Jens Frühauf

Surface Dopant Interactions in Ultra-Shallow Junctions

Fakultät für Physik
Lehrstuhl für Experimentalphysik E16

Surface Dopant Interactions in Ultra-Shallow Junctions

Dipl.-Phys. Univ. Jens Frühauf

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. M. Kleber
Prüfer der Dissertation:	1. Univ.-Prof. F. Koch, Ph.D.
	2. Univ.-Prof. Dr. D. Schmitt-Landsiedel

Die Dissertation wurde am 8.11.2004 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 05.09.2005 angenommen.

to my parents

Summary

Modern semiconductor manufacturing in the sub-100 nm technology nodes uses ultra-shallow source and drain junctions in order to avoid short-channel effects that deteriorate the performance of a transistor. With the introduction of each new technology node, the junctions become shallower, and thus the influence of the surface on the dopant atoms increases. One of the major effects observed is a segregation of large amounts of the implanted dopants to the silicon-oxide interface during a spike anneal.

For this thesis, sheet resistance measurements, Secondary Ion Mass Spectrometry (SIMS) and high-resolution Elastic Recoil Detection (ERD) together with other techniques were used to investigate the physical and electrical behavior of As and B dopants in the vicinity of the interface. The results show that up to 70% of the dopants that remain in the junction after anneal are actually segregated to the interface, depending on implant and anneal conditions and the type of screening oxide used. For As, the pile-up thickness is limited to the thickness of the interface, i.e. to a few monolayers. For B, however, the pile-up is seen to extend slightly into the oxide in some cases, depending on the processing conditions.

A small part of these pile-up dopants was found to be electrically active. Active concentrations above $10^{21}\,\mathrm{cm}^{-3}$ are observed, well above the active concentration in the bulk part of the junction. At least for boron, these values are also well above bulk solid solubility.

The results indicate that the pile-up formation is strongly enhanced by oxidation during anneal. The interfacial atomic rearrangement during the oxidation process and the (anomalous) uphill diffusion of the dopants towards the surface during anneal are necessary for the creation of a pile-up at the interface. Oxide induced stress and interfacial trapping are believed to allow active concentrations above bulk solid solubility.

The first chapter of this thesis will briefly introduce the International Technology Roadmap for Semiconductors (ITRS) and the necessity of semiconductor device scaling, and then give an overview on the process steps in modern IC man-

ufacturing. In the second chapter, various analysis techniques are presented, with a special focus on the physical and electrical characterization of ultra-shallow dopant profiles.

In the third chapter, an introduction to the pile-up phenomenon together with an overview on existing literature is given, and the usability of different analysis techniques for pile-up profiling is evaluated. ERD and SIMS profiles are used in chapter four to investigate the pile-up shape and dose, and their dependence on implant and anneal parameters. Finally, the fifth chapter will discuss the electrical properties of the dopants trapped at the interface.

For convenience, numbers with exponents (especially for doses) will in some cases be written in the form 2e15, instead of $2 \cdot 10^{15}$. The standard units are cm^{-2} for doses and cm^{-3} for concentrations. Also, evident units will sometimes be omitted for better readability; e.g. in “a 1 keV, 3e15 B implant”, the 3e15 refers to the implanted dose, given in units of cm^{-2}, while 1 keV is the implant energy.

Publications

A part of the results of this work has already been published in the references [1–8]:

[1] J. Frühauf, R. Lindsay, A. Bergmaier, W. Vandervorst, G. Tempel, K. Maex, G. Dollinger, and F. Koch. Electrical activity of B and As segregated at the Si-SiO_2 interface. In *Mat. Res. Soc. Symp. Proc.*, volume 717, page C3.4, San Francisco, CA, 2002

[2] J. Frühauf, R. Lindsay, W. Vandervorst, K. Maex, A. Bergmaier, G. Dollinger, and F. Koch. Characterization of the B and As pile-up at the Si-SiO_2 interface. In *Proc. USJ-2003*, pages 399–404, Santa Cruz, CA, 2003

[3] R. Lindsay, A. Lauwers, J. Frühauf, M. de Potter, and K. Maex. Effect of implant oxide on ultra shallow junction formation. In *Proc. USJ-2001*, Napa, CA, 2001

[4] R. Lindsay, A. Lauwers, J. Frühauf, M. de Potter, and K. Maex. Effect of implant oxide on ultra shallow junction formation. *J. Vac. Sci. Tech. B, Microelectron. Nanometer Struct.*, 20(6):2225–8, 2002

[5] W. Vandervorst, T. Janssens, R. Loo, M. Caymax, I. Peytier, R. Lindsay, J. Frühauf, A. Bergmaier, and G. Dollinger. An (un)solvable problem in SIMS: B-interfacial profiling. In *Proc. SIMS-XIII*, Nara, 2001

[6] W. Vandervorst, T. Janssens, R. Loo, M. Caymax, I. Peytier, R. Lindsay, J. Frühauf, A. Bergmaier, and G. Dollinger. An (un)solvable problem in SIMS: B-interfacial profiling. *J. Appl. Surf. Sci.*, 203-204:371, January 2002

[7] W. Vandervorst, T. Janssens, J. Frühauf, I. M. Ross, A. Cullis, J. A. Van den Berg, A. Bergmaier, and G. Dollinger. Assessment of the near-surface profiling capabilities of SIMS. In *Proc. USJ-2003*, Santa Cruz, CA, 2003

[8] W. Vandervorst, T. Janssens, J. Frühauf, I. M. Ross, A. Cullis, J. A. Van den Berg, A. Bergmaier, and G. Dollinger. Assessment of the near-surface profiling capabilities of SIMS. *J. Vac. Sci. Tech. B*, (to be published)

Table of Contents

Summary **i**

Publications **iii**

Table of Contents **v**

1 Introduction **1**
1.1 History 1
1.2 An overview over the IC fabrication process 3
1.2.1 Front-End-Of-Line process 3
1.2.2 Back-End-Of-Line process 7
1.2.3 Testing and Packaging 8
1.3 The ITRS requirements for ultra-shallow junctions 9
1.4 State-of-the-art junction formation technologies 12
1.4.1 Ion Implantation 12
1.4.2 Rapid thermal processing (RTP) 14
1.4.3 Novel ultra-rapid annealing concepts (Flash, LTA) 18
1.4.4 Solid phase epitaxial regrowth (SPER) 19
1.4.5 In-situ deposition 21

2 Characterization methods **23**
2.1 Introduction 23
2.2 Definition of the sheet resistance 23
2.3 FPP (Four Point Probe) 24
2.3.1 In-line tip arrangement 24
2.3.2 Van-der-Pauw structure 26
2.4 SRP (Spreading Resistance Profiling) 27
2.5 PS (Probe Spacing) 29
2.6 Hall Effect 29
2.7 SIMS (Secondary Ion Mass Spectrometry) 31
2.8 High-resolution ERD and RBS 33
2.8.1 ERD (Elastic Recoil Detection) 34
2.8.2 RBS (Rutherford Backscattering Spectrometry) 36
2.8.3 MEIS (Medium Energy Ion Scattering) 37

2.9 TEM (Transmission Electron Microscopy) . . . 37
2.10 XPS (X-Ray Photoelectron Spectrometry) . . . 39
2.11 Ellipsometry . . . 40

3 Pile-up fundamentals 43
3.1 Initial experiments . . . 43
3.2 The interfacial pile-up . . . 45
3.3 Pile-up profiling techniques . . . 46
3.3.1 ERD/RBS results . . . 48
3.3.2 SIMS results . . . 49
3.3.3 TEM and TEM-EDX results . . . 51
3.4 Dopant diffusion . . . 54
3.4.1 Enhanced diffusion . . . 54
3.4.2 Uphill diffusion . . . 56
3.5 Effects mentioned in the literature . . . 57
3.5.1 Theoretical work . . . 57
3.5.2 Experimental Results . . . 58
3.5.3 Contribution of this work . . . 60

4 Pile-up formation mechanisms 61
4.1 Introduction . . . 61
4.2 Sample preparation . . . 61
4.3 Dopant loss mechanisms . . . 63
4.4 Dependence on implant conditions . . . 67
4.4.1 Species dependence . . . 67
4.4.2 Dose dependence . . . 69
4.4.3 Energy dependence . . . 79
4.4.4 Different screening oxides . . . 82
4.5 The role of oxidation during anneal . . . 86
4.5.1 Oxygen in the annealing ambient . . . 86
4.5.2 Influence of oxidation on the pile-up . . . 89
4.5.3 Anneals without oxidation . . . 93
4.6 Influence of the annealing temperature . . . 95
4.7 Surface treatment to locate the pile-up . . . 98
4.8 A model for the pile-up formation . . . 99
4.9 Diffusion mechanisms . . . 102
4.10 Pile-up on SPER annealed samples . . . 103

4.11 Summary 104

5 Electrical effects in the pile-up **107**

5.1 Introduction 107
5.2 Accuracy of sheet resistance measurements 107
5.3 Surface treatment methods 108
5.3.1 Oxide removal by HF dips 109
5.3.2 Surface quality after oxide removal 111
5.3.3 Etching selectivity 111
5.3.4 Repeatability 117
5.3.5 Conclusions on accuracy and repeatability 119
5.4 Observation of native oxide growth 120
5.4.1 Experimental conditions 122
5.4.2 Measurable electrical effects 123
5.4.3 Band bending at the surface 124
5.4.4 Side effects reducing mobility 125
5.5 Conductivity of the pile-up 126
5.5.1 Quantitative results 126
5.5.2 Interpretation of the measured concentration 128
5.5.3 Repeated HF dips 129
5.5.4 Overview and comparison 131
5.5.5 SRP measurements 132
5.6 Influence of implant and anneal parameters 133
5.6.1 Different screening oxides 134
5.6.2 Influence of oxidation 134
5.6.3 Influence of the thermal budget 136
5.7 Summary 138

6 Conclusions **141**

Acknowledgements **145**

Bibliography **149**

Index **161**

1 Introduction

1.1 History

In 1965, nine years before the first processor, Gordon E. Moore observed that the number of transistors integrated on a single chip was doubling every year. He predicted a continuation of this exponential growth for at least ten years, achieving 65,000 transistors per chip in 1975 [9]. This goal was missed only by a factor of 2, and his principal statement of a continuing exponential growth is still valid today, almost 40 years later. The number of transistors per chip is doubling about every two years (fig. 1.1), while the price of one transistor is reduced by a factor of two every 18 months. Moore's law also applies to the costs for new chip fabrication plants ("fabs") that used to double every two years. However, this trend has recently slowed down, after having reached $1B per fab in the mid-90ies [10].

The main driving factors that make today's semiconductor industry still follow Moore's Law are production cost and performance. By reducing the size of every individual transistor, more dies can be produced on the same wafer. This leads to an important reduction in unit production costs, even though the production tools for new technologies are becoming increasingly expensive. At the same time, the decreasing size of transistors and interconnects together with the use of new production techniques and materials reduces electrical capacity and resistance, and allows higher operation frequencies [11].

International Sematech, a consortium of the world's most important semiconductor companies, organizes every two years the publication of the "International Technology Roadmap for Semiconductors" (ITRS) [12]. It gives a prediction about the future development of semiconductor technology, the needs of the industry and the physical and technological challenges to be overcome. In the ITRS, so-called "technology nodes" are defined as a set of parameters to produce state-of-the-art ICs of a certain size. The DRAM memory cell is used as standard: The name of the respective technology node corresponds to half the distance of two

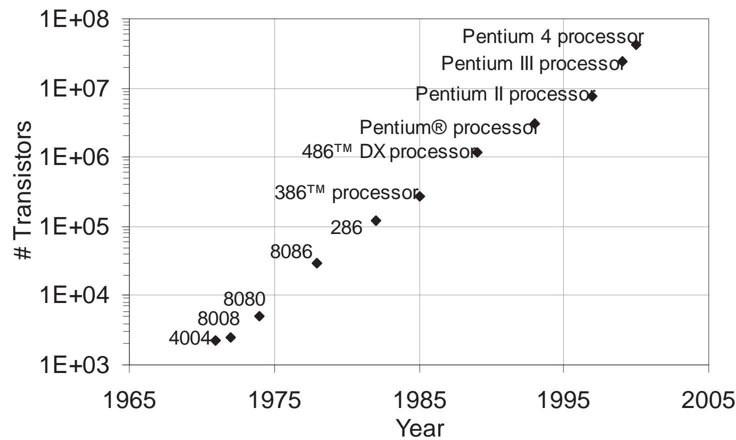

Figure 1.1: Number of transistors per chip on Intel processors [13]

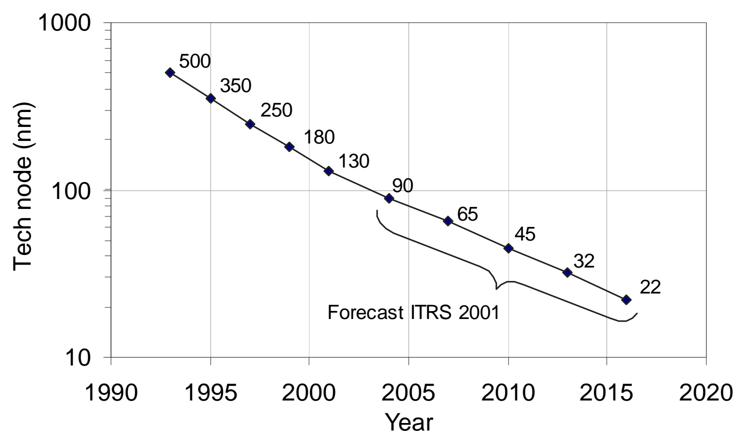

Figure 1.2: Technology nodes since 1993

word- or bitlines ("DRAM half-pitch"). A standard DRAM cell typically occupies an area of $8F^2$ with F being the technology node.

Between two nodes, the DRAM half-pitch is scaled by a factor of around 0.7, thus the cell size is halved. While in the 1990's, a new tech node was introduced every two years, the development has slowed down to one node in three years since the beginning of the new century (cf. fig. 1.2). In order to improve (or at least not deteriorate) the device's electrical characteristics from one tech node to the next, new materials and manufacturing techniques have to be developed. Along with the node and its definition of the structures' sizes, the ITRS describes a wealth of different requirements for physical properties and production accuracy.

The work described in this thesis was done to contribute to the development of FEOL (Front-End-Of-Line) technology: If the lateral dimension of a transistor shrinks, also its vertical size must be reduced to avoid detrimental electrical effects[1]. The doped regions of source and drain are made shallower, hence the name "Ultra-Shallow Junctions". As a consequence, the dopants come closer to the surface of the silicon, and the surface's influence on the electrical properties of the entire layer becomes increasingly important.

This work has been focused on the behavior of dopants in close proximity to a free surface or an interfacial material. Distribution and electrical activation of these dopants have been examined using various conventional and novel analysis techniques, in order to get

> "A more fundamental understanding of what's happening at the surface".

1.2 An overview over the IC fabrication process

The fabrication of a CMOS die consists of two main parts: The FEOL processing, which builds the individual transistors (also called "devices"), and the BEOL (Back-End-Of-Line) processing, which creates one or several metal layers to connect the transistors to each other and to the final package's exterior connectors. In the following, a typical deep-submicron CMOS process flow is briefly described, as an example for a state-of-the art CMOS process. The steps are illustrated in figs. 1.3 and 1.5. Details such as the materials used for adhesion or barrier layers, or the combination of species, doses and energies for the various implants are the intellectual property of each manufacturing company and can differ significantly from the process described here.

1.2.1 Front-End-Of-Line process

1. The process starts with a blanket, p-doped (100) wafer. After a cleaning step, which removes any native oxide, a thin protective oxide layer is grown.

[1] Ultra-shallow junctions are mainly introduced to avoid short channel effects and to correctly adjust the overlap capacity between source/drain and the gate [14].

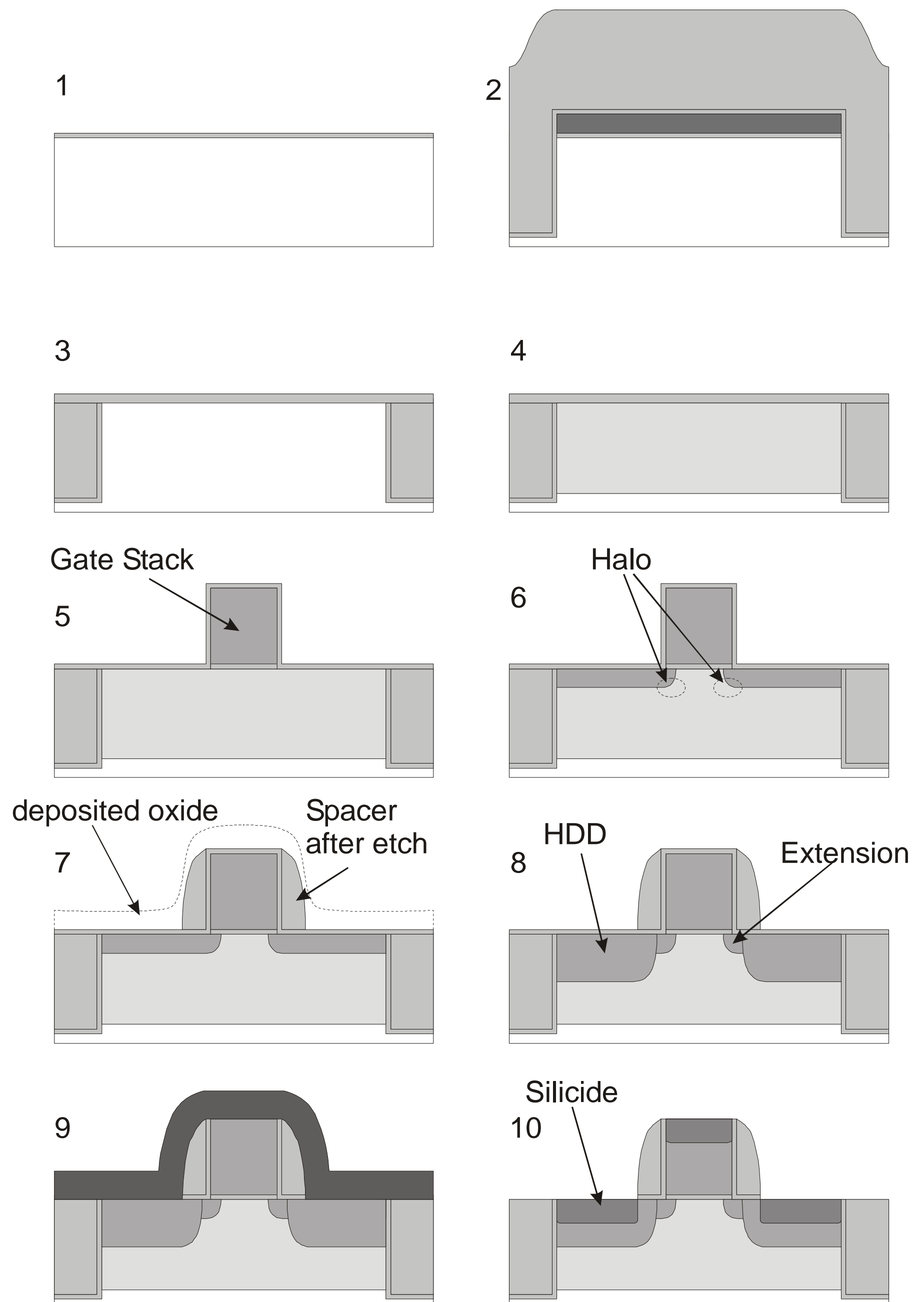

Figure 1.3: Major FEOL Processing steps

2. STI (Shallow Trench Isolation): A nitride layer is deposited and patterned, to be used as a hardmask. In the areas not covered by nitride, a trench is etched into the silicon.

3. On the trench walls, a carbide (SiC) layer is grown in a furnace, and the trench is filled with a CVD oxide. A CMP step removes excessive oxide and planarizes the surface, then the nitride is removed in a wet etch step. With this network of oxide-filled trenches, the wafer is now divided in isolated "active areas", each of which will later contain one transistor. Finally, an oxide layer is grown to cover the entire wafer.

4. Well implant: All areas that are supposed to contain an nMOS transistor are covered with resist. A deep, n implant is done to create the well for the pMOS transistors. An additional low dose shallow implant ("V_t-adjust") is done to fine-tune the channel doping, in order to obtain the desired threshold voltage V_t. After covering the the pMOS areas with resist, an analogous procedure is repeated for the nMOS transistors. The dopants are activated in a furnace anneal.

5. Gate stack formation: After removing any resist, the oxide layer on the entire wafer is removed in a short etch step. In a furnace, a high-quality oxide is grown that will later serve as gate dielectric. By DPN nitridation, the dielectric constant of the oxide is increased to a value between 5 and 6. On top of the gate oxynitride, a polysilicon ("poly") layer is deposited in a CVD process. A SiON layer is deposited to serve as a hardmask After lithography, polysilicon and gate oxide are removed in a dry etch step (patterning), leaving only the gate stack of each transistor. Resist and hardmask are removed in a strip and a selective wet etch step. This damages the sides of the gate oxide by slightly etching under the gate. To repair this damage, a thin oxide layer ("reox", 1 nm) is grown over the whole wafer, including the gate stack sidewalls.

6. Extension implant: With again all nMOS transistors covered with resist, a medium energy, low dose As implant with a high tilt angles (25°) is used to create the Halo[2]. The idea of the halo is to increase the well doping below the channel, and thereby to reduce the width of the depletion zones in this region. Short channel effects that arise from overlapping source and drain depletion

---

[2] All implants are done at four quadrants: Four individual implants (with 1/4th of the total dose each), with the wafer being twisted in the four directions 0°, 90°, 180°, 270°. This guarantees a uniform dopant distribution when using tilted implants.

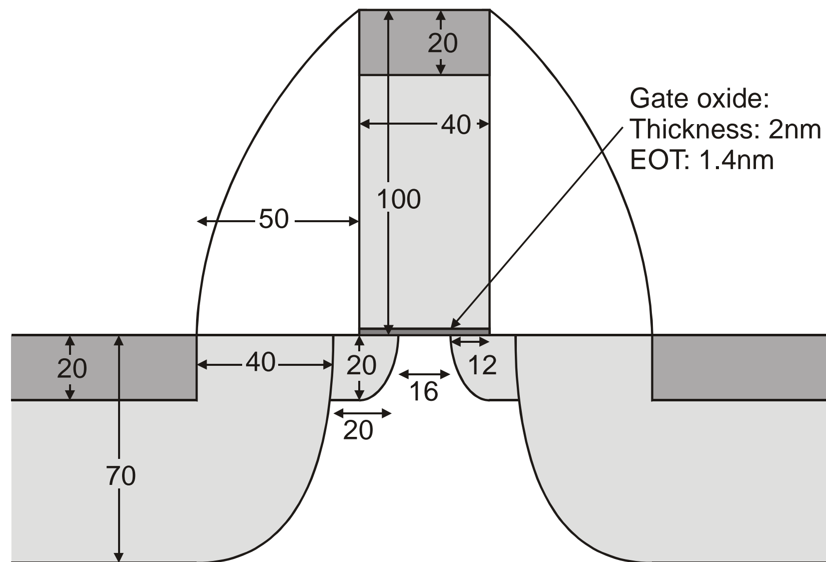

Figure 1.4: Transistor of the ITRS 90 nm technology node, drawn to scale. All dimensions are approximate values, given in nm. The lateral diffusion of the extension under the gate and of the HDD under the spacer are around 0.6 times their respective depths. A heavily nitrided oxide with a dielectric constant between 5 and 6 is used as gate oxide, therefore the EOT (equivalent oxide thickness) is smaller than the physical thickness.

zones are avoided. An ultra low energy, high dose Boron implant forms the highly doped extension. The extension implant is done at a tilt angle of 7° to avoid channelling (cf. p. 13). The procedure is repeated accordingly for the nMOS transistors, using a medium energy B implant and a ultra low energy As implant for halo and extension, respectively.
Since this work will focus mainly on the extension characteristics and less on the HDD areas, the word "junction" will be from now on used synonymously to "extension".

7. Spacer: A CVD oxide layer and a thick TEOS oxide are deposited over the whole wafer and partly removed in an RIE step. Since this etch is anisotropic, the oxide is entirely removed on all flat regions (on top of the gate and on the active areas), but some oxide remains in the corners on both sides of the gate stack.

8. HDD implant: After spacer formation, the transistors are implanted with low energy, high dose Boron or arsenic, respectively, to create deep source and drain areas ("HDD", highly doped drain). They are necessary to provide a good

contact between the silicide and the extension. The gate stack and the spacer serve as mask for this step. Finally, HDDs and extensions are annealed in a spike anneal. These implants also serve for gate doping.

9. Silicidation: In order to create a low-resistance contact between the silicon and the metal lines connecting the transistors, a silicide is formed on source and drain and on the poly-Si gate. For this, the surface is cleaned with a dip in diluted HF. Then, a nickel layer is deposited over the whole wafer.

10. During a low-temperature soak anneal ("RTP1", 320 °C), a part of the nickel reacts with the underlying silicon and forms a Ni_2Si silicide. There is no reaction between nickel and Oxide, i.e. the silicidation only occurs on the silicon regions. Hence the name "Salicide" for Self-aligned silicide. After RTP1, all Ni which has not reacted to NiSi is removed in a selective etch step. Finally, a second soak anneal ("RTP2", 450 °C) transforms the Ni_2Si into NiSi, which has a significantly lower resistivity. Fig. 1.4 shows the final transistor, drawn to scale.

1.2.2 Back-End-Of-Line process

The BEOL part of a chip consists of up to 9 layers of metal. The lower metal layers typically contain short and narrow lines for local connections, while the lines in higher layers become increasingly large to allow low-resistive connections over the entire chip. The last metal layer contains large square areas that are used as pads for external contacting (bonding).

1. A protective SiC layer and a HDP-CVD oxide are deposited over the whole wafer and flattened by CMP. An additional CVD oxide layer is deposited.

2. Lithography and a dry-etch process are used to open contact holes in the oxide above the source, drain and gate of each transistor. A strip removes the resist. To fill the holes, a thin Ti/TiN is deposited as diffusion barrier and contact layer. Then, Tungsten is CVD-deposited to fill up the contact holes. All material outside the contact holes is removed by CMP. This procedure is called a "damascene" process.

3. To form the lines in the first metal layer, a thin Ti barrier and seed layer is deposited, followed by an AlCu alloy, and a Ti / TiN cover. The metal layer is patterned by a lithography / metal etch / resist strip step. The openings are filled with a CVD oxide, and finally a CMP step removes excessive oxide.

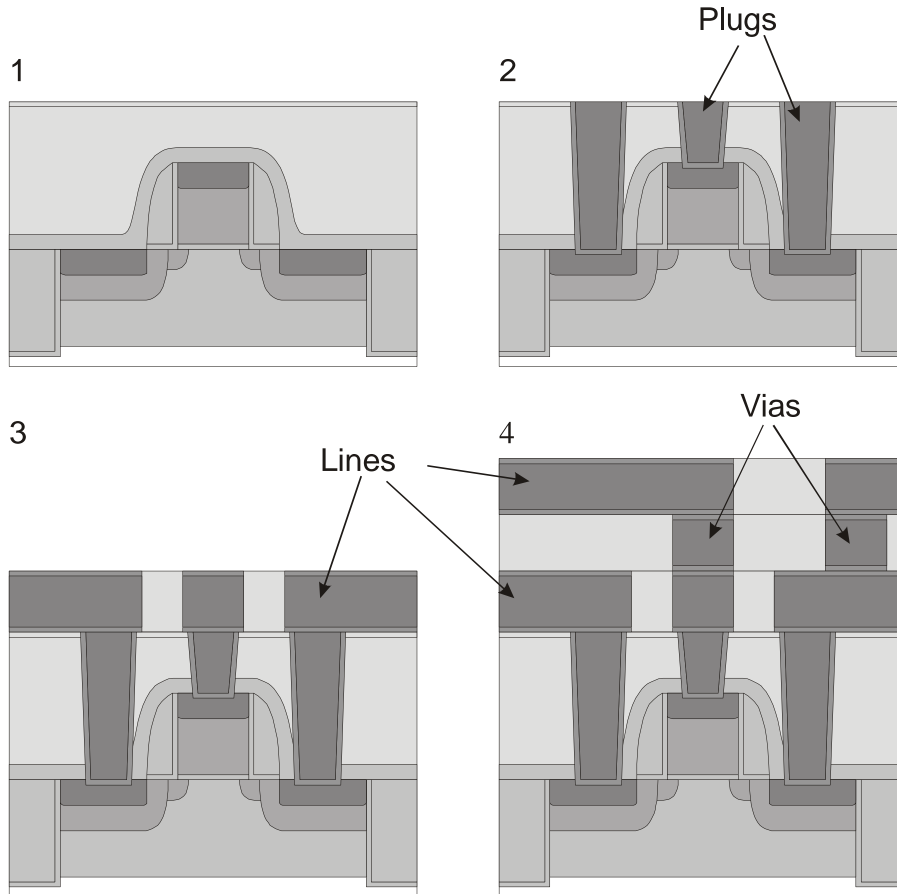

Figure 1.5: BEOL Processing steps

4. Above the first metal layer, the first via layer is created either by the method described in step 3 or by a damascene process (step 2). Then the second metal layer is formed analogous to step 3. This procedure is repeated several times. Typical chips contain four to nine metal layers.

5. At the end of the BEOL process, the last metal layer is covered with a protective SiN layer. Only the bond pads are left open. Finally, the wafer is annealed in a H_2 or forming gas ambient. The hydrogen diffuses through the wafer and terminates dangling bonds.

1.2.3 Testing and Packaging

During the described FEOL and BEOL process, entire wafers are handled that can contain up to several thousand identical dies. The wafers are then sawn, and the individual chips are packaged into plastic or ceramic packages that allow an easy

year	2001	2004	2007	2010	2013
node (nm)	130	90	65	45	32
min. phys. gate length (nm)	65	37	25	18	13
junction depth (X_j) (nm)	27-45	15-25	10-17	7-12	5-9
max. R_s for pMOS ($\Omega/\square$)	400	660	760	830	940
max. R_s for nMOS ($\Omega/\square$)	190	310	360	390	440

Table 1.1: Requirements for shallow extensions as defined by the 2001 edition and the 2002 update of the ITRS.

handling and electrical connection. Several testing steps on wafer level and after packaging ensure that only functional chips are sold.

1.3 The ITRS requirements for ultra-shallow junctions

The International Technology Roadmap of Semiconductors ("ITRS") describes the present and predicted future timeline for the introduction of new technology nodes into mass production [12]. For each node, detailed specifications are given on the requirements of FEOL Processing, Lithography, Interconnects, etc. It is set up and updated by the International Sematech ("ISMT") consortium in Austin, TX. Since this work will focus on the issues associated with the ultra-shallow junction ("USJ") processing, a brief introduction to ITRS requirements on the junction extension will follow, as described in the present ITRS edition (2001 edition with the 2002 update). A general overview on the challenges of manufacturing transistors in the deep sub-micron regime is given in [15, 16].

The main physical problem that leads to the development of ever-shallower junctions are the short-channel effects ("SCE"). Since the gate (and channel) length is reduced for each technology node (see table 1.1), the source and drain extension regions approach each other. As soon as the depletion regions start to touch, the off-state leakage of the transistor increases dramatically. Two approaches are used to avoid this effect: Firstly, one tries to reduce the junction depth X_j, in order to increase the separation of source and drain. As such, the ITRS set very stringent limits to the junction depth. Secondly, the halo implants (cf. fig. 1.3, no. 6) increase the background doping in the problematic area and thereby reduce the width of the depletion zone.

There is also a lower limit defined for the junction depth. The lateral extension of the junction under the gate is proportional to the junction depth (the proportionality factor is typically around 0.6). A too short lateral extension reduces the overlap between the channel and the junction, which increases the overall resistance. However, this limit is rarely discussed, as it is difficult enough to produce a junction as shallow as the maximum allowed X_j.

The other important parameter is the resistance of the junction, discussed here in the form of a sheet resistance R_s. A detailed definition of the sheet resistance is given in section 2.2 (page 23).

The extension contributes around 10% to the total series resistance of the transistor. Even though this appears to be only a very small part, still an optimization is crucial to fulfil the stringent ITRS requirements. Due to the intrinsic difficulty to manufacture junctions that are shallow *and* have a low sheet resistance, the ITRS 2001/02 requirements have been significantly relaxed, compared to the earlier editions. While junctions get shallower with every technology node, a little increase in sheet resistance is permitted.

A trade-off exists between the sheet resistance R_s and the junction depth X_j: By reducing the thermal budget (e.g. the temperature and time) of the junction anneal, X_j can be lowered due to reduced diffusion. However, this normally deteriorates the activation of the implanted dopants, which again increases R_s. Secondly, it is possible to increase the dopant dose in the junction to reduce the resistance. But an increased dose leads again to increased diffusion (especially for Boron: BED, Boron Enhanced Diffusion), and to reduced mobility of the carriers. Furthermore, a certain solid solubility limit of the dopants cannot be exceeded. Additional dopants form inactive clusters that even lead to a reduced mobility. By varying the implant and anneal parameters it is therefore possible to trade a low sheet resistance against the shallowness of the junction [17, 18].

Fig. 1.6 displays the requirements for sheet resistance and junction depth for a pMOS transistor, as defined in the ITRS for the upcoming technology nodes. The lines indicate the limits in R_s and X_j that can be reached with a specific implantation/anneal technology for pMOS junctions. By fine-tuning the implant and anneal parameters, any point on (or above) the respective line can be achieved. However, it is very difficult or impossible to reach the area below and to the left of the line.

The line designated "B" (also called "universal curve" or "Sematech curve") shows the best published results using conventional Boron ion implantation and spike anneal. It touches the ITRS requirements for the 130 nm node, but for the

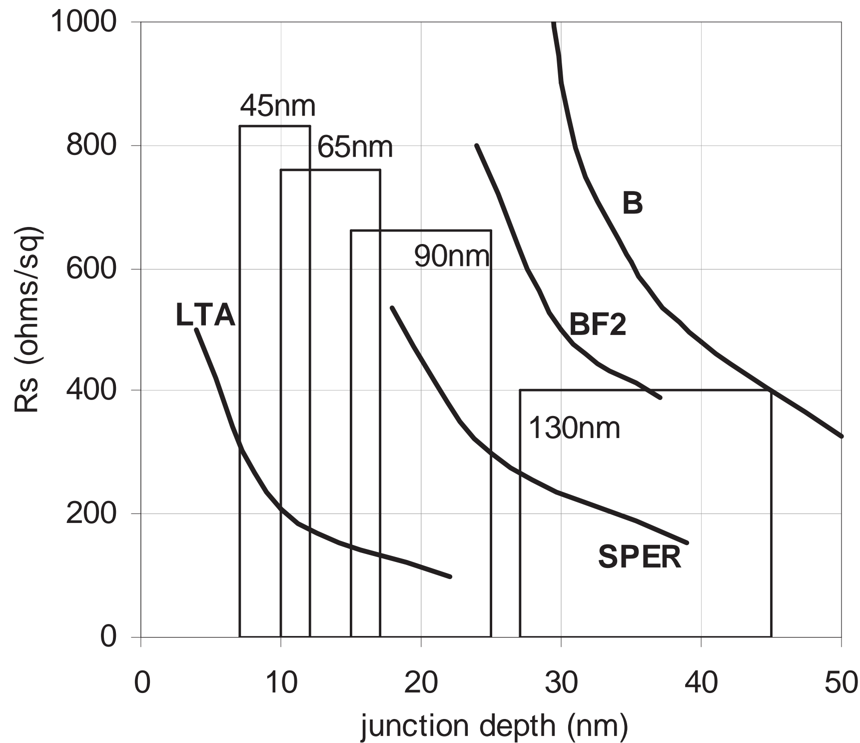

Figure 1.6: R_s and X_j pMOS Requirements of the ITRS for present and future technology nodes. The limits achievable with different junction formation concepts are indicated as lines.

sub-100 nm nodes, significant improvements are necessary. Several new junction formation concepts allow a junction performance better than the universal curve; estimated limits for the most important are shown as additional lines in fig. 1.6. The details of these concepts will be assessed in the following sections.

For nMOS, it is typically easier to achieve good results, since the widely used dopant arsenic has better properties than boron: It diffuses less and higher active concentrations are possible. However, since the ITRS requirements on sheet resistance are more stringent for nMOS than for pMOS (see table 1.1), similar difficulties are seen, and the mentioned novel junction fabrication techniques have to be assessed with respect to both pMOS and nMOS. An overview on modern USJ formation technologies and their practical physical limits is given in [19, 20].

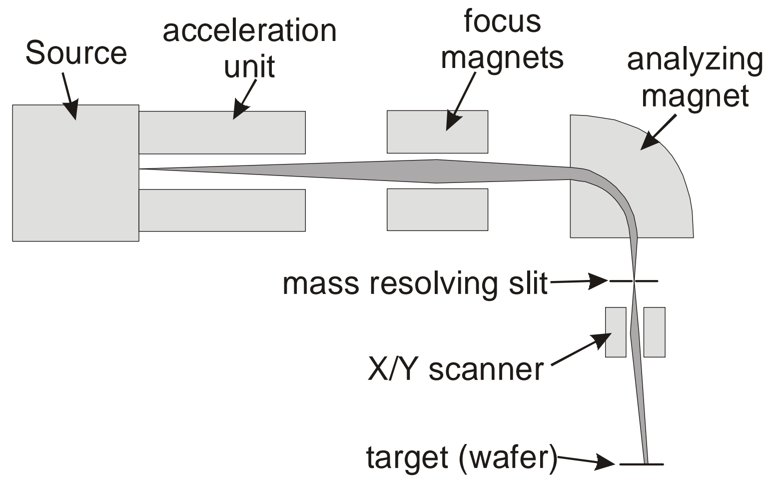

Figure 1.7: Schematic drawing of an ion implantation tool

1.4 State-of-the-art junction formation technologies

The classical formation of a junction consists of the following two steps: First, dopants are added to the pure or pre-doped silicon. Secondly, the dopant atoms need to be activated by placing them into substitutional sites in the silicon lattice. This is done in a thermal step (anneal), which at the same time removes silicon interstitials from the bulk lattice and heals other implant damage.

1.4.1 Ion Implantation

Ion implantation ("I/I") is the main commercially used technique for doping. Dopant atoms or molecules (provided as gas) are ionized in the "source", then extracted, accelerated and focused into an ion beam. A mass filter (e.g. a 90° magnet) selects only one species of dopants. The beam is then scanned across the wafer to provide a homogenous doping (figure 1.7).

Typical dopants used include B and In (p doping) or P, As and Sb (n doping). Typical energies range from 1 to 150 keV. For the upcoming sub-100 nm technologies that require shallower junctions, energies down to 500 eV are investigated. Table 1.2 shows projected ranges[3] for some implant conditions, a typical profile is given in fig. 2.5 (page 32).

The energy of the implanted ions is transferred to the target atoms in a collision cascade. 10-1000 target Si atoms are displaced for every implanted dopant atom,

[3] The projected range is the depth of the implant profile's maximum.

Energy / keV	0.5	1	2	5	10
B	3.2	5.3	9.0	19.7	37.3
BF_2	1.2	1.9	3.0	5.8	10.0
As	2.5	3.4	4.8	7.9	12.1

Table 1.2: Projected range in silicon (R_p, given in nm) for some implant conditions. In SiO_2, R_p is around 20% lower. [21, 22]

most of which recombine at room temperature with a vacancy. After implant, about one interstitial Si atom per implanted atom is left in the crystal. Due to the high number of collisions, the displacements are mainly isotropical.

Implants suffer from an effect called “channelling”. If the implant angle is close to the angle of a major lattice axis, a fraction of the implanted ions is diverted into the direction of this axis. Travelling along this axis, their probability of being scattered is significantly reduced, and they penetrate deeper into the substrate before being stopped (see also fig. 2.5). To avoid this, an implant angle has to be chosen that lies as far from any major lattice axis as possible. The implants done for this work were all done at an angle of 27° twist and 7° tilt. The acceptance angle of the channelling directions becomes larger at low implant energies. Therefore, the channelling effect becomes more important at low implant energies. Below a few 100 eV, it cannot be avoided any more.

A common problem when using implanters in the ultra-low energy regime (below 1-2keV) is a rapid drop in beam current, leading to extremely long implantation times. This is due to the space-charge effect (maximum possible current $\propto$ energy$^{3/2}$), which also causes an unwanted widening of the beam at low energies [23]. One way to circumvent this is to extract the beam at a relatively high energy (3-4 keV) and then decelerate the ions to the desired energy just before hitting the wafer. However, if an ion gets neutralized before deceleration, it hits the wafer with the original extraction energy. Even though this happens typically with less than a percent of the implanted dose, this “energy contamination” can lead to a significant increase in junction depth and a deterioration of the abruptness. In a SIMS graph, energy contamination is not discernible from channelling (see fig. 2.5).

Another way to achieve high beam currents with low energies is to implant clusters (BF_2, SiB, SiB_2...) instead of atoms [23]. For example, BF_2^+ can be implanted with an energy of 2.2 keV, resulting in the same B profile as a 0.5 keV B^+ implant. In the case of boron, the co-doping of fluorine is often desired, since it re-

duces the diffusion during anneal and thus leads to shallower junctions (compare fig. 1.6).

An alternative doping method that is currently investigated for future generations requiring even shallower profiles is Plasma Doping ("PLAD"). In this technique, a plasma is ignited above the wafer. By applying a bias voltage on the wafer, dopant ions are accelerated towards the wafer and implanted. The advantages are a high throughput (the entire wafer is implanted at once, no beam scanning is necessary) and, of course, the possibility to implant at extremely low energies, down to the mere deposition of dopants at the wafer surface. On the negative side, it is difficult to control the implanted dose and to achieve a good uniformity over the wafer. The implant angle is almost impossible to control, and since there is no possibility of selectively implanting one dopant species, the risk of contamination is very high.

1.4.2 Rapid thermal processing (RTP)

After ion implantation, the lattice structure in the wafer's top layer is heavily damaged. A thermal annealing step is necessary to repair the damage and to activate the dopants by placing them on substitutional lattice sites. Diffusion of the dopants is desired in some cases (e.g. well formation) and avoided in other cases (e.g. USJ formation).

After an USJ anneal, the resulting dopant profile should mainly meet three requirements: It should be shallow, the active concentration should be as high as possible, and the profile should be as abrupt as possible. Therefore, the USJ anneal is optimized for good activation and low diffusion.

The main physical problem for annealing is the fact that the activation energy for diffusion is lower than the one for dopant activation [24, 25]. In order to optimize the trade-off between a good activation and as little diffusion as possible, high annealing temperatures have to be used. Also, the solid solubility of the dopants in the lattice is better for high temperatures.

The ITRS specifications for sub-100nm technologies therefore require an anneal temperature of 1100 °C or above, where the solid solubility is around $1.5 \cdot 10^{21}\text{cm}^{-3}$ for As and $2.5 \cdot 10^{20}\text{cm}^{-3}$ for B, close to the maximum[4]. Since the diffusion length l_d follows the proportionality

[4] The maximum solid solubility is $1.8 \cdot 10^{21}$ for As (reached at around 1230 °C) and $8 \cdot 10^{20}$ for B (1400 °C). [26]

$$l_d \propto \sqrt[3]{Dt} \tag{1.1}$$

with the diffusivity $D \propto e^{-\Delta E/kT}$ at a given Temperature [27], extremely short anneal times are required to obtain a shallow and abrupt profile. This can be achieved in a spike anneal, with not more than 1 s dwell time at the peak temperature.

Unwanted diffusion (TED, see section 3.4.1) sets in already at around 750 °C. The ramp-up and ramp-down rates are pushed as high as possible, such that the implant damage causing TED is removed as quickly as possible. However, according to recent results [28, 29], the beneficial effect of fast ramp-up and -down rates is more due to the reduction of the overall thermal budget than to the suppressed TED effect.

For these extremely short anneals, the classical furnace (~1 hr, 800 - 900 °C) is by far not sufficient. Instead, RTP tools[5] use an array of powerful, individually controlled lamps to heat up the wafer. The wafer temperature is constantly monitored with a pyrometer, and the lamp power is adjusted many times per second. With this technique, maximum ramp-up rates of 250 °C/s can be achieved; higher rates are not controllable. The ramp-down rate is typically lower (70 °C/s), because the cooling process is based on radiation and limited by the emissivity of the wafer surface.

Fig. 1.8 shows a typical temperature profile of a 1070 °C, 1s spike anneal, as measured on a SHS 2800 tool by AST (now Mattson). Fig. 1.9 displays the ramp rate vs. temperature.

Figure 1.10 shows SIMS profiles of samples implanted with B, BF_2 and As before and after anneal. The junction depth values, 60 nm (B), 45 nm (BF_2) and 33 nm (As) were determined from the profiles at a concentration of $10^{18}\,cm^{-2}$. These values are typical for all samples prepared for the work described in this thesis, but they vary as a function of implant and anneal conditions. For a detailed description of the SIMS process, see section 2.7, for an assessment of the different characteristics and limitations of a SIMS profile, see section 3.3.2.

Several methods exist to achieve higher ramp rates: For example, in the Levitor (made by ASM), the wafer is put between two hot blocks. Since the distance between wafer surface and the hot block is only 150 μm, the heat is transferred by conduction through an inert gas, and not by radiation or convection. Ramp-rates

[5] An RTP anneal process using an inert ambient is also called Rapid Thermal Annealing (RTA), as opposed to Rapid Thermal Oxidation (RTO) or Nitridation (RTN).

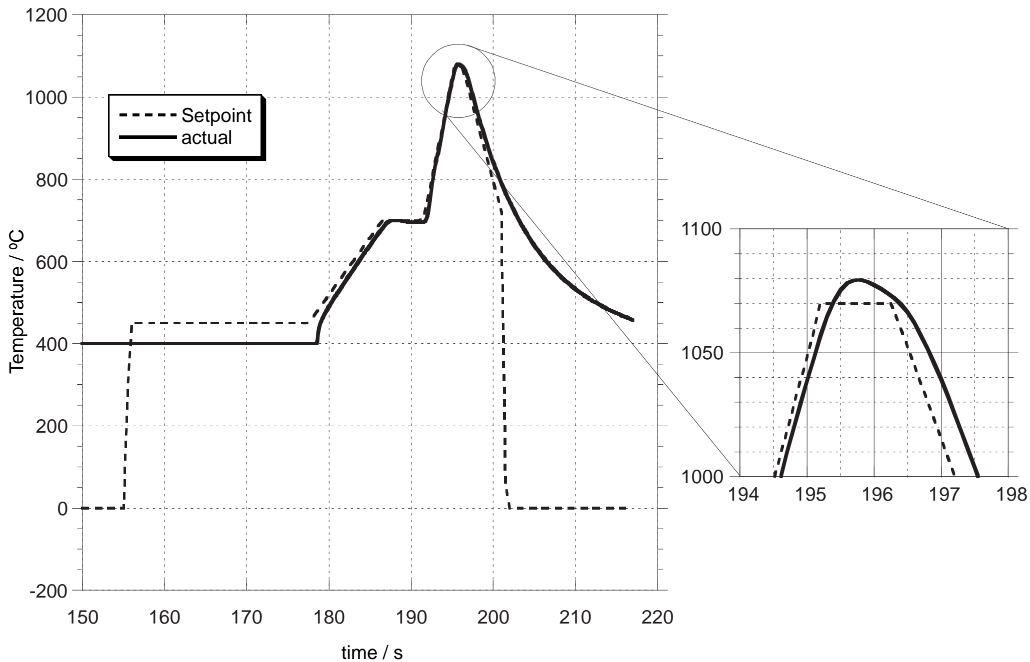

Figure 1.8: Temperature profile of a 1070°C, 1s spike anneal. Displayed are the Setpoint temperature as defined by the recipe, and the actual temperature measured by a pyrometer. The profile includes a 150 s purge of the annealing chamber and two stabilization steps at 400°C and 700°C. The pyrometer is not able to correctly measure temperatures below 450°C

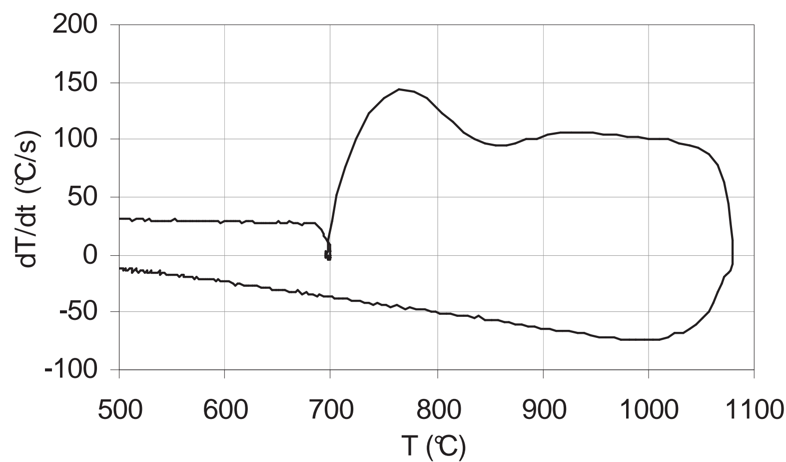

Figure 1.9: Temperature time derivative (ramp rate) vs. temperature for the 1070°C, 1s spike anneal (measured data from pyrometer).

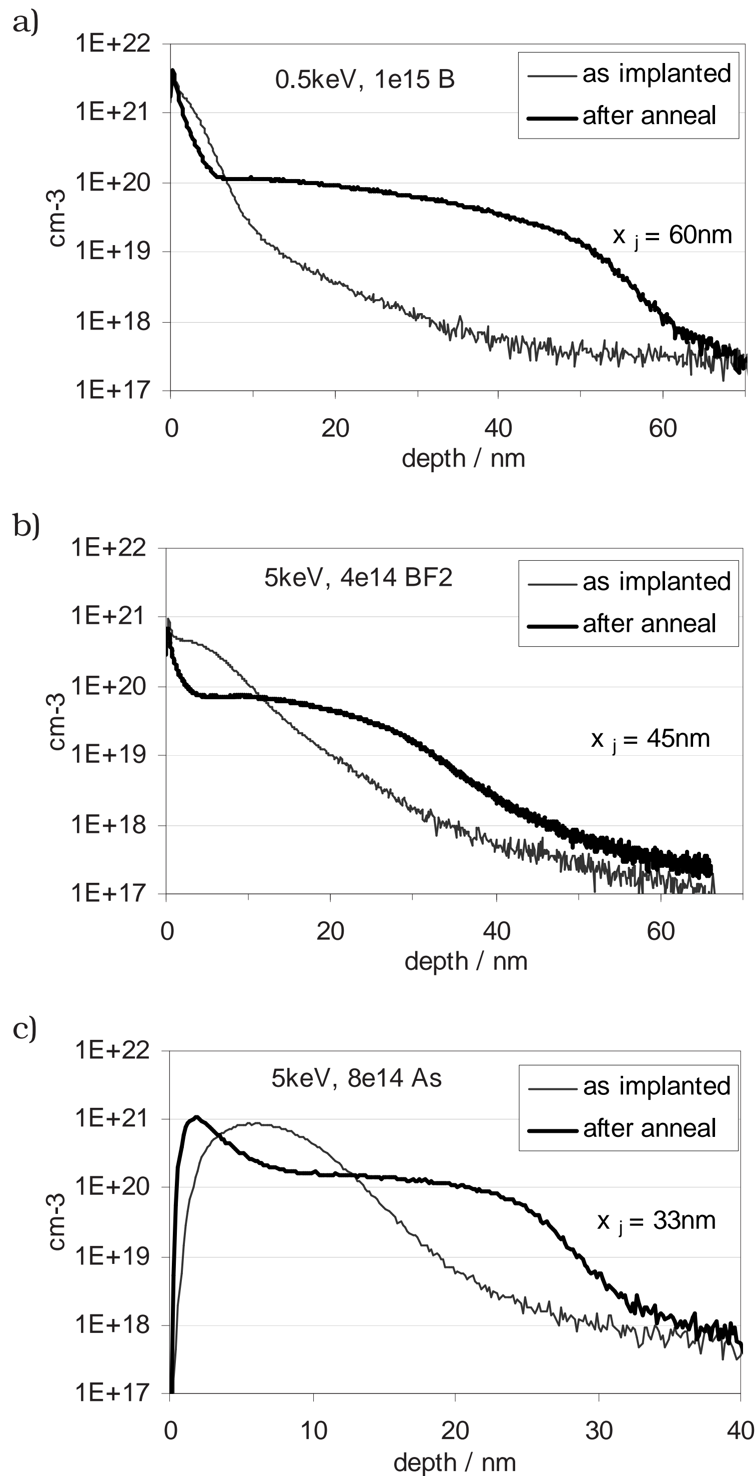

Figure 1.10: Comparison of as-implanted and annealed profiles of B, BF_2 and As implants. In each graph, the junction depth as measured at a concentration of 10^{18} cm^{-2} is given. (Samples from lots E010731, P020144 and T020208)

of up to 900 °C/s can be achieved. In combination with co-implants this leads to improved junction performance [30, 31].

In a hot-walled RTP tool (e.g. by Axcelis), a vertical tube is kept at different temperatures. The wafer is inserted in a low-temperature area and then moved on a stage towards zones of higher temperatures. This technique does not significantly affect the ramp rate, but improves the temperature control as well as the uniformity of the anneal [32].

To avoid loss of the surface oxide during anneal, a little amount of oxygen is added to the annealing ambient. The oxygen is supposed to adsorb to the surface to prevent the desorption of SiO_2 [33]. The anneals in this work were done in an N_2 ambient with 133ppm of oxygen, unless mentioned. However, this also causes a slight oxidation during anneal.

1.4.3 Novel ultra-rapid annealing concepts (Flash, LTA)

Several novel concepts exist to overcome the inherent limits of the RTP spike anneal approach. Whereas a lamp-based RTP tool heats the entire wafer uniformly to the respective temperature, LTA and Flash anneals only heat the surface of the wafer, using a high-intensity light pulse. This process is so fast that a feedback for power control is not possible. The pulse energy must be chosen beforehand, which requires a good knowledge of the optical properties of the wafer's surface. Due to the slow thermal response of the wafer, only the surface is heated. The cooling is mainly controlled by thermal conduction towards the bulk silicon, which assures a very steep ramp down. This helps to reduce the thermal budget and to avoid diffusion.

The Canadian company Vortek Industries Ltd. develops a "Flash Annealing" tool that heats the wafer to an initial temperature (e.g. 700 °C) as for a conventional, low-temperature spike anneal. When the temperature is reached, a single arc lamp pulse heats the wafer within some 0.5 milliseconds up to the final anneal temperature. Compared to an RTP spike anneal, Flash annealing offers a further reduction of the thermal budget and consequently a significant improvement in the junction performance. However, it requires a careful calibration of the pulse energy for every single anneal step [34]. A similar approach is pursued by Applied Materials under the name "Dynamic Surface Annealing" (DSA).

For full-melt Laser Thermal Annealing (LTA), typically the wafer's surface is amorphized (e.g. by a Ge implant) prior to the dopant implantation. A laser beam is then stepped over the wafer, heating up the surface within nanoseconds to a

temperature of around 1300 °C. Since the melting point of amorphous silicon is about 200 °C lower than the one of crystalline silicon (1410 °C), only the amorphized areas are molten. In the liquid silicon, a high diffusivity assures a homogenous distribution of the dopants. Since the temperature is still below the melting point of c-Si, immediately a recrystallization process sets in, with the bulk substrate as nucleation layer. This process is too fast to allow a thermal equilibrium to be reached, so that very high active dopant concentrations are possible in the resulting crystal [35]. The bulk wafer, being a highly effective thermal sink, rapidly cools the top layer down, such that almost no normal diffusion or activation occurs.

The profile of LTA junctions is easily controllable (by choosing an appropriate amorphous depth) down to junction depths <10nm. Also the abruptness is excellent, since in the recrystallized part, a quite homogenous and highly active dopant concentration is found. Beyond the junction depth, some dopants are present, but they are not activated. As seen in Fig. 1.6, LTA would perfectly fulfill any requirements defined in today's ITRS.

The integration of LTA, however, poses major problems: Amongst many issues, any oxide layer present on the surface is transparent for the laser light, so that the silicon below is molten. This can lead to delamination of the oxide layer. Also, a transistor's gate stack is thermally better insulated than the surrounding junction. Thus it reaches higher temperatures, which can cause deformation. Furthermore, different types of structures on the wafer can lead to variations in the absorption of the laser light and thus to a non-uniform anneal. One way to overcome most of these problems is to deposit a metal absorption layer. However, the removal of this layer is very difficult. Due to theses issues, LTA is still far from being an alternative annealing method in CMOS processing, even though it has been investigated for many years.

1.4.4 Solid phase epitaxial regrowth (SPER)

For SPER, a pre-amorphization implant (e.g. Ge) is used to create an amorphous layer (a-Si) with a thickness of 10-20 nm. Then, the dopant and possibly co-dopants are implanted. For heavy dopants (e.g. As), the dopant implantation itself is amorphizing the wafer surface, therefore no separate amorphization implant is necessary.

The amorphized layer is regrown during a low-temperature anneal (e.g. 650 °C, 1 min). Starting with the bulk silicon as a seed layer, the a-Si layer is recrystallized

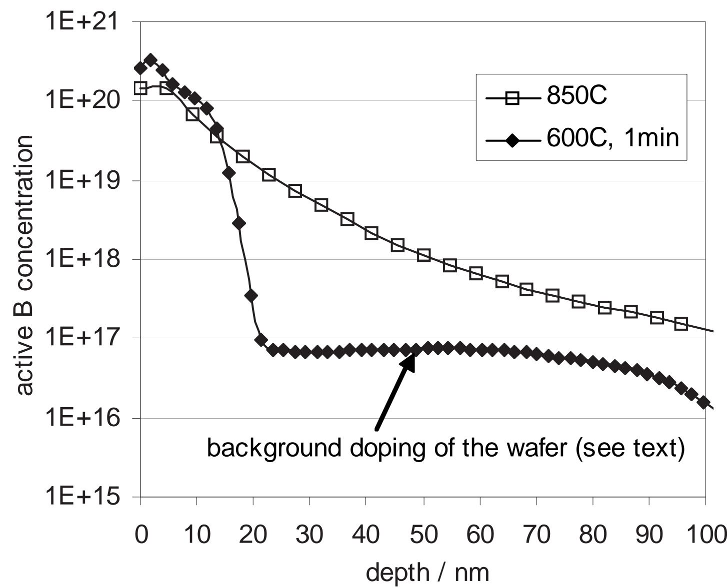

Figure 1.11: SRP profiles (active dopant concentration) of two SPER samples, annealed at 600°C or 850°C, resp. (Courtesy of B. Pawlak)

with a speed in the order of 1-10 nm/s, depending on the temperature. The a-Si/c-Si interface moves towards the surface, and the dopants are quenched into the lattice structure. Since the process doesn't constitute a thermal equilibrium, active dopant concentrations above the solid solubility limit can be reached [36]. At the low regrowth temperature, only little diffusion (TED) takes place, and the dopant profile remains very similar to the as-implanted shape. The junction depth is determined by the thickness of the amorphous layer. The dopant atoms beyond the pre-amorphization depth are neither diffused nor activated, therefore they do not contribute to the electrical junction profile. This leads to an excellent junction abruptness, while the junction depth can be easily controlled by adjusting the pre-amorphization depth [37].

For boron, the co-doping of fluorine reduces the diffusion in the c-Si, but enhances diffusion in the amorphous layer. This effect helps not to activate the tail of the implant, as well as to homogenize the dopant distribution in the junction. Recently, this has been attributed to the effect that the presence of F is retarding the regrowth [38]. Therefore, the dopants in the a-Si layer have more time to diffuse.

The main disadvantage of SPER is the amount of end-of-range (EOR) defects present. The fraction of the implant damage (e.g. dopant or Si interstitials, clusters, vacancies. . .) that lies beyond the amorphized layer is not completely healed

in the annealing step. As a consequence, a high defect density is found just below the metallurgical junction, still within the depletion region, causing increased leakage. Also, during subsequent anneals at higher temperatures, these defects enhance dopant deactivation in the junction [39].

While still many integration issues need to be addressed, SPER seems to be a promising candidate for the fabrication of junctions in the 45 nm to 32 nm technology nodes and beyond.

Figure 1.11 shows SRP profiles of two SPER samples. They were pre-amorphized to a depth of around 20 nm and then recrystallized for 1 minute at 600°C or 850°C, respectively. The SRP profile shows only the active dopant concentration, inactive dopants are not seen. The 650°C sample exhibits a box-like profile with a very steep gradient in active concentration at a depth of 16-20 nm. Close to the surface, very high active concentrations are observed, while the tail of the profile (below 20 nm) is not activated at all. The $10^{17}\,\mathrm{cm}^{-3}$ level of active dopants seen in the profile is the background doping of the silicon substrate used. It is not relevant for this experiment. (This is the "probe junction", described in section 2.3.1.)

For the sample annealed at 850°C, on the contrary, thermodynamical activation has taken place, similar to an RTP anneal. The profile's tail is activated, and additionally the profile depth has been increased due to normal diffusion and TED. Also, the peak active concentration close to the surface is lower than in the 600°C case.

1.4.5 In-situ deposition

For in-situ deposition, openings are etched in the wafer and filled with pre-doped silicon by CVD or PVD. This allows optimal control over the doping profile, resulting in very good junctions. However, there are unsolved issues with masking, faceting under the gate, linewidth dependence on activation, maximum activation with selective deposition, and deposition rate [33].

2 Characterization methods

2.1 Introduction

As described in chapter 1, the accelerating progress from one technology node to the next requires tremendous efforts to improve the silicon processing techniques. In parallel, a continuous improvement of the available analysis methods is needed, to fulfil the increasing requirements for spacial resolution, measurement range or sensitivity.

This chapter will give an introduction into the various characterization methods that are used to analyze ultrashallow doping profiles. The first methods discussed (FPP, SRP, PS and Hall) are used to determine the sheet resistance or resistivity of the material, and give information about the active dopant concentration. The following methods (SIMS, ERD, RBS, MEIS and XPS) are mainly used to determine elemental concentrations, without giving information about the electrical activity of the dopants. Finally, TEM and Ellipsometry are presented as means of determining thickness and quality of a layer. TEM also has the capability of measuring a dopant concentration profile.

2.2 Definition of the sheet resistance

The resistance of the implanted junctions is discussed here in the form of a sheet resistance R_s, which is defined as

$$R_s = \frac{\rho}{t} \tag{2.1}$$

with the material's resistivity ρ and thickness t of the layer used. Its value is given in units of $\Omega/\square$ (Ohms per square). This unit, mathematically identical to Ω, is used in order to distinguish R_s from a resistance, given in Ohms.

To explain the origin of the denomination "ohms per square", we consider a wire with length l and width w (as printed by lithography), and thickness t (here the junction depth). The surface of this wire can be divided in a identical squares

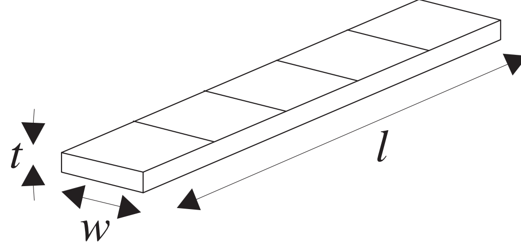

Figure 2.1: Wire with 5 squares (a=5), to explain the sheet resistance concept (see text).

$l = a{\cdot}w$ (fig. 2.1). The number of squares, a, is a characteristic geometry factor that can be freely chosen when designing the wire. The resistance can be written as

$$R = \frac{\rho \cdot l}{w \cdot t} = a \cdot \frac{\rho}{t} = a \cdot R_s \quad \Rightarrow \quad R_s = \frac{R}{a} \tag{2.2}$$

The sheet resistance turns out to be the resistance of each square, hence the unit "ohms per square". The wire's resistance can be written as a function of only sheet resistance and the geometry factor a [40]. This also implies that for an actual wire of length l and width w, R is invariant to scaling ($l \to c{\cdot}l$, $w \to c{\cdot}w$), as long as the scaling doesn't affect the value of the sheet resistance R_s.

Equation (2.1) is given here for a box-like profile, i.e. a junction in the form of a homogenously doped layer with uniform resistivity. In reality, the resistivity varies as a function of doping concentration over the depth of the implanted profile. As a consequence, (2.1) becomes more complex, but the conclusion drawn in (2.2) is still correct:

$$\frac{1}{R_s} = \int \frac{dx}{\rho(x)} \quad \Rightarrow \quad \ldots \quad \Rightarrow \quad R_s = \frac{R}{a} \tag{2.3}$$

The advantage of the sheet resistance concept is that if R_s is known, one can calculate the electrical properties of a layer without knowing the exact profile shape and thickness. R_s is therefore a property of a *layer*, not of a *material*.

2.3 FPP (Four Point Probe)

2.3.1 In-line tip arrangement

To measure the sheet resistance of an implanted layer, normally the Four Point Probe method (FPP, also 4PP) is used. Four tips touch the surface of the wafer with little force (20-100g per tip). The tips are arranged in one line, with equal spacing.

Tool	Tencor RS75	SSM-240	SSM-150
Type	4 points	4 points	2 points (PS or SRP)
Tip separation	500 μm	500 μm	25-1000 μm
Load per tip	100 g	20 g	5 g
Imprint depth	130 nm	30 nm	5 nm
Contact radius	15.0 μm	2.3 μm	1.7 μm

Table 2.1: Some core data of the main tools to measure sheet resistance available at IMEC [41].

While applying a current I through the outer two contacts, the potential difference V between the inner two contacts is measured. Measurements with different currents in both directions are combined to improve the precision. This configuration avoids any problems with the unknown contact or spreading resistance. The sheet resistance is calculated as

$$R_s = \frac{\pi}{\ln 2} \cdot \frac{V}{I}. \tag{2.4}$$

This formula is correct if the probe tip separation is much larger than the thickness of the investigated layer [40]. Also, the sample must be sufficiently large: The distance between the measurement spot and the closest edge of the sample should be at least five times the probe tip separation. Table 2.1 shows some core data for the tools available in IMEC.

While FPP provides a very easy and fast method to measure sheet resistance (e.g. for a quick confirmation that implant and anneal were done correctly), several limitations need to be considered: To establish a good electrical contact, the tips need to be pressed to the surface with a certain force. If their penetration into the surface layer is bigger than the junction depth, they can short-circuit the implanted junction. In this case, a parallel electric current through the substrate causes an underestimation of the junction's sheet resistance. The sheet resistance of the wafers used in this work was around 200 $\Omega/\square$, only slightly less than the typical R_s of the junctions (300-600 $\Omega/\square$)

In all of the investigated samples, the electrical junction depth is deeper than the metallurgical p/n-transition, which helps to avoid probing the substrate [41–43]. This shift is caused by the carrier spilling effect, which is observed at junctions between a highly and a lowly doped layer (fig. 2.2).

For the junctions used in this work, the SSM-240 tool provided a sufficiently low probe tip pressure. For shallower junctions a special "probe junction" can be implanted: Before the USJ implantation, a very deep and lowly doped junction is

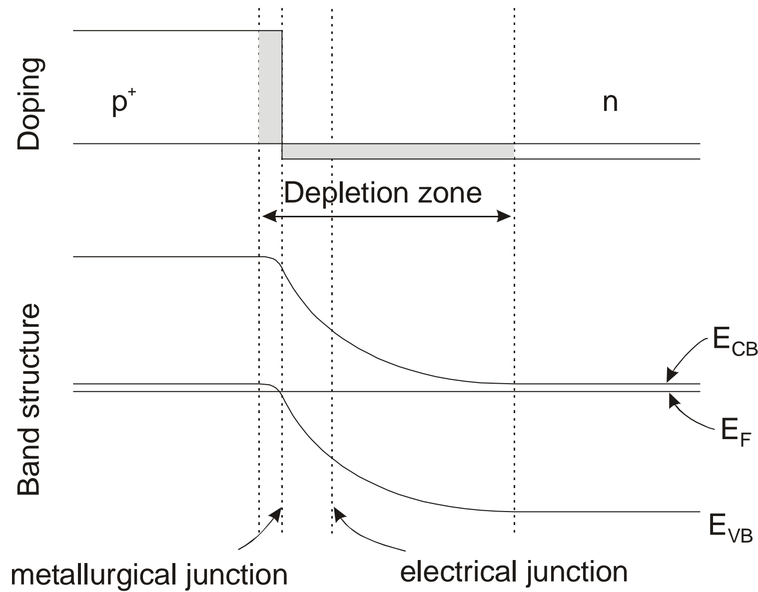

Figure 2.2: Carrier spilling effect: The electrical junction depth is defined as the point where the Fermi level is in the center of the band gap, whereas the metallurgical junction is the reversal point of the bad edge's curvature. At a junction between a lowly and a highly doped material, the the poisson equation $\partial^2 V/\partial x^2 = -\rho/(\varepsilon_0 \varepsilon_r)$ *results in a shift of the electrical junction towards the lowly doped layer. [44]*

implanted and annealed. Like this, the actual junction depth is pushed deep into the wafer, but the doping of the probe junction is so low that it doesn't contribute significantly to the measured sheet resistance.

2.3.2 Van-der-Pauw structure

As an alternative to the in-line tip arrangement discussed above, a van-der-Pauw structure can be used to measure a sheet resistance [45]. This requires a square sample. The sample is either cleaved (with a typical size of a few cm), or can be structured by lithography and etching of the surrounding material (typical size: 100 μm). The four contacts are placed in the corners or in the middle of the edges (see fig. 2.3). A current is applied through two adjacent contacts, the voltage is measured on the two opposite contacts. Four different contacting schemes are possible, from which individual resistance values can be calculated:

$$R_{1243} = \frac{I_{12}}{V_{43}}, \quad R_{4312} = \frac{I_{43}}{V_{12}}, \qquad R_{1423} = \frac{I_{14}}{V_{23}}, \quad R_{2314} = \frac{I_{23}}{V_{14}}, \tag{2.5}$$

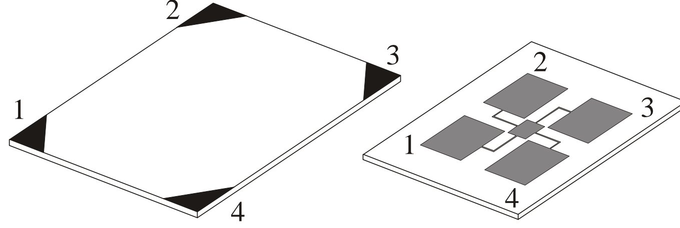

Figure 2.3: Van-der-Pauw structure, to measure sheet resistance or Hall effect. Left: Cleaved sample from a blanket wafer. Right: Patterned structure with four large contact pads around a small active area in the middle.

$$R_A = \frac{R_{1243} + R_{4312}}{2} \qquad R_B = \frac{R_{1423} + R_{2314}}{2} \tag{2.6}$$

R_A and R_B are two characteristic resistances that are related to R_s by the van-der-Pauw equation

$$e^{-\pi R_A/R_s} + e^{-\pi R_B/R_s} = 1, \tag{2.7}$$

which can be solved numerically for R_s.

In order to obtain good results, the measured sample should be as close to a square as possible, and the contact points should be located very close to the corners. The better the geometry, the closer the two resistances R_A and R_B. For manually cleaved samples on blanket (i.e. non patterned) wafers, this is difficult to achieve. Therefore, the in-line FPP measurement on larger samples is easier and more precise. For patterned wafers, however, on which no large blanket areas are available, the van-der-Pauw technique offers a possibility to measure sheet resistance on very small areas. Many test wafer mask sets include these structures.

2.4 SRP (Spreading Resistance Profiling)

For SRP, a corner of the sample is carefully bevelled to an angle of down to 7’ (0.12°). Along the bevelled surface, the vertical profile of the sample can be observed, with a one-dimensional magnification of a factor of up to 500.

SRP uses a tool with only two probe tips (at 25 μm separation) between which the resistance is measured. The measurements are done at 1 μm distance, corresponding to around 2nm in depth. The spreading resistance technique is based on the fact that the largest contribution to the measured resistance is caused by the immediate surroundings of both tips. If the contact geometry is know sufficiently well, an integral, but weighted sheet resistance information about the material

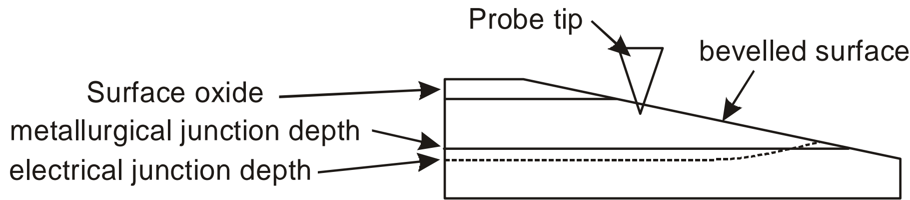

Figure 2.4: SRP measurement principle. Due to the carrier spilling effect, the electrical junction depth is not identical to the metallurgical one.

under the tips is measured [46]. After applying a variety of correction models, a depth-resistivity profile is calculated, which finally leads to a depth profile of the active concentration.

The starting point of the profile is normally determined by manual inspection of the sample through a microscope, where the bevelling edge can easily be seen.

The accuracy of the obtained profiles depends on many factors: While the theoretical models are well known at least for implanted silicon samples, physical uncertainties like surface damage after bevelling or worn probe tips can significantly influence the profile's quality. Furthermore, since the measured signal is always an integral signal over all the material below the measurement point, the calculation quality of every point depends on how well the profile was determined below that point. Measurement errors are accumulated. This procedure also limits the depth resolution to around 5 nm per decade in active concentration; sharper changes cannot be resolved.

Since the surface material is removed during bevelling, no carrier spilling occurs around the exposed metallurgical junction. On the contrary, a slight reverse carrier spilling effect is observed (see fig. 2.4). Therefore, the junction depth obtained from SRP profiles is typically shallower than the metallurgical one, as seen in SIMS. A comparison between SRP and SIMS profiles is given in fig. 4.3 on page 66.

The imprint depth of the SRP probe tips is around 5 nm, much deeper than the 2 nm oxide that was mainly used within this work. This comparison shows that SRP cannot be very accurate within the first nanometers from the surface. To improve the quality it is possible to deposit an oxide of several 100 nm thickness on the sample prior to bevelling. With this method, the profile's starting point can be easily determined from the electrical data as the first point with a nonzero conductivity. This results in a better resolution close to the surface, and also allows a better depth calibration.

The Nanoprofiler [47] is an IMEC project that aims to extend the resolution of SRP to below 1 nm. The probe tips, shaped by surface micomachining, will be made of silicon with a protective coating and have a curvature radius of 20 nm. Similar tips can also be used for the two-dimensional methods SSRM (Scanning Spreading Resistance microscopy) or SCM (Scanning capacitance microscopy). For these methods, a thin vertical cross section is scanned with a probe tip. SSRM measures the resistance between the tip and the sample to evaluate the local conductivity [46], whereas SCM investigates the local carrier density from the capacitance between the tip and the sample.

An overview on SRP and other 1D and 2D profiling techniques that use microscopic tips to probe cross section or bevelled surfaces is given in various publications [43, 47–50]. An quantitative standard for the relation between doping concentration and resistivity for B and P doped bulk silicon is published in [51].

2.5 PS (Probe Spacing)

Probe Spacing was not used for this work, but will be briefly described as an alternative to the FPP method. Using the same tool as for SRP, the resistance is measured between two probe tips. In this configuration, the resistance is not only due to the material's sheet resistance, but also influenced by the contact resistance.

To compensate for the contact resistance influence, the measurement is repeated with different tip spacings, typically between 50 and 1000 μm. The function

$$R = c_0 \cdot \ln(spacing) + c_1 \tag{2.8}$$

is fitted to the results. The sheet resistance can then be calculated as

$$R_s = \pi \cdot c_0 \tag{2.9}$$

Since the PS tool is optimized for spreading resistance measurements, it uses a very small tip load and thus creates extremely shallow imprints of only 5 nm.

2.6 Hall Effect

The Hall effect is based on the fact that moving carriers in a magnetic field experience a force perpendicular to the moving direction. This creates an electric field

which can be measured as a voltage perpendicular to the applied current. The force equilibrium is given by

$$F = q \cdot v \cdot B = q \cdot E \tag{2.10}$$

In the easiest case, a van-der-Pauw structure is used for a measurement (fig. 2.3). A current is applied through contacts 1 and 3, and the Hall voltage V_H is measured perpendicularly, on contacts 2 and 4. For a theoretical sample with exactly perpendicular current (contacts 1-3) and voltage (2-4), V_H is independent of the sample shape. For box-shaped dopant profiles with a depth d and a constant doping concentration n_p or n_e, one finds:

$$V_H = \frac{IB}{dqn_h} - \frac{IB}{dqn_e} \tag{2.11}$$

with the applied current I, the magnetic flux B, the sample thickness (here junction depth) d, the elementary charge q and the hole and electron densities n_h and n_e. It is interesting to note that the sign of the Hall voltage allows to distinguish between p- and n-type samples (see also [52]).

If the carrier concentration is not constant within the conducting layer, equation 2.11 transforms to

$$V_H = \frac{IB}{qD_p} - \frac{IB}{qD_n} \tag{2.12}$$

with the total active dose D_p and D_n.

In case of a non-perfect sample shape, a geometry-depended ohmic voltage drop proportional to I is measured, additional to V_H. To compensate for this effect, the Hall measurement is repeated with different magnetic fields and in all four contacting variations, (analogous to the van-der-Pauw technique).

The main limitation for Hall measurements is the implanted dose: For high carrier concentrations, the ratio V_H/I becomes low. If V_H is similar to the measurement noise of the ohmic voltage, no reliable results can be obtained any more. Reducing the temperature to liquid nitrogen or helium improves the results, if the carrier concentration is high enough to avoid freezing out.

The available Hall measurement tool was able to measure samples with doses below a limit of around $10^{13}\,\mathrm{cm}^{-2}$. Since the samples used for this work had doses around two orders of magnitude higher, no reproducible results could be obtained, thus Hall measurements were not used for sample evaluation.

2.7 SIMS (Secondary Ion Mass Spectrometry)

SIMS is a method for one-dimensional chemical depth profiling. Its properties for analyzing silicon samples have been studied extensively, and are well understood [53].

A beam of primary ions is used to sputter atoms or molecules from the sample surface. The secondary ions are analyzed in a mass spectrometer. If the sputter yield (sputtered atoms or molecules per incident ion), the ionization yield (fraction of all secondary particles that are ionized) and a few more parameters are sufficiently well known, one can calculate the concentration of any impurity in the silicon as a function of depth. To calibrate the depth scale, the depth of the sputter crater is measured after the analysis is finished.

In order to obtain a signal with good depth resolution, the sputter beam is scanned across an area of typically 100x100 μm^2. The resulting crater has a square shape with a flat bottom in the middle. At the border of the crater, the beam (having a gaussian shape) sputters not only material from the bottom of the crater, but also from its side walls. This leads to a mixing of material from different layers. To avoid this, an electronic window selects only sputtered ions from the middle of the crater for analysis (between 6% and 25% of the total crater surface). Due to the large crater size, SIMS can only be used for analyzing very large structures or unpatterned wafers.

Several issues limit the resolution of SIMS: A small fraction of the material sputtered from the crater's side walls is redeposited in the middle of the crater, leading to a noise signal, typically 3-5 orders of magnitude below the concentration at the point from where the material originated. The result is a finite "dynamic range" between 3 and 5 orders of magnitude. The value depends on many parameters like sputter beam energy, crater size, size of the electronic window etc.

The sputtering process not only removes material, but also mixes the atoms close to the surface, leading to a degradation of the depth resolution. To avoid this effect, one can reduce the beam energy. The disadvantage of low energies is a superproportional reduction of the sputter yield and hence an increase in sputtering time [54]. Also, with low energies, beam control becomes more difficult, resulting in a reduced dynamic range. The lowest useful energy of an O_2^+-beam is around 500 eV. At lower energies, only a slight improvement in resolution is obtained at the cost of extremely long measurement times (10 hours or more).

Another important problem are matrix effects at the surface or at interfaces between different material layers: The sputtering characteristics can change sud-

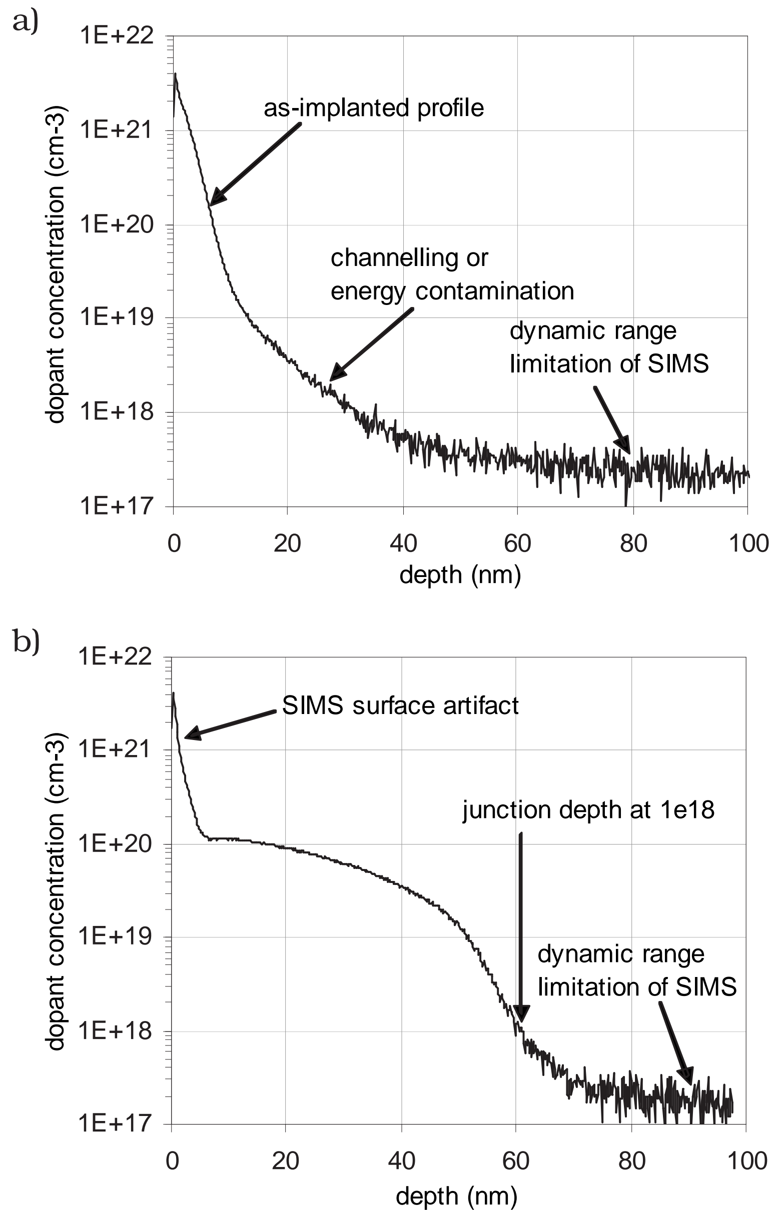

Figure 2.5: SIMS profiles of a typical 0.5 keV, 1e15 B implant. a) As-implanted profile. b) Sample annealed in a 1070°C, 1s spike anneal. (Samples from lot T020208)

denly, resulting in a difficult depth and concentration calibration. To avoid this, most of the samples in this work were analyzed using an O_2^+ beam [55, 56]. The oxygen atoms first oxidize the surface and then sputter it. Hence, for a silicon sample with a surface oxide (this was the case for all samples of this work), the sputtered material is uniquely oxide throughout the entire analyzed depth. Therefore, the sputtering conditions don't change during the analysis, which almost avoids any matrix effects at Si/SiO_2 interfaces.

For precise depth calibration, one needs of course to take into account that the effective sputter rate for silicon is by a factor 0.44 lower than for oxide, since 1 nm of oxide contain as many Si atoms as 0.44 nm of silicon. This correction is difficult to apply in a physically correct way, especially for thin oxides of 1-2 nm. In order to avoid any deterioration of the information contained in the SIMS profiles, the correction was not done for the profiles presented in this work. The presence of dopants in the material is no problem for SIMS, since the typical concentrations are always below a few percent.

One of the unsolved problems of SIMS is a dopant peak seen at the surface of each sample. It is probably created by a rearrangement of atoms close to the surface during the initial sputtering, where no equilibrium has been reached yet [57]. The height and dose of this peak are, however, not yet well understood. The effect will be discussed in more detail in section 3.3.2.

With the exception of this surface artifact, the SIMS process is very well understood for the Si-SiO_2 system. Dopant profiles with a depth resolution of around 1 nm are possible under optimal conditions, and a concentration accuracy of around 5% (B) or 10% (As) can be achieved. SIMS tools are commercially available and widely spread as a standard profiling method. During this work, SIMS was mainly used to obtain information about the distribution of the dopants over the total junction depth and about the total dose.

2.8 High-resolution ERD and RBS

Elastic Recoil Detection and Rutherford Backscattering are methods with an very high depth resolution in regions close to the surface. This advantage is obtained at the cost of relatively low sensitivity (data are only good for high dopant concentrations above $10^{20}cm^{-3}$) and by a very costly measurement process. The ERD and RBS results for this work were obtained by the group of G. Dollinger at the Munich University of Technology [58–60].

2.8.1 ERD (Elastic Recoil Detection)

For the results cited in this work, ERD was used to determine B, N and O profiles. 40 MeV Au^+ ions were used as primary beam. The ions hit the target under a low angle of 4°. Elastically recoiled secondary ions that leave the sample under an angle of 11±3° are analyzed (15±3° recoil angle). While travelling through the material, the primary as well as the secondary ions lose energy. The final energy of the secondary ions is measured, and the original depth of the secondary ion is calculated from the energy lost. To optimize the ratio of actual depth to energy loss, low incident and analysis angles are used. With this technique, a resolution down to individual atomic layers has been achieved in special cases [60, 61].

The angle of the primary ion beam and the acceptance angle of the analyzing magnet and detector are also optimized to facilitate the measurement setup. It is possible to focus the secondary ions to a detector in such a way that every channel of the detector corresponds to a certain originating depth of the measured ion. The elemental identification of the secondary ions is done by tuning the magnet to a certain charge state of the desired element, and by using a gas-filled $\Delta E/E_{rest}$ detector [59].

To improve the ratio of measurable signal to lattice damage, it is desirable to optimize the recoil cross section. This is again achieved by running the setup with very low incident and recoiled beam angles, as well as by choosing heavy Au ions as primary beam.

For most of the ERD profiles in this work, all measured profiles (oxygen, nitrogen, dopant) are displayed together in one graph. The left scale of each graph corresponds to the O and N concentration, while the dopant profile is aligned to the right scale.

The concentration is mostly given in atomic percent. For silicon, 1 at% corresponds to $5.00 \cdot 10^{20}\,\mathrm{cm}^{-3}$, for oxide to $6.83 \cdot 10^{20}\,\mathrm{cm}^{-3}$. The different conversion factors for oxide and silicon introduce a small artificial step at the Si/SiO_2 interfaces. This step is seen in most profiles presented in this work, when given in units of cm^{-3}. One could avoid this step by using a smoothing algorithm that takes the resolution of ERD and a possible interface roughness into account. However, since the interface region is of special interest for this work, we abstain from any smoothing in order not to lose any details of the profile.

ERD is capable of profiling samples with a resolution of 0.6-0.8 nm (FWHM of a delta peak) at the surface. This value, however, deteriorates with depth, because the longer travelling path of the primary and secondary ions through matter in-

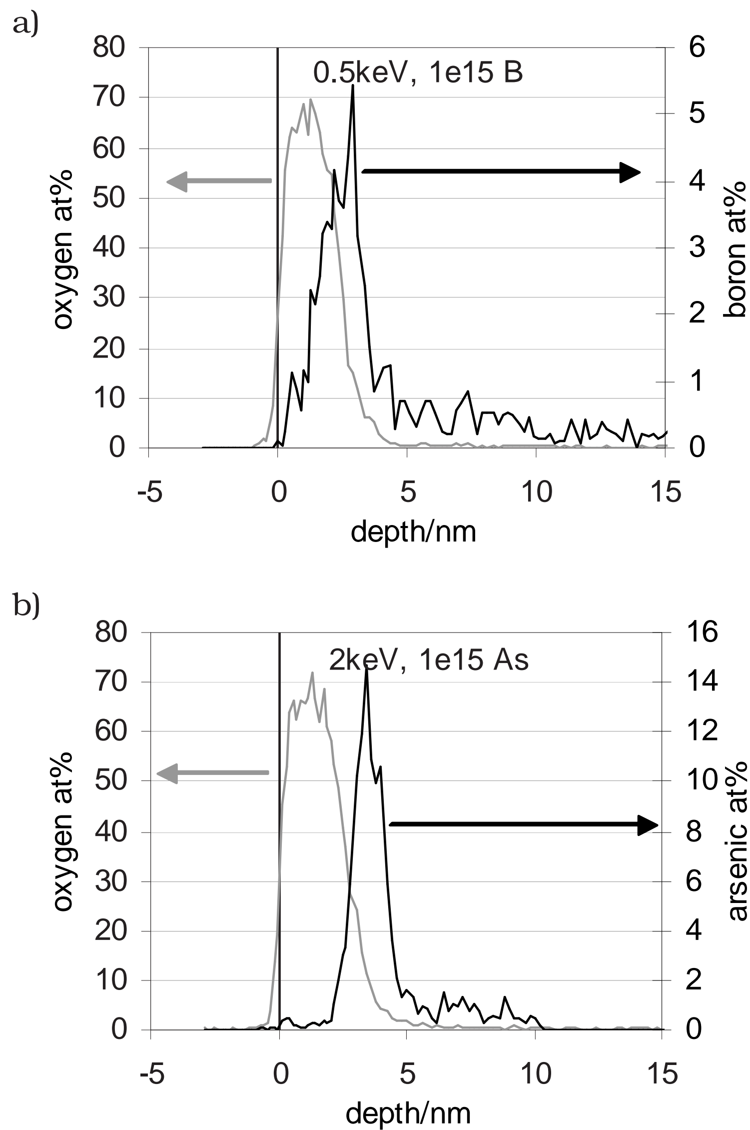

Figure 2.6: ERD/RBS profiles of a B and an As sample, annealed in a 1070°C, 1s spike anneal. The concentration is given in atomic percent (1 at% $\approx 5 \cdot 10^{20}$ cm^{-3} for silicon). (Samples from P020465)

creases the statistical error. This limits the useful profiling depth of ERD to around 10-15 nm.

Furthermore, the depth resolution of the measured profiles is not only limited by the physical constraints of the measurement process, but also by surface adsorbates as well as sample non-uniformity issues like oxide thickness variations and interface roughness. As a result, the depth resolution is limited to values between 1 and 2 nm (FWHM) for most of the investigated samples.

The concentration resolution of ERD is deteriorated by statistical effects due to the quite low number of counts per measurement[1], seen as noise in the profiles. It follows a $\sqrt{n}/n$ law and amounts to values between around ± 0.2 percentage points in the bulk part of the junction up to around ± 1 percentage point in the high-concentration pile-up.

Longer measurements would improve the statistics, however the measurement time is limited by practical reasons (limited available time on the measurement facility, costs etc.) and the sample size: Since the measurement destroys the sample, a large sample area is needed (around 0.5-1 cm^2) for each profile. Furthermore, larger samples would decrease the depth resolution due to processing variations over the sample.

It has to be noted that these limitations are mainly due to the fact that the ERD setup was optimized for high surface resolution, as required for this work. By increasing the primary beam energy and adjusting the setup geometry, ERD can also be used to investigate deeper profiles with a concentration sensitivity down to a few ppm, at the cost of a reduced depth resolution.

2.8.2 RBS (Rutherford Backscattering Spectrometry)

While ERD was used to measure B, N and O profiles, the RBS technique is preferred for heavier elements such as arsenic. The measurement is done with the same accelerator, analyzing magnet and detector setup as for ERD, but using a primary beam of 40 MeV Cu ions. The energy loss during the scattering process is characteristic for the involved target atom species. Again, the Cu ions additionally lose energy while travelling through the material. From the energy of each measured Cu ion, the depth of the scattering event is extracted. The depth resolution of an arsenic RBS profile is typically slightly better than the one of a boron or oxygen ERD profile.

---

[1] A B profile consists typically of less than 1000 counts.

Identical to the ERD setup, RBS uses angles of 4° for the incident beam and 11±3° for the scattered ions. Compared to a setup with angles close to 90°, this forward-scattering RBS has the advantage of a much higher scattering cross section. This improves the trade-off between detected scattering events and sample damage. The disadvantage of the low scattering angle is a worse target atom mass resolution. This method is therefore only useful for detecting atoms with an atomic mass that is sufficiently different from the surrounding matrix.

With RBS it is very difficult to investigate elements that are lighter than the matrix atoms – in this case, boron or oxygen are lighter than silicon. The small signal peaks of B and O are impossible to separate from the overwhelming signal of deeper backscattering at Si atoms.

2.8.3 MEIS (Medium Energy Ion Scattering)

MEIS is a variant of RBS that is capable of profiling light atoms such as B, N or O [62–65]. For the profiling of light ions, two main problems of RBS need to be solved. Firstly, the beam ions must be lighter than the analyzed target atoms for RBS. Therefore, MEIS uses p^+ or He^+, with an energy in the range of 50 to 400 keV. The low energy is necessary to provide a good depth resolution when using these very light primary ions, because the relative energy loss per nanometer in material is higher for low ion speed.

The second problem of RBS is the mentioned difficulty to measure small amounts of atoms lighter than the matrix species. Therefore, MEIS makes use of channelling along lattice axes, which suppresses the bulk Si signal. This makes MEIS an ideal method to investigate lattice damage, or to differentiate between substitutional and interstitial or clustered dopant atoms. However, for ultra-shallow profiling of light elements, ERD is preferred.

2.9 TEM (Transmission Electron Microscopy)

For cross section TEM, the sample needs to be thinned to around 50 nm for electron transparency. To achieve this, a carrier is glued on the top side of the sample; then the sample is cut into thin slices. The slices are first thinned by polishing, then by ion milling: An ion beam sputters the material away. The milling process is stopped as soon as a hole is opened in the center of the sputter crater. The optimal sample thickness of ~50 nm is then found on a ring around the hole.

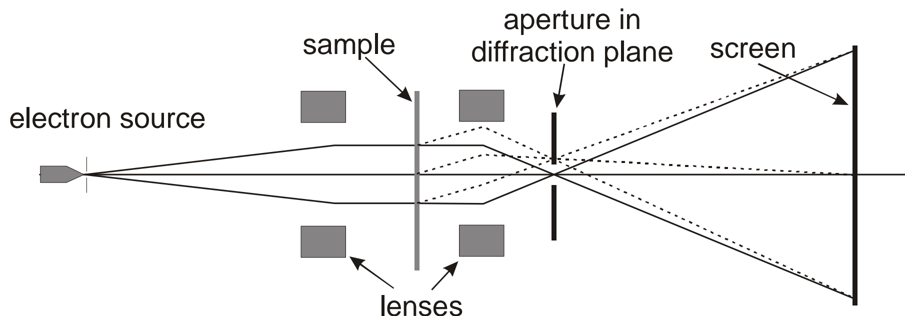

Figure 2.7: TEM microscope

Then the sample illuminated by a parallel electron beam. Using electromagnetic lenses, the transmitted beam is magnified and observed on a fluorescent screen or recorded on film or a CCD camera (“TEM imaging”, see fig. 2.7)

An objective aperture is used to remove all scattered electrons. In the “bright-field” mode, only the transmitted electrons reach on the screen; the aperture is used to remove elastically scattered electrons and limits the angular acceptance of the inelastically scattered electrons. Different scattering probabilities, caused by different density or electronic properties of different materials, are seen as contrast. A contrast at an interface between two materials can, however, also be caused by artificial effects like steps in the sample thickness, if the polishing is not homogenous due to different hardness.

A ring-shaped slit can be used to investigate only scattered electrons (“dark-field” mode). This allows to differentiate between different lattice orientations of individual grains.

Alternatively, it is possible to place the screen in the diffraction plane, where normally the aperture is placed. A diffraction pattern is seen on the screen that gives information on the lattice parameters of the investigated sample, averaged over the total illuminated area (“diffraction mode”).

In a high-resolution TEM image, individual lattice atom columns are discernible, if the sample is properly oriented with one lattice axis parallel to the electron beam (normally in a <110> direction). Information about the lattice orientation and quality is obtained, individual lattice dislocations can be seen. Different materials can be distinguished by lattice size, density and quality (e.g. crystalline–amorphous).

In an STEM (Scanning TEM), the wide beam illumination is replaced by a highly focused beam which is scanned across the sample. In the diffraction mode, this allows the investigation of diffraction patterns with high spatial resolution. When

using the imaging mode, the screen and camera can be replaced by an electron detector. The image then is calculated electronically [66]. The digital capture allows a higher magnification, but typically lower resolution, compared to a normal TEM setup.

Several TEM modifications exist for additional material characterization:

EFTEM (Energy filtered TEM) – An energy filter is placed before the screen of a TEM or STEM in imaging mode. The filter allows only those electrons that have lost a certain energy to pass. By choosing an element-specific energy loss value, a density map of a specific element is produced. The background can be determined by measuring two more maps while filtering to slightly lower or higher energy. EFTEM can be used on a TEM or STEM system.

EELS (Electron Energy Loss Spectroscopy) – The energy distribution of electrons passing a specific point of the sample is analyzed. From characteristic peaks, the elements present in the sample can be determined.

EDX (Energy Dispersive X-ray) – Secondary X-rays are analyzed for energies characteristic for specific elements. When using an EDX setup on a SEM, the spacial resolution is deteriorated by scattering of the primary electrons into a large volume of the sample. TEM-EDX avoids this problem, since the sample is only a few tens of nanometers thick.

While TEM has the advantage of an extremely high spatial resolution, EDX and EELS can only detect relatively high dopant concentrations above 1% ($5 \cdot 10^{20}\,\mathrm{cm}^{-3}$). This is due to the very thin sample: At a thickness of 50 nm, a concentration of 1% corresponds to only a few dopant atoms per lattice column.

For this work, cross-section TEM and STEM imaging were used to investigate the quality of the interfacial oxide. EDX was used to determine the dopant concentration at a series of points across the interface.

2.10 XPS (X-Ray Photoelectron Spectrometry)

XPS illuminates the sample with X-rays and analyzes the secondary electrons created by the photo effect. For a given monochromatic X-ray source, the energy of the electrons is determined only by their binding energy in the crystal. The intensity of a peak in the spectrum is a function of the number of atoms of any element and of their depth profile – the escape probability decreases with depth.

In this work, XPS was only used to determine the thickness of the surface oxide by investigating the silicon 2p electrons. Their binding energy is 99 eV for pure (bulk-)silicon and 103 eV for perfect SiO_2. Depending on the oxide quality,

two more intermediate oxidation states can be observed. The oxide thickness and quality can be calculated by comparing the areas under the different peaks. The thicker the oxide, the higher the peak at 103 eV, but also the lower the 99 eV peak, because the oxide attenuates the pure-Si signal.

The intermediate oxidation states are normally observed as shoulders of the two main peaks. A clear difference between SiO_2 and SiO_x can be made only for $x < 1.6$.

The escape depth for SiO_2 is around 2.2 nm, therefore the precision is best for oxide thicknesses of less than ~5 nm. During this work, it has been shown that for samples without any oxide (“0.00 nm”), the result is accurate to about 0.1 monolayer of oxide. For this work, an accuracy of around 0.02 nm is assumed for an oxide layer thickness of less than 1 nm.

2.11 Ellipsometry

Ellipsometry is an optical technique to determine the thickness and refractive index of a film. Under different angles and polarization directions, a laser beam is reflected from the sample. Interference between the reflections from the surface and the bottom interfaces of the film causes a change in polarization from linear to elliptic, hence the name.

The reflectivity for light polarized parallel (p) and perpendicular (s) to the incident plane are measured. The change in polarization during reflection can be expressed as two angles, Ψ and Δ:

$$\tan(\Psi)e^{i\Delta} = \frac{R_p}{R_s} \tag{2.13}$$

Since only the ratio of R_p and R_s is needed, the measurement can be done with high accuracy. The fraction R_p/R_s is minimal if the angle of incidence is the Brewster angle. From the definition of the Brewster angle,

$$\tan \Phi_B = \frac{n_{film}}{n_{air}}, \tag{2.14}$$

the refractive index of the the reflecting material is calculated [67]. If one or more thin films are deposited on a silicon substrate, the original silicon Brewster angle is modified by the film(s). From the resulting “pseudo-Brewster angle” and the reflection properties over the whole range of incident angles, the refractive index and thickness of the film(s) can be determined independently [68].

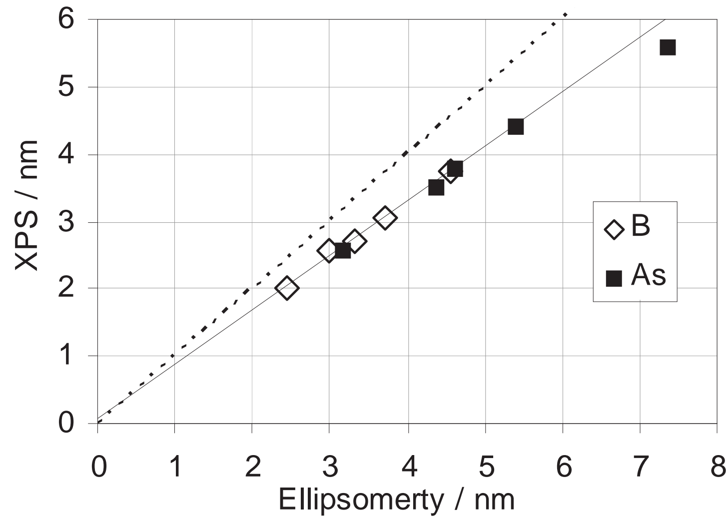

Figure 2.8: Comparison of ellipsometry and XPS results when measuring oxide thickness on implanted and annealed samples. (Lot P020465 and P030131)

For this work, ellipsometry was used to determine the thickness of a surface oxide. With the available tool (Tencor ASET F5), reliable results can be obtained for an oxide thickness above ∼5 nm, while the measured samples had mostly an oxide in the range of 1-3 nm. To improve the thickness measurement quality, the refractive index of the oxide was predefined to a value of 1.46. This is not always precise for the thin oxides used, however it improves the measurement repeatability. Various side effects like a damaged or doped oxide typically increase the measured value. For example, directly after implant, a 2 nm oxide is measured to anything between 5 and 10 nm. However, the ellipsometry results are well reproducible, and a comparison of different samples with presumably the same oxide quality is easily possible.

Figure 2.8 compares ellipsometry to XPS results for samples implanted with 0.5 keV, 1e15 B or 2keV, 1e15 As and annealed in a 1070 °C spike. Due to different oxygen content in the annealing ambient (133ppm to 100%), different amounts of oxide have grown during anneal. Apart from the As sample with the thickest oxide, all ellipsometry values are about 20% too high, when compared to XPS data as a high-precision reference. However, this depends strongly on the implant and anneal conditions and should rather be regarded as an estimation than as a precise result.

3 *Pile-up fundamentals*

3.1 Initial experiments

The initial approach to the topic of this thesis was to investigate the influence of the presence of a screening oxide during implant and/or anneal, mainly using sheet resistance measurements. During a couple of initial experiments, five main types of samples were investigated:

OO An oxide was grown in the beginning, such that it was present during implant and anneal.

O- An initially grown oxide is removed by an HF dip after implant, and the anneal is done without screening oxide.

-O The oxide is grown only after implant, such that it is present only during anneal

- - No oxide is grown at all, both implant and anneal are done without screening oxide.

-O- An oxide is grown after implant an removed immediately, to separate the influence of the oxidation from the one of oxide presence during anneal, as compared to the "-O" sample type.

All oxides were 2 nm thick. The experiment was done with pure oxide ("O") and a nitrided oxide ("NO"). The results are shown in fig. 3.1. The sample with the lowest sheet resistance is used as reference. For the other samples, the relative difference in sheet resistance is evaluated, in order to obtain information on the influence of the oxide. The percentile sheet resistance increase is converted into a dose loss scale, assuming an inverse linear dependence $D \propto 1/R_s$.

Several conclusions can be drawn from these results, e.g. that the growth and removal of an oxide after implant ("-O-") leads to a dramatic loss of active dopants that are consumed during oxide growth and then removed together with the oxide.

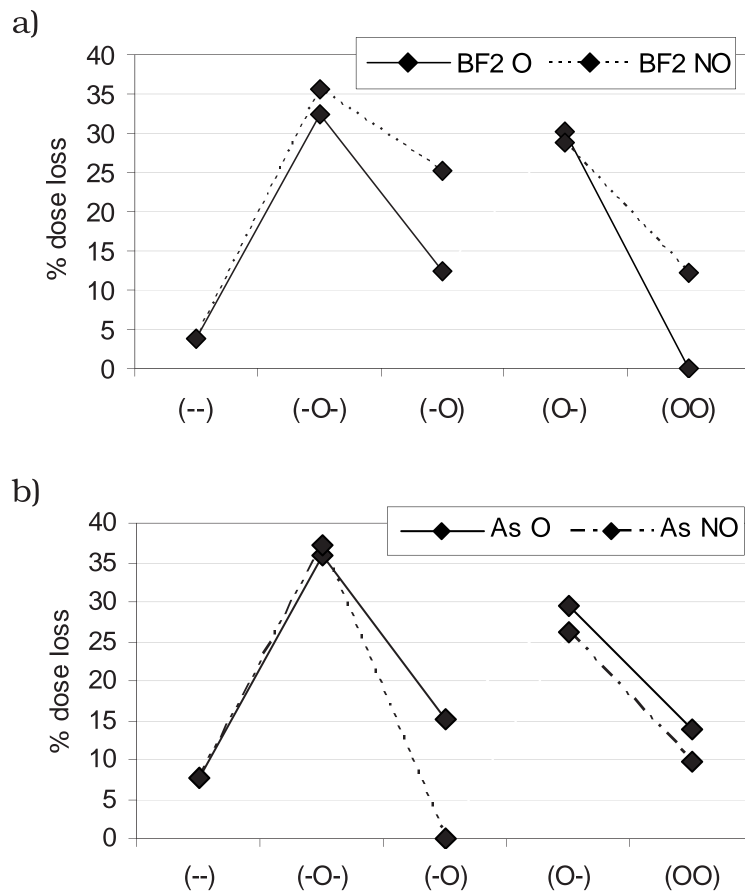

Figure 3.1: Dose loss results, calculated from sheet resistance increase for differently prepared samples (see text). (Samples from lots P000575 and P000657)

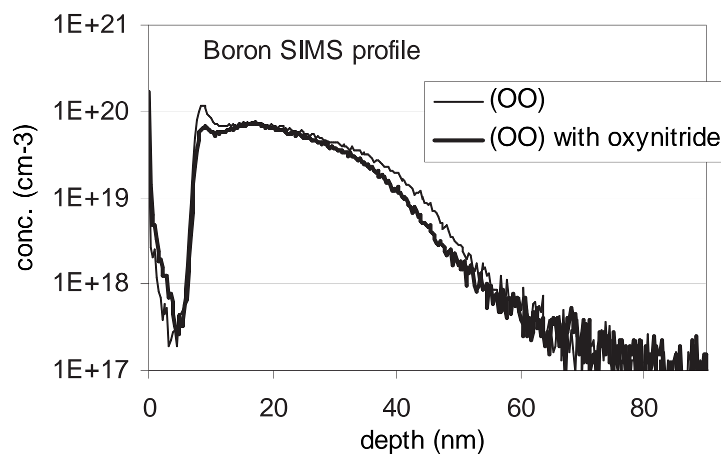

Figure 3.2: SIMS profiles of two samples. After anneal, the screening oxide was removed, then the sample was covered with around 11 nm of a-Si. (Samples from lot P000575)

Instead, if the oxide is left on the sample during anneal (“-O”), the dopants can diffuse back into the junction and contribute to conductivity. Similarly, implanting through an oxide and then removing the oxide leads to the loss of all dopants implanted into the oxide (“O-”). Leaving the oxide on the sample allows them to diffuse back into the junction (“OO”). For the samples with oxynitride, a similar behavior is observed.

A careful examination of the experimental conditions, however, showed several side effects that made it very difficult to draw final conclusions form these sheet resistance data: Firstly, the mobility of the carriers is not known precisely, hence any dose estimation based on R_s data s very coarse. Secondly, samples without oxide develop a native oxide of similar thickness within hours, and, especially during anneal, significant oxidation occurs. The comparison of samples with and without oxide is therefore virtually impossible. The observed effects have to be explained by a wealth of different oxidation-influenced phenomena, rather than by the mere presence or absence of oxide during implant and anneal.

Additional SIMS profiles were obtained to examine the samples, two of them are shown in fig. 3.2. These profiles already exhibit a hint on the existence of a pile-up at the interface between oxide and silicon. Before SIMS profiling, the samples were etched and cleaned to remove the surface oxide, then 11 nm of amorphous silicon were deposited. This was done to avoid SIMS surface artifacts and matrix effects, but obviously any interfacial features are significantly disturbed by this procedure.

During the subsequent experiments, these methods were continuously improved to obtain more precise results. ERD/RBS mainly replaced SIMS for high-resolution near-surface profiling, and several small accompanying experiments helped understanding the various phenomena of the sample processing as well as of the measurements, to interpret all results in sufficient detail.

3.2 The interfacial pile-up

Classically, the word “segregation” describes an effect that causes different concentration levels of an impurity species on both sides of an interface in thermal equilibrium. This is due to different solid solubility in the two adjacent materials. A concentration step occurs at the interface. The segregation coefficient is defined as

$$m = \frac{C_{Si}}{C_{Oxide}} \tag{3.1}$$

In a profile, this step appears smoothed due to limited profiling resolution as well as layer thickness variations.

For boron, Colby et al. [69] experimentally found an easy formula describing the temperature dependence of the segregation coefficient observed in their experiments:

$$m = 0.03 \cdot e^{0.52/k_B T} \tag{3.2}$$

This model can be refined to not only include a dependence on temperature, but also on pressure (e.g. induced by stress) and other parameters [70].

Charitat et al. [71] give an overview over the theoretical description of m. Their experiments indicate that m lies in the region of 0.5 to 0.6 for temperatures of around 1000 °C, being of interest for this thesis. The boron concentration in the oxide is thus around twice as high as in the silicon. Arsenic, on the contrary, is known not to be soluble in oxide. Segregation coefficients of up to 30 are observed [72].

The samples investigated during this work were annealed in a spike anneal. The results indicate that the fast ramp rates and the short dwell time at peak temperature are not sufficient to reach thermal equilibrium. Furthermore, the surface oxide of mostly 2 nm is too thin to exhibit full bulk properties. Therefore, the concentration step described above was not observed on most of the samples.

Instead, the dopant profiles revealed a more or less pronounced pile-up at or around the interface between silicon and the surface oxide. Fig. 3.3 shows as-implanted and annealed profiles for typical B, BF_2 and As implants, measured by ERD. This pile-up can contain a very high amount of dopants. Between 25% and more than 60% of all implanted dopant atoms are found in the pile-up after a spike anneal, depending on the implant and anneal parameters and the type of screening oxide used.

3.3 Pile-up profiling techniques

This section will focus in more detail on advantages and disadvantages of the main profiling techniques used for this work. ERD/RBS, SIMS and TEM will be compared with respect to their contribution to the characterization of the interfacial pile-up.

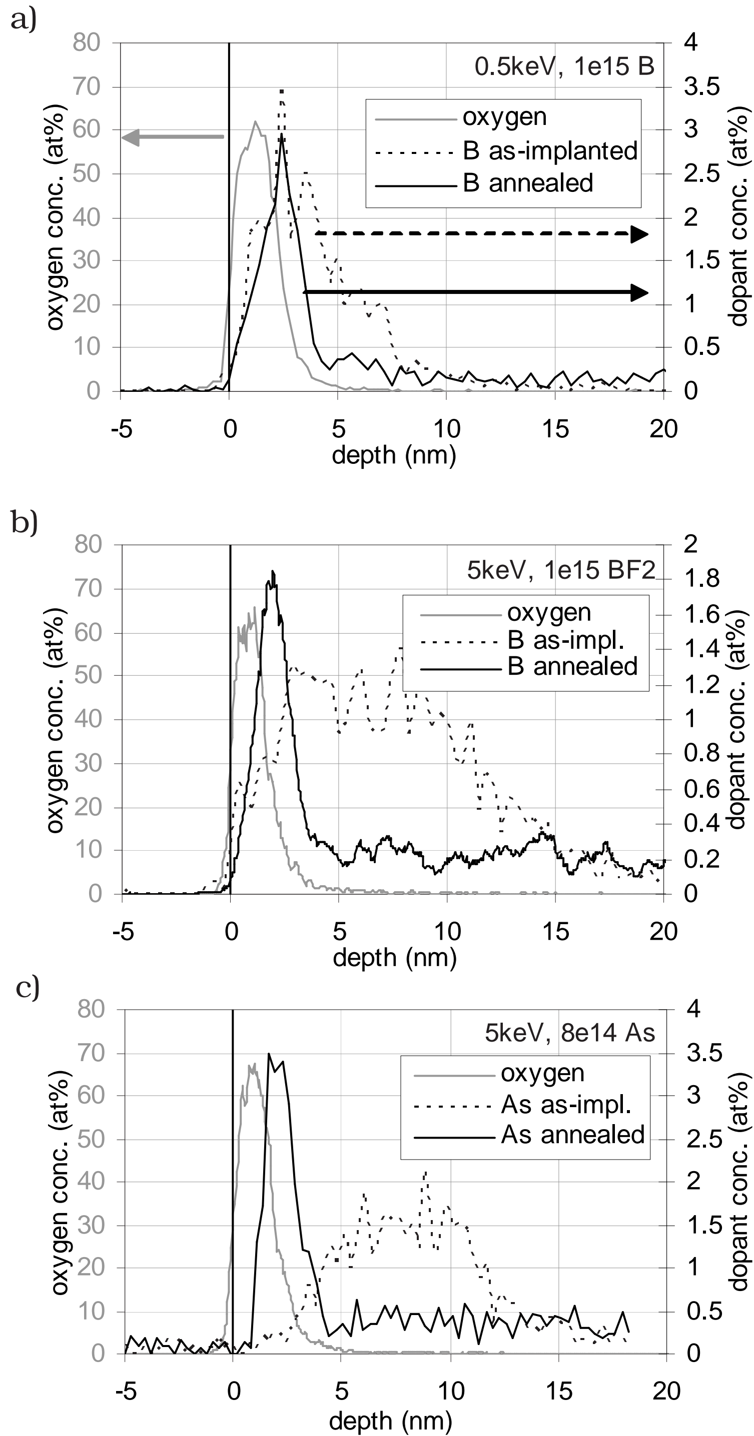

Figure 3.3: ERD/RBS profiles of different samples implanted with B, BF_2 and As. For each implant, as-implanted (dashed line) and annealed profiles (solid line) are compared. The oxygen profiles are drawn in gray. (Samples from lots T020208 and E010731)

3.3.1 ERD/RBS results

The results shown in fig. 3.3 were obtained using high-resolution ERD and RBS. These methods are able to resolve the pile-up in unprecedented detail. Therefore they contributed the most important information on the pile-up for this work, and they will be used as main basis for the analyses presented in the following chapters.

The ERD/RBS profiles primarily reveal that in all investigated samples, a pile-up is built at the silicon/oxide interface. The pile-up is observed for B as well as for As implants, reaching concentrations of 4 atomic percent (for these samples). The presence of fluorine seems to reduce the boron pile-up height.

The profiles reveal clearly that a pile-up exists at the interface, but it is difficult to judge precisely whether it is located more on the silicon or the oxide side of the interface, or how far it extends into each direction. In chapter 4, a detailed investigation and comparison of numerous samples will be given to extract the best possible information.

Definition of the depth resolution

The depth resolution of ERD/RBS can be easily determined by examining the slopes of the oxygen profile. The easiest value to determine is the distance between the 10% and 90% points of the slope. This corresponds quite well to the FWHM (Full Width Half Maximum) of the profile of a theoretical delta peak with zero thickness, subject to the same gaussian broadening[1].

For this work, we define the resolution of the ERD/RBS technique as *half of the 10%-to-90% distance of the oxygen slope.* At the surface, an excellent resolution of 0.3-0.4 nm is observed. The right-hand slope of the oxygen profile shows for most samples a deteriorated resolution of 0.5-1 nm. Since in this work, the ERD results were used to determine dopant profiles close to the surface with a special interest in the region around the Si/SiO_2 interface, the right-hand slope of the oxygen profile gives a good indication of the resolution of the dopant profile around the interface.

For a discussion of the pile-up width, its HWHM (Half Width Half Maximum) will be measured. Assuming that the interfacial pile-up is very narrow, we can directly compare the HWHM to the resolution – if the pile-up's HWHM is larger

[1] To be precise, the 10%-to-90% distance of a broadened step function corresponds exactly to the width of a gaussian peak measured at 43.5% of the peak height.

than the resolution, it can be deduced that the pile-up is thicker than a delta peak.

These values correspond also reasonably well to one-sigma deviations, and should thus be considered as "±-values".[2] Given the limited spatial resolution of the ERD detector (in most profiles 5 points per nm), it is useful to treat all three discussed values (profile resolution, HWHM and one-sigma) as identical for simplicity.

For a direct comparison, however, the different contributions to the total depth error have to be kept in mind:

- *Limited depth resolution of the measurement* – This depends on the precise adjustment of the tool and might be different for the three investigated elements boron, arsenic and oxygen. While B and O have similar atomic masses and are subject to very similar measurement variations, a greater difference is expected towards As, since it is measured using an RBS setup instead of ERD. In the measured profiles, the depth resolution is typically better for As than for B and O.
- *Interface roughness* is a physical property of the sample and contributes identically to the widening of As, B and O profiles.
- *Oxide thickness variations* over the sample are considered in two ways: For small scales (up to a few mm), they appear as a deterioration of the depth resolution and can be treated identically to the above-mentioned interface roughness. In larger scales (above 1 cm), they can lead to a shift between the profiles of the different investigated elements, since each of them is measured on a different spot on the sample.

Two possible alternative profiling methods, SIMS and TEM/EDX, are presented in the following sections. Their results provide important complementary information to the ERD profiles, but do not have sufficient resolution in the near-surface region to replace the ERD/RBS profiles.

3.3.2 SIMS results

Next to ERD/RBS, SIMS was the main profiling methods used in this work. SIMS has become a wide-spread standard means of profiling with good speed, tool avail-

[2] The HWHM is measured at 50% of the peak height. This is by a factor of 1.18 higher than the one-sigma distance (measured at 60.7% of the peak height).

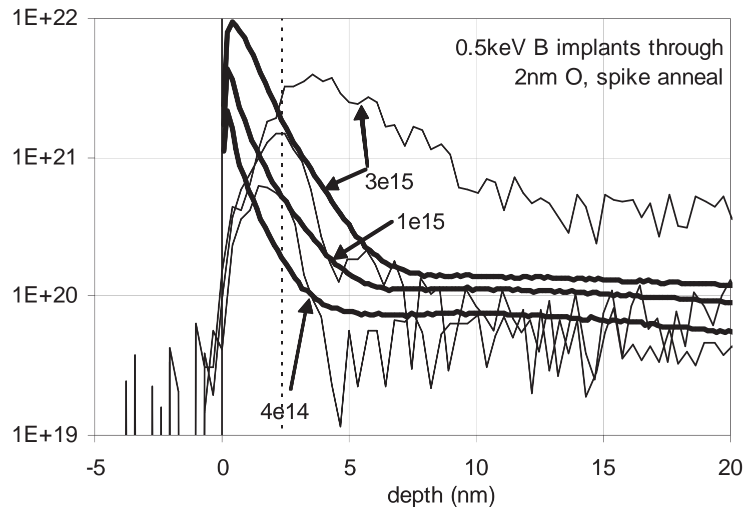

Figure 3.4: Comparison of SIMS and ERD profiles of three samples implanted with 4e14, 1e15 and 3e15 B (at 0.5 keV) and annealed in a 1070°C, 1s spike anneal. The dashed line indicates the approximate position of the oxide/silicon interface. (Samples from lot T020208)

ability and versatility. It provides high resolution (down to ∼1 nm per decade) and good dose accuracy (∼5% for B and 10% for As, [73]). However, for profiling extremely shallow profiles within a few nm of the surface, the SIMS surface artifacts are the dominant feature of the profile and prohibit any usable results from that region [7, 8].

The origin of the SIMS surface peak is not entirely understood yet [74]. It was suspected that the peak is created by B deposition on the wafer during exposure to the cleanroom air[3]. This explanation was proven wrong by examining the isotopic composition of the B seen in SIMS: While in nature, the ^{10}B:^{11}B ratio is around 20:80, the implants only use ^{11}B. The ^{10}B concentration in the SIMS surface peak is, however, several orders of magnitude too low [75].

One possible explanation of the SIMS surface peak is a badly resolved interfacial dopant pile-up. While the pile-up is found in a depth of 1-2 nm, depending on the sample, the SIMS peak is measured exactly at the surface, with a width of several nm (figure 3.4). In some cases, the dose measured by SIMS corresponds well to the one measured by ERD, in other cases differences of up to 30% are seen. Even with a SIMS process optimized for <1 nm depth resolution, the surface artifact is still seen (see also [7, 8]); the same holds for samples on which the pile-up

[3] Amongst other sources, boron in the air might originate from the air filters in the cleanroom ceiling that contain borophosphosilicate glass (BPSG).

is removed by etching. The pile-up is therefore possibly contributing to the SIMS surface peak, but this effect cannot be the only explanation.

It is known that during sputtering or ion implantation, a rearrangement of the atoms close to the surface occurs. This can also enhance dopant self-sputtering during implant [75]. The SIMS surface artifact seems to be mainly due to this dopant rearrangement, along with a non-equilibrium sputtering behavior during the initial stages of the process [57]. It is concluded that, for the purpose of this work, the quantification of SIMS close to the surface is not sufficiently precise.

As a consequence, SIMS was used during this work mainly to determine the junction depth, the overall profile and the total dose. Dose information from SIMS also helped calibrating the ERD profiles.

It should be noted that for a direct comparison of ERD and SIMS results, the different sputtering speed of SIMS in oxygen and in silicon should be compensated for. As explained in section 2.7, this can be done by stretching the depth scale and compressing the concentration scale of the oxide part of the profile by a factor of 2.3 or 0.44, respectively. Various smoothing algorithms can be used to generate a smooth transition from oxide to silicon, which however introduce various artifacts into the profile. To avoid a deterioration of the profile information, the SIMS results in figure 3.4 are not compensated. Thus in principle, the surface peak is wider and less high than displayed in fig. 3.4. This, however, does not affect the principal judgement of the SIMS surface peak given above.

3.3.3 TEM and TEM-EDX results

Figure 3.5 shows TEM micrographs of two samples implanted with high dose B or As, and annealed in a standard spike anneal. The As sample exhibits a dark line at the interface. This line is probably caused by the pile-up, but it is difficult to distinguish the pile-up from the various side effects that can be seen at interfaces. For B, the contrast seen at the interface is not considered significant.

For a more detailed analysis, TEM-EDX (fig. 3.6) was used to generate an arsenic dopant profile across the interface. EDX was measured at several separate points across the interface, with a distance of 10 nm (a) or 5 nm (b), respectively. Each of the shown points is an average over 5 or 6 individual measurements at identical distance to the interface. The results are given in counts, thus in arbitrary units. A boron-implanted sample was investigated by the same method, however the EDX setup was not able to detect a discernible boron signal.

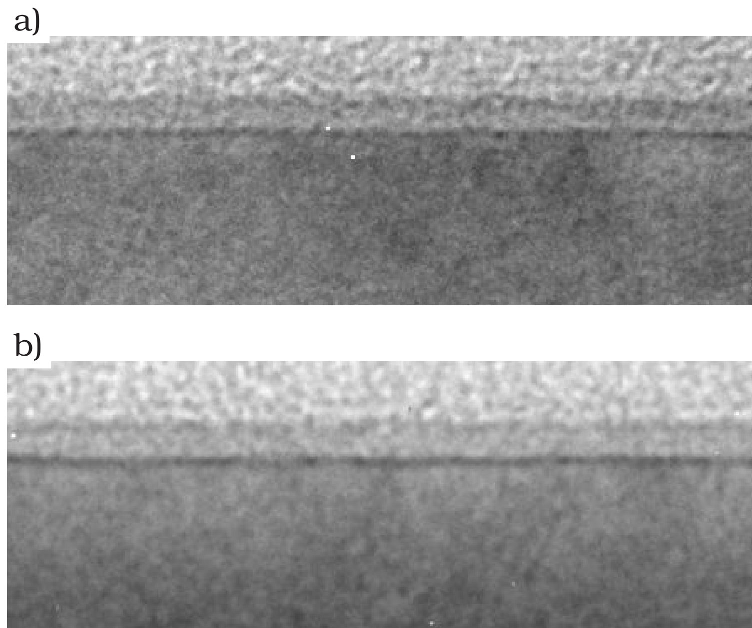

Figure 3.5: STEM micrographs (bright field imaging) of a 0.5 keV, 3e15 B (a) and a 2 keV, 3e15 As sample (b). (Samples from lots T020208 and P030122)

The leftmost measurement points of both samples show a significant carbon concentration - this is the glue on top of the sample, used for TEM specimen preparation. For the deeper measurement points on the right side of each graph, still a high amount of oxygen is seen. This is due to the native oxide growing on the very thin TEM specimens during preparation.

For both samples, a clear transition is seen from a region with high oxygen content (the oxide layer) to one of almost pure silicon on the right (the bulk silicon). At the interface, an increased arsenic concentration is observed. This peak is, however, much wider than the thin pile-up at the interface, seen in ERD. This widening is caused by several effects, like a too large beam diameter (a few nm), beam widening by scattering in the sample, or beam drift during the measurement (up to 6 nm). Since the pile-up contains a very high concentration of arsenic, already a small gaussian widening of the beam can cause a significant detection of arsenic in the neighboring measurement points.

The STEM-EDX results confirm the primary result of the ERD investigations – a pile-up of quite high concentration is observed at the interface between oxide and silicon. The STEