacancies and self-interstitials) and extrinsic point defects (impurities).
- 1-dimensional defects represented by all kinds of dislocations.
- 2-dimensional defects (planar defects) which include stacking faults, grain boundaries and phase boundaries.
- 3-dimensional defects, namely precipitates (usually involving foreign atoms), voids (agglomeration of vacancies in three-dimensional form) and special cases like stacking fault tetrahedra and dislocation clusters.

In the following, an overview of defects in silicon will be given, with emphasis on those types of defects introduced during wafer bonding.

2.3.1. Point defects

Point defects are localized defects of atomic dimensions occurring in an otherwise perfect crystal lattice. As already mentioned, they include intrinsic point defects (involving only atoms of the host lattice) and extrinsic defects (involving foreign atoms, i. e. impurities).

The simplest imaginable intrinsic point defect is the vacancy, which occurs whenever an atoms leaves its lattice site due to thermal fluctuations or vapour pressure of the species surrounding the material. In the event that the atom migrates to the crystal surface, we have the so-called Schottky defect whose energy of formation is 2.6 eV in the case of silicon. A single vacancy is created by breaking of four covalent bonds whereas two neighbouring vacancies require only six bonds to be broken. Due to this reason, di-vacancies can form as well, with higher probability than pairs of individual vacancies.

The second type of intrinsic point defect is the interstitial (or, more exactly, the self-interstitial), which occurs when an atom becomes located in one of the many interstitial positions within the crystal structure. The energy of formation of this defect in silicon is around 1.1 eV. An atom leaving its regular site has now an additional possibility – that is, ending up in an interstitial position. Thus, a Frenkel defect is created, having an energy of formation of about 9 eV.

In compound semiconductors e.g. GaAs, interstitials and vacancies can be encountered in two ways, involving either gallium or arsenic atoms. Moreover, a gallium atom has the additional possibility to take the position normally occupied by an arsenic atom (and vice-versa), thus generating an antisite defect.

The concentration of intrinsic point defects depends on their energy of formation and on the temperature. Taking into account only the possibility to

form Frenkel defects, their equilibrium concentration at RT is given by [56] as follows:

$$n = Ne^{-\frac{E_f}{2kT}} \tag{2.6}$$

where N is the total number of atoms in a crystal of unit volume (roughly $5x10^{22}$ cm^{-3} for silicon), E_f is the energy of formation of a Frenkel defect (≈9 eV), k is the Boltzmann's constant and T is the absolute temperature. The factor 2 in the denominator of the exponent appears because the formation of a Frenkel defect requires an interstitial and a vacancy to form independently.

The extrinsic point defects (also called chemical point defects) are represented by impurities, either deliberately introduced (for doping purposes or in the case of fast switching devices) or inadvertently as contamination during processing. They can occupy either lattice sites (substitutional impurities) or interstitial sites (interstitial impurities). In many cases, the former are electrically active while the latter are inactive. Among impurities which are electrically active, donors and acceptors which have energy levels within 3kT of the conduction and valence band edges, respectively, are referred to as shallow impurities, while those with energy levels outside this range are referred to as deep, i. e. levels lying in the interval (E_c-0.1 eV)… (E_v+0.1 eV). They have the special (and most of the time undesired) property of reducing the minority carrier lifetime through the generation-recombination (G-R) mechanism. For this reason they are also called G-R centers, trapping centers or simply traps.

Shallow impurities are well known for their doping properties. Using the effective mass theory their ionization energies are found to be 0.04-0.07 eV away from the corresponding band edges. Their ionization probability depends on the concentration and on the temperature; at RT they are fully ionized for low and moderate concentrations. They are not within the scope of this work, therefore the term "defects" will only refer to deep levels henceforth.

The effective mass theory does not apply in the case of deep impurities because their electronic wave functions are more highly localized than for shallow impurities [57]. In order to illustrate the general character of deep levels in semiconductors, consider a monovalent impurity introduced into a substitutional site. Such an atom, in the neutral state, has only one attached electron which provides covalent bonding with its neighbouring lattice atoms. As soon as electrons are attached to it, it is successively transferred to a more and more negatively charged state, giving rise to as many as three acceptor levels. Additionally, the atom can lose an electron and be promoted to a positive (donor) charge state. However, in most cases only a few of these levels are identifiable due to the fact that the addition/extraction of electrons leads to such a decrease/increase of energy until the corresponding energy level eventually moves beyond the bandgap.

The concentration of these impurities depends on their availability and their solid solubility in the semiconductor material. The latter depends upon the position of the Fermi level in the semiconductor, i. e. the background doping level and type.

A special category of point defects is represented by *complexes*, which can result from the following interactions between point defects:

- intrinsic-intrinsic
- extrinsic-extrinsic
- intrinsic-extrinsic

Some of the most commonly observed complexes are those that involve association between oppositely charged donors and acceptors. When the donor and the acceptor are both singly charged, the quasi-chemical equation that describes the complex formation would be:

$$D^{+} + A^{-} = (DA)^{0} \tag{2.7}$$

This neutral complex can then be ionized by donating an electron to the conduction band or accepting a hole from the valence band. The ionization energy would be:

$$E_i = E_g + \frac{q^2}{4\pi\varepsilon}\left(\frac{1}{r_{da}} - \frac{1}{2r_d} - \frac{1}{2r_a}\right) \tag{2.8}$$

where r_d and r_a are the orbital radii of the electron in the ground state of the isolated donor and the hole in the ground state of the isolated acceptor, respectively and r_{da} is the donor-acceptor separation [58]. ε represents the absolute permittivity and q is the elementary charge. When

$$r_{da} < \frac{2r_d r_a}{r_d + r_a} \tag{2.9}$$

the ionization energy of the pair is greater than the energy gap. That is, the donor and acceptor energy levels are removed from the energy gap as the two ions are brought together. Since the Bohr radii of electrons and holes around isolated hydrogenic donors and acceptors are typically large, the formation of a complex will generally remove states from the forbidden band. There are, however, situations when a complex might be electrically active such as:

- when one of the ions is doubly charged. In this case the deeper second ionization state can remain in the forbidden gap.
- when the ionization energy of one or both of the ions is large. In this case the orbital radius of the electron or hole will be small and it might be difficult for the complex to fulfil the inequality (2.9).

In summary, both intrinsic and extrinsic point defect concentrations can be maintained in principle low enough to ensure a negligible influence on device characteristics. The use of prime quality wafers combined with semiconductor grade chemicals and a cleanroom environment would suffice, since today's silicon can be grown with metallic impurities concentrations of less than 10^{10} cm^{-3}. However, subsequent processing steps such as ion implantation, high temperature annealing, and metal contacting tend to introduce higher concentrations.

In wafer bonding, point defects might become important because they can be easily incorporated at the bonded interface. Moreover, point defects might as well play a role when interacting with dislocations.

2.3.2. Dislocations

Dislocations are one-dimensional defects which can extend throughout large regions of the crystal generating a geometric fault of the lattice. It occurs when the crystal is subjected to stresses in excess of the elastic limit. They are characterized by a core region with a diameter in the nm range, where the crystal symmetry is broken and a far-field region where an elastic field changes slowly in space. If the dislocation has dangling bonds, it can be viewed as a linear array of acceptors that scatter mobile carriers. The dangling bonds can also attract impurities and interstitials or act as an internal source and sink for vacancies. Due to the many-electron nature of the dislocations, they might have a quasi-continuous spectrum of electronic states in the bandgap [59].

Since the dislocations are one-dimensional defects, the lattice is disturbed (only) along the dislocation line. Thus, a first element to describe the dislocation at any point is the line vector $\vec{l}$. In order to describe the state of slip that generates a dislocation or has been generated by a dislocation, a second element, called Burgers vector $\vec{b}$ is introduced. It specifies the direction and distance by which atoms in one plane have moved with respect to atoms in another plane. Alternatively, the Burgers vector is defined as a closure failure arising in a cycle in the perfect material surrounding the dislocation. The slip (glide) plane is a plane along which the dislocation can move under applied stresses, and it is determined by the line and Burgers vectors.

Although the nature of dislocations is quite complex, they are usually composed of combinations of two basic types, namely the screw dislocation and the edge dislocation which will be briefly discussed in the following. A simple cubic lattice will be considered. For a detailed analysis the reader is referred to textbooks [60], [61].

Fig. 2.6 depicts the schematic representation of a screw and an edge dislocation.

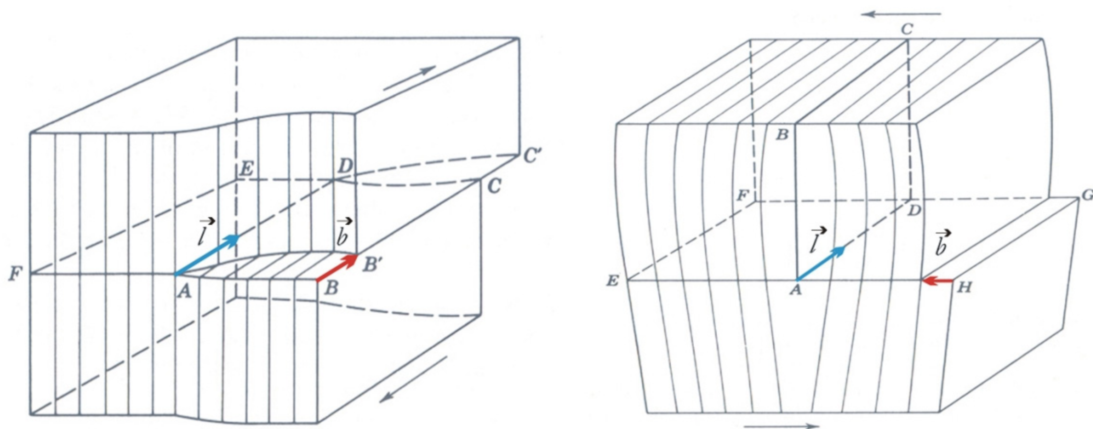

Figure 2.6: Schematic representation of a screw (left) and edge (right) dislocation.

In the first case, consider an imaginary cut in the crystal along the plane ABCD that ends on the line AD. Applying sufficiently high shear stresses on the two halves of the crystal separated by the plane ABCD, will cause them to shift by one atomic spacing. Thus, a screw dislocation is created along the line AD which marks the boundary in the plane ECBF which divides the perfect crystal from the imperfect. The line and Burgers vectors are parallel, so there is no well defined slip plane, i. e. the screw dislocation can move, in principle, in any plane. The screw dislocation doesn’t have dangling bonds but perturbs the lattice by a spiral rearrangement of bonds around its axis. Consequently, it only shifts the band edges locally without introducing deep levels in the band gap of silicon. Using elasticity theory one can get an approximate formula for the strain energy associated with a screw dislocation:

$$E_{screw} = \frac{Gb^2 l}{4\pi} \ln\left(\frac{R}{r}\right) \tag{2.10}$$

where G is the shear modulus, b is the magnitude of the Burgers vector, l represents the dislocation length, r is the core radius and R is an undetermined external radius corresponding to the distance at which strain fields of the dislocations cancel each other (typically 10^5 atom spacings). With a value of the ratio R/r of about 10^4, this energy is roughly estimated to 10-19 eV/atom length, quite high compared to the energy of formation of a point defect.

Fig. 2.6b shows how an edge dislocation can be constructed by inserting an extra half-plane of atoms, ABCD in a regular lattice. Most of the distortion will be concentrated around the line AD. An edge dislocation is created by applying a shearing force along the face of the crystal parallel to a major crystallographic plane, high enough to cause plastic deformation and move the upper half of the crystal. As observed, the Burgers vector is perpendicular to the line vector determining a glide plane on which the dislocation can move. The strain energy associated with an edge dislocation is

$$E_{edge} = \frac{Gb^2 l}{4\pi(1-\nu)} \ln\left(\frac{R}{r}\right) \tag{2.11}$$

where ν is the Poisson ratio (about 0.3 for Si and GaAs). Thus, it is approximately 60% larger than that for a screw dislocation, with important consequences on the direction that real dislocations like to assume. Two dangling bonds are present in each atomic plane along the direction of the axis; therefore a continuum of states can be expected in the band gap if no reconstruction occurs.

Due to the high energy of formation for both screw and edge dislocations, they cannot be thermally generated like point defects and thus their equilibrium concentration is negligible. They can move relatively easy by slip (along the glide plane) or climb (outside the glide plane with the help of point defects). Provided that dislocations can move within the crystal, they might come close enough to interact, tending to attract or repulse each other depending on whether their resulting strain field is smaller or larger than the individual strain

fields, respectively. Thus they can create knots (point where at least three dislocations meet) and networks. The real dislocations encountered in a semiconductor material usually are mixed, having both screw and edge character. They will tend to be straight meanwhile trying to have as large screw component as possible because the total energy is proportional to both the dislocation length and the Burgers vector according to (2.10) and (2.11). If enough energy is provided to the crystal, dislocations can also multiply forming closed loops.

Dislocations cannot end inside the crystal, but only at a surface, an internal surface or interface (e.g. grain boundary), a dislocation knot or on itself (forming a closed loop).

So far only perfect dislocations in a simple cubic lattice were considered. In fcc (face centered cubic) lattices dislocations have an additional option to lower their energy, by splitting into partials (dislocations having Burgers vectors smaller than a translation vector of the lattice) accompanied by stacking fault formation. The sphalerite structure (to which both Si and GaAs belong), has two atoms in the base of the crystal. Consequently, the extra lattice plane defining an edge dislocation may come in two modifications called “glide” and “shuffle” set, because it might end in two distinct atomic positions. The dislocations of the two sets have different core structures and dangling bond configurations. A partial belonging to the glide set has dangling bonds that can disappear upon reconstruction, while the one belonging to the shuffle set doesn’t have this property [62]. For silicon, the glide set seems to prevail, while for GaAs the situation is less clear. Upon reconstruction the core atoms form bonds pair-wise and states are removed from the band gap [63]. The same effect is expected when the dangling bonds are passivated with hydrogen atoms.

High quality single crystalline silicon comes nowadays (almost) dislocation free, while GaAs is somewhat poorer in this respect with values up to 5000 dislocations/cm^2 (which can be considered still a low value). For this reason dislocations don’t represent a problem by themselves, provided that the core is reconstructed. However, if there is an increased content of both point defects and dislocations, the former might interact with the strain field of the latter, often leading to segregation of metallic impurities around the dislocation core.

In summary, if there is to be any electrical activity associated with dislocations, it must be determined by point defects or by the interplay of point defect levels and dislocation states.

2.3.3. Grain boundaries

A grain boundary is a two-dimensional defect, defined as a thin layer of atomic dimensions in a crystalline material corresponding to the transition from one crystallographic orientation to another. A grain boundary extends over the whole bonded interface and governs the electrical activity of the interface. In the following, a rather intuitive description of the grain boundaries will be given in order to emphasize the main structural features.

Imagine a diamond structure in a <110> projection, where a cut along a [111] plane has been made such as to separate the crystal in two parts (Fig. 2.7, left).

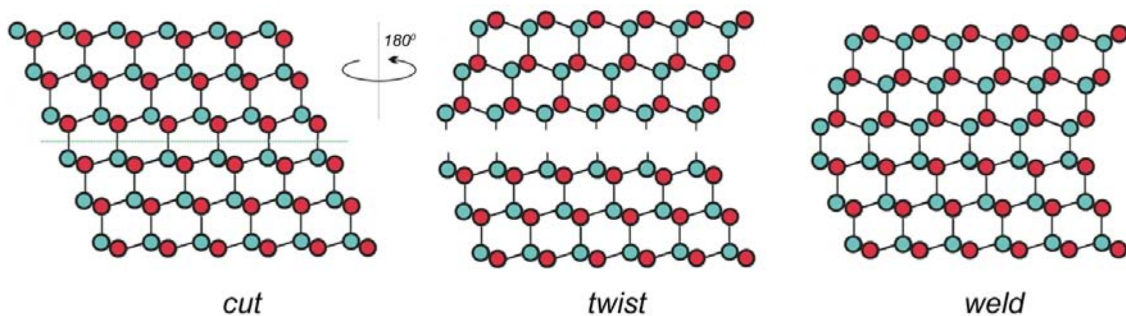

Figure 2.7: Formation of a twin boundary.

If the upper part is rotated by 180° around an axis perpendicular to the cut plane (Fig. 2.7, center) and then welded to the lower part again (Fig. 2.7, right), we obtain a structure that preserves the coherence of the lattice, therefore being termed as (coherent) twist boundary. We call it twist because the boundary was formed only by a twist operation. In a similar manner, the upper part can be tilted clockwise by a proper angle (70.53° around the <110> axis), and welded to the lower part after the necessary material removal, generating a (coherent) tilt boundary. The structure obtained in this way is identical to the previous one, and because of the mirror symmetry of the crystal along the cut/weld plane it is usually termed as twin boundary. The energy associated with a twin boundary (i. e. the energy spent in order to form one unit area of twin boundary) is quite low, since this structure preserves both the bond length and angle, and all dangling bonds find partners upon welding.

If the same exercise would be repeated using different twist (and tilt) angles, one observes that the two parts of the crystal won't fit with the same accuracy: a small number of bonds will directly find partners, some other bonds will be distorted while many bonds remain unsatisfied. Consequently, an increase in energy is expected to occur. Thus the twin boundary is a special case of a grain boundary, having the highest symmetry and the lowest possible energy. An important parameter that changes continuously upon twist and/or tilt of the two grains is the degree of coincidence (overlapping) between the lattices sites of the two crystals. This is described in terms of a coincidence site lattice (CSL) which shows that there are preferred orientations between grains that lead to high number of coincidence lattice points. Generally, low energy configurations correspond to a high number of CSL points [64], [65].

Of special interest are the small-angle grain boundaries formed when the grains have small misorientations with respect to those of a low energy configuration. In this case the grain boundary may decrease its energy by introducing grain boundary dislocations so that the dislocation free parts are now in a precise CSL orientation and the misalignment is taken up by the grain boundary dislocations. These grain boundary dislocations have the same properties as individual and separated dislocations, i. e. they posses strain and stress fields, energy, they can interact and form networks but they are generally confined to the boundary (except for the case when they split into partials). Additionally, their Burgers vectors take values belonging to the set of all possible displacement vectors which preserve the CSL.

Of course, upon bonding the wafers cannot be perfectly aligned. If we additionally consider the tilt associated with the wafer miscut, the steps and the possible foreign atoms at the wafer surface, we end up with a quite complicated grain boundary that posses energy, yields a high number of dangling (and distorted) bonds and consequently might have a quasi-continuous distribution of deep levels in the band gap by the same token as dislocations.

The determination of a dislocation structure needed to transform a near-coincidence boundary into a true coincidence boundary is a difficult task. There are two extreme cases where the misalignment is resolved in a simple manner: for a pure tilt boundary (by a sequence of edge dislocations) and for a pure twist boundary (by a network of screw dislocations). Fig. 2.8 depicts how edge dislocations can accommodate the misfit relative to a low energy orientation.

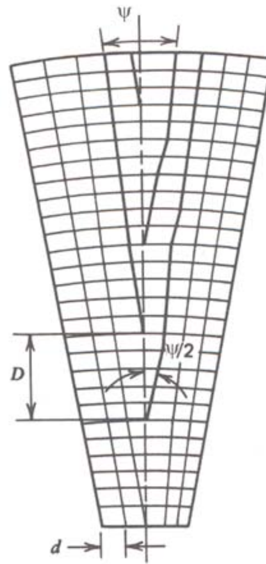

Figure 2.8: Model of a low-angle grain boundary.

If d is the atomic spacing, the distance between dislocations is given by:

$$D = \frac{d}{2\sin\left(\frac{\psi}{2}\right)} \approx \frac{d}{\psi} \tag{2.12}$$

where ψ is the tilt angle between the grains. The approximation in (2.12) holds only for small tilt angles (ψ expressed in radians). For $\psi \leq 1°$, the term lineage boundary is used, denoting a negligible loss of coherence. For $1° \leq \psi \leq 5°$ a low-angle grain boundary forms, preserving only partially the lattice coherence.

The electronic properties of a low-angle grain boundary can be perceived also by looking at the situation in a different way. It is well known that a freshly cleaved semiconductor surface exhibits a high number of deep-lying levels throughout the energy gap, due to the presence of unsaturated covalent bonds that are acceptor-like in character [66]. In an internal surface (or interface),

many of these bonds are terminated by a partial coherence of the lattice, leading to a reduction in the density of deep levels. Thus, the less dangling bonds the less space charge around the grain boundary.

2.4. Electrical characterization of bonded interfaces

First, the properties of the bonded interfaces in thermal equilibrium are described. Afterwards, the steady state condition in case of both unipolar and bipolar current flow is discussed in terms of a voltage dependent barrier height. The electrical conductivity is explained using the well-known thermionic emission theory. In the last part of this section, an overview of the drift-diffusion model extended to incorporate deep levels is given.

2.4.1. Grain boundary barrier in thermal equilibrium

As outlined in the previous section, the bonding process causes a continuum of interface states in the band gap of silicon. These states can be either acceptor-like, i. e. causing a negative interface charge in the ionized state, or donor-like, causing a positive interface charge in the ionized state. For charge neutrality, the interface charge is compensated by a complementary space charge at both sides of the interface. Generally, the interface causes both a potential barrier and superficial distribution of recombination centers.

Taylor et al. [9] observed that grain boundaries in germanium bicrystals behave like a rectifying element in blocking direction. They developed the model of anti-serial Schottky barrier (or double-Schottky barrier) used extensively in the literature since then to describe the properties of grain boundaries [67], [68], [69], [70], [71], [72], [73], [74].

In the following, an acceptor-like grain boundary in n-type silicon will be assumed. For simplicity, the full depletion approximation will be used, which asserts that the depletion region created by the interface charge does not contain mobile charges.

According to Fig. 2.9, the relation between the interface charge Q_S and the equilibrium depletion width x_{d0} is:

$$Q_S = -2qN_D^+ x_{d0} \tag{2.13}$$

The corresponding electrical field strength $\vec{E}$ and the electrostatic potential ψ at the interface are evaluated using the Poisson equation:

$$\Delta\psi = -\frac{\rho}{\varepsilon} \tag{2.14}$$

which gives:

$$E_{\max} = E(0) = \frac{qN_D^+ x_{d0}}{\varepsilon}$$
$$\psi_{\max} = \psi(0) = \Phi_B = \frac{qN_D^+ x_{d0}^2}{2\varepsilon} \quad (2.15)$$

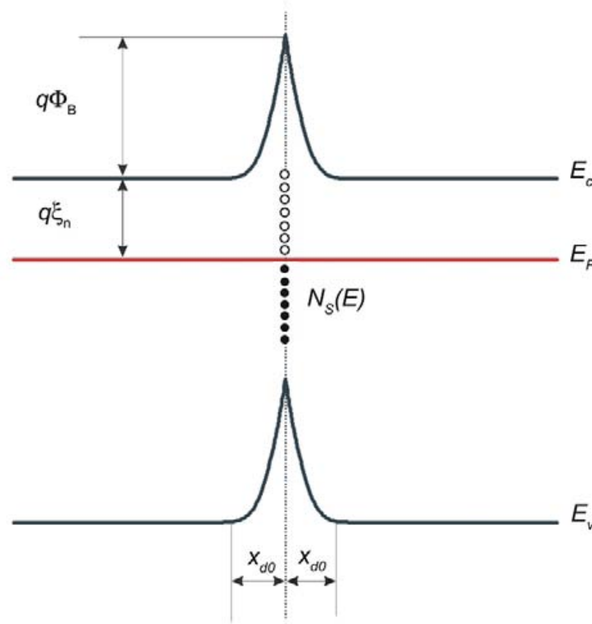

Figure 2.9: Grain-boundary barrier in thermal equilibrium.

Considering a more general situation when Q_S is represented by both donor and acceptor type levels, we have:

$$Q_S = q(N_{SD}^+ - N_{SA}^-) \quad (2.16)$$

where N_{SD}^+ represents the amount of ionized donors and N_{SA}^- the amount of ionized acceptors. In order to relate these quantities with the total available donor and acceptor states one can use the Fermi-Dirac distribution function:

$$f(E) = \frac{1}{1 - e^{\frac{E-E_F}{kT}}} \quad (2.17)$$

Thus the interface charge becomes:

$$Q_S = q\int_{E_{v,s}}^{E_{c,s}} N_{SD}(E)[1 - f(E)]dE - q\int_{E_{v,s}}^{E_{c,s}} N_{SA}(E)f(E)dE \quad (2.18)$$

The subscript “s” refers to the positions of valence and conduction band edges at the interface. One can see that the trap occupation strongly depends

on the relative position of the Fermi level in the band gap. For levels situated several *kT* below the Fermi level, all states are filled with electrons whereas trap levels several *kT* above the Fermi levels are free. Consequently, acceptor type levels are negatively charged for $E_T \ll E_F$ and neutral for $E_T \gg E_F$. Similarly, donor type traps are neutral for $E_T \ll E_F$ and positively charged for $E_T \gg E_F$.

Depending on the types and densities of the interface traps, the grain boundary can cause:

- depletion layers, when $E_F - E_{i,s} > 0$
- inversion layers, when $E_F - E_{i,s} < 0$
- accumulation layers, when $E_F - E_{i,s} < E_F - E_{i,b}$ (E_i denotes the intrinsic Fermi level and the subscripts "b" and "s" refer to bulk and interface values, respectively)

2.4.2. Unipolar current through the interface

The current flow above the potential barrier is modelled using the thermionic emission theory [75]. Thus, only carriers having energy greater than ($E_C + \Phi_B$) can contribute to the current. In thermal equilibrium, the barrier height that carriers need to overcome is identical for both sides of the grain boundary, yielding zero current.

As soon as a potential difference *U* is applied across the barrier, the energy bands on the right hand side of the grain boundary are lowered compared to the left exactly by this amount as depicted in Fig. 2.10.

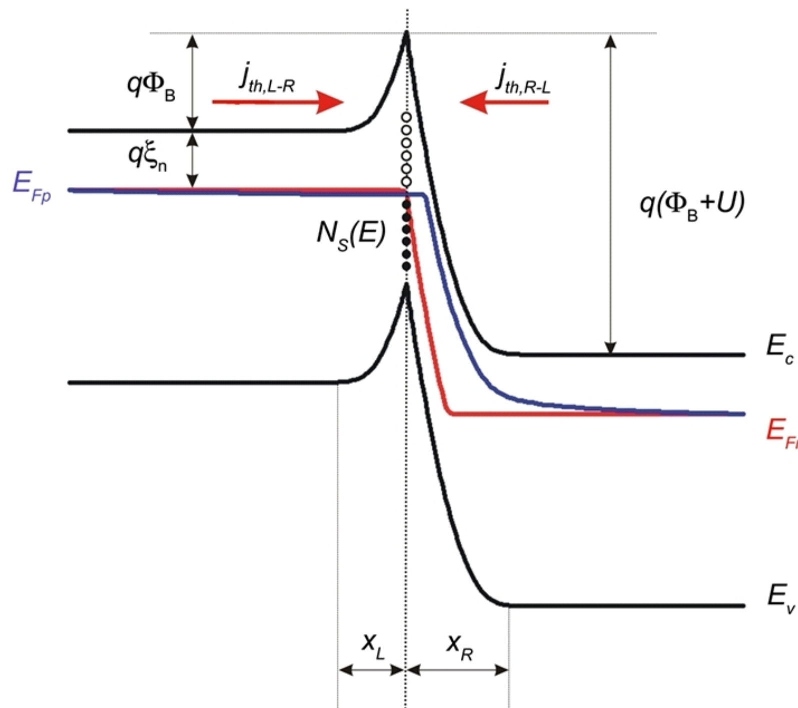

Figure 2.10: Barrier height and trap occupancy in the unipolar case.

The Fermi level splits into the quasi-Fermi levels (QFL) for holes and electrons, E_{Fn} and E_{Fp}, respectively, while the carrier densities outside the depletion region remain at their equilibrium values.

The thermally emitted current is given by the difference between the current components flowing into the interface from either side of the grain boundary:

$$\begin{aligned} J_{th,L\rightarrow R} &= -A_n^* T^2 e^{-\frac{q(\xi_n+\Phi_B)}{kT}} \\ J_{th,R\rightarrow L} &= -A_n^* T^2 e^{-\frac{q(\xi_n+\Phi_B+U)}{kT}} \end{aligned} \tag{2.19}$$

Thus:

$$J_{th,n} = J_{th,R\rightarrow L} - J_{th,L\rightarrow R} = A_n^* T^2 e^{-\frac{q(\xi_n+\Phi_B)}{kT}}\left(1-e^{-\frac{qU}{kT}}\right) \tag{2.20}$$

The Richardson constant for electrons A_n^* is given by:

$$A_n^* = \frac{4\pi q m_n^* k^2}{h^3} \tag{2.21}$$

where m_n^* is the effective mass of electrons in the crystal, h is Planck's constant and ξ_n is the demarcation between the conduction band edge in the bulk and the QFL for electrons. Since m_n^* depends on the wavevector $\vec{k}$, A_n^* depends on the crystallographic orientation to some degree.

The current increase is accompanied by a decrease of the potential barrier Φ_B. Consequently, the depletion width at the left of the grain boundary decreases, while the depletion width at the right increases according to:

$$\begin{aligned} x_L &= \sqrt{\frac{2\varepsilon}{qN_D^+}\Phi_B} \\ x_R &= \sqrt{\frac{2\varepsilon}{qN_D^+}(\Phi_B+U)} \end{aligned} \tag{2.22}$$

Due to the decreasing potential barrier height, the QFL for electrons approaches the conduction band edge, leading to an additional occupation of interface states:

$$\Delta Q_S = -q\int_{E_F}^{E_{Fn}} N_S(E) f(E)\, dE \tag{2.23}$$

For this reason, the maximum voltage drop at the interface corresponding to a collapse of the barrier can reach values as high as 100 V in the case of unipolar current flow for moderate doping levels.

2.4.3. Bipolar current through the interface

In the case of a bipolar current, both electrons and holes are involved in the transport process. The position of the associated QFLs is given by [76]:

$$n = n_i e^{\frac{E_{Fn}-E_i}{kT}}$$
$$p = n_i e^{\frac{E_i-E_{Fp}}{kT}} \tag{2.24}$$

An increasing electron concentration is accompanied by a rise of the QFL for electrons towards the conduction band edge, and an increasing hole concentration causes a rise of the QFL for holes towards the valence band edge (the direction of energy increase for holes is opposite to that of electrons). As a consequence, the trap occupancy will no longer be given solely by the position of the E_F and QFL for electrons. Fig. 2.11 depicts the situation in the case of bipolar current flow, where the trap occupancy changes according to the position of both QFLs.

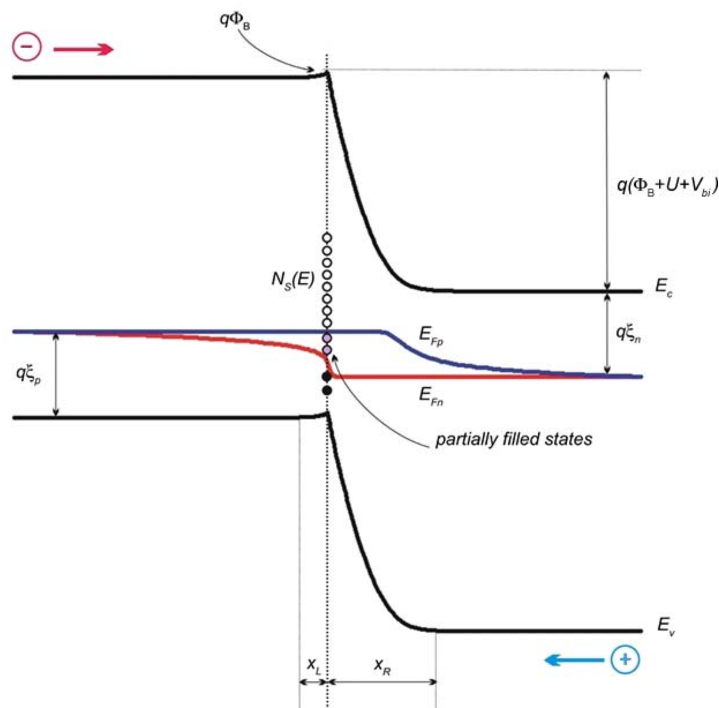

Figure 2.11: Barrier height and trap occupancy in the bipolar case.

In order to predict the trap occupancy, two simplifying hypotheses are made:

- no interaction between different trap levels is considered.
- the emission of electrons into the conduction band is neglected for trap energies lower than E_{Fp} and the emission of holes into the valence band for energies higher than E_{Fn}, respectively.

According to [77] the occupation probability in the case of a bipolar current can be approximated to:

$$f(E)=\begin{cases}1 & for \quad E<E_{Fp}\\ \dfrac{R_{th}n}{p+R_{th}n} & for \quad E_{Fp}<E<E_{Fn}\\ 0 & for \quad E>E_{Fn}\end{cases} \tag{2.25}$$

where R_{th} is the ratio between the capture coefficients of electrons c_n and holes c_p, respectively. Equation (2.25) tells us that the occupation probability decreases between E_{Fn} and E_{Fp}, leading to a decrease of the absolute total value of the interface charge:

$$\Delta Q_S = q\int_{E_{Fp}}^{E_{Fn}} N_S(E)\left[1-f(E)\right]dE \tag{2.26}$$

In contrast to the unipolar case, the decreasing potential barrier is accompanied by a decrease of the negative interface charge when applying a potential difference *U* across the interface.

Recalling (2.20) which describes the current flow across the interface in the case of unipolar transport, an extra contribution of holes thermionic current is added in a similar fashion:

$$J_{th,p} = A_p^* T^2 e^{-\frac{q\xi_p}{kT}}\left(1-e^{-\frac{qU}{kT}}\right) \tag{2.27}$$

where $\xi_p = E_{Fp} - E_{v,b}$ is the effective barrier for holes. For low injection conditions in an isotype junction (same type of doping in both sides of the interface), $\xi_p > \xi_n$ and the overall current is a pure electron current.

2.4.4. Generation-recombination statistics at traps

The above derived formulas for transport at bonded interfaces rely on the full depletion approximation, which gives inaccurate results in some cases. Moreover, the current across the interface as given by (2.20) depends on the barrier height Φ_B which in turn depends on the trap density (and occupancy) at the interface. By using the drift-diffusion model coupled with equations describing the charge conservation at traps, one can obtain a set of six coupled partial differential equations which can be numerically solved to give accurate simulation of bonded interfaces. This model covers almost anything except for tunnelling and impact ionization, provided that the necessary boundary conditions are given.

First, the generation-recombination process involving a trap and the conduction and valence bands will be briefly described. With a single trapping level E_T close to the middle of the bandgap (Fig. 2.12), four processes can occur.

In a first step, the trap can capture an electron from the conduction band at a rate given by:

$$r_c = \sigma_n v_{th,n} n N_T (1-f) \tag{2.28}$$

which depends on the capture cross-section for electrons σ_n, the thermal velocity of the electrons in the crystal $v_{th,n}$, the number of available traps N_T and the trap occupancy by the (presently unknown) distribution function f. The reverse process is called electron emission, given by:

$$g_c = e_n N_T f \tag{2.29}$$

e_n being the emission coefficient for electrons.

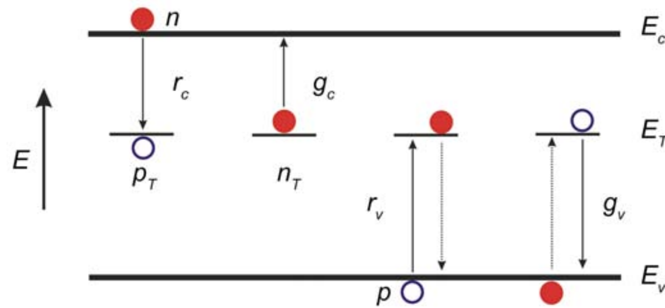

Figure 2.12: Capture and emission processes between the level E_T and the conduction and valence bands.

In a similar manner the fall of a trapped electron to the valence band (viewed as a hole capture from the valence band) is described:

$$r_v = \sigma_p v_{th,p} p N_T f \tag{2.30}$$

with σ_p and $v_{th,p}$ analogously defined as the capture cross-section for holes and the thermal velocity of holes, respectively. The reverse electron excitation from the valence band to the trapping level (viewed as a hole emission to the valence band) is described by the following equation:

$$g_v = e_p N_T (1-f) \tag{2.31}$$

e_p being the emission coefficient for electrons.

The difference between the captured and emitted electrons gives the net electron recombination rate. The net hole recombination rate is defined similarly:

$$\begin{aligned} U_n &= r_c - g_c \\ U_p &= r_v - g_v \end{aligned} \tag{2.32}$$

Finally, we are in a position to describe the trap occupancy in time (charge conservation):

$$\frac{\partial n_T}{\partial t} = U_n - U_p \tag{2.33}$$

We add the continuity equations for electrons and holes:

$$\begin{cases} \dfrac{\partial n}{\partial t} = \dfrac{1}{q}\nabla J_n - U_n \\ \dfrac{\partial p}{\partial t} = -\dfrac{1}{q}\nabla J_p - U_p \end{cases} \tag{2.34}$$

where J_n, J_p are the total electron and hole current densities (summing drift and diffusion components):

$$\begin{aligned} J_n &= q\mu_n n + qD_n \nabla n \\ J_p &= q\mu_p p - qD_p \nabla p \end{aligned} \tag{2.35}$$

μ_n and μ_p represent the electron and hole mobilities and D_n, D_p the diffusion coefficients for electrons and holes, respectively. The Poisson equation is needed for overall charge balance:

$$\Delta\psi = -\frac{q}{\varepsilon}\left(p - n + N_D^+ - N_A^- + n_T\right) \tag{2.36}$$

The term n_T accounts for charged traps in the volume of the semiconductor.

Although the system of equations (2.33)-(2.36) can only be solved numerically, the trap occupancy can be obtained in an analytical form for particular cases. In steady state U_n=U_p, i. e. the trap occupancy doesn't change any more with time. Moreover, in thermal equilibrium each emission process is balanced by the corresponding capture process. Then a crucial assumption (not necessarily valid in all cases) is made: the capture and emission coefficients in non-equilibrium remain equal to their equilibrium values $e_n \approx e_{n0}$, $e_p \approx e_{p0}$, $c_n \approx c_{n0}$, $c_p \approx c_{p0}$ [58], [78]. Letting $v_{th,n}$=$v_{th,p}$=v_{th}, we have:

$$e_n = \sigma_n v_{th} n_0 \frac{1 - f_0}{f_0} = \sigma_n v_{th} n_t \tag{2.37}$$

f_0 denoting the thermal equilibrium Fermi-Dirac distribution as given in (2.17) and

$$n_t = n_i e^{\frac{E_T - E_i}{kT}} \tag{2.38}$$

which is the concentration of electrons that would be in the conduction band if E_F=E_T. In a similar fashion:

$$e_p = \sigma_p v_{th} p_0 \frac{f_0}{1-f_0} = \sigma_p v_{th} p_t \,, \quad p_t = n_i e^{\frac{E_i - E_T}{kT}} \tag{2.39}$$

The non-equilibrium distribution function *f* can now be obtained recalling the steady-state condition, with the result:

$$f = \frac{\sigma_n n + \sigma_p p_t}{\sigma_n (n + n_t) + \sigma_p (p + p_t)} \tag{2.40}$$

The steady-state net recombination rate can be calculated:

$$U_n = U_p = U = \frac{\sigma_n \sigma_p v_{th} N_T (np - n_i^2)}{\sigma_n (n + n_t) + \sigma_p (p + p_t)} \tag{2.41}$$

this is the well-known Shockley-Read-Hall (SRH) formula for band-to-impurity recombination [79]. With additional definitions,

$$\frac{1}{\tau_{p0}} = \sigma_p v_{th} N_T \quad \text{and} \quad \frac{1}{\tau_{n0}} = \sigma_n v_{th} N_T \tag{2.42}$$

(2.41) becomes

$$U = \frac{np - n_i^2}{\tau_{p0}(n + n_t) + \tau_{n0}(p + p_t)} \tag{2.43}$$

For $\tau_{n0}=\tau_{p0}=\tau$, the expression for the net G-R rate can be further simplified:

$$U = \frac{np - n_i^2}{n + p + 2n_i \cosh\left(\frac{E_i - E_T}{kT}\right)} \frac{1}{\tau} \tag{2.44}$$

This tells us that the recombination efficiency increases as the trap level lies closer to the intrinsic Fermi level E_i. Therefore, midgap traps are very efficient G-R centers while shallower traps usually contribute to the band shift via traping-detraping mechanism involving only one type of carriers.

In the case of surface/interface traps, a surface recombination velocity is defined analogously to the minority carrier lifetimes given by (2.42) as:

$$S_{n,p} = N_S \sigma_{n,p} v_{th} \tag{2.45}$$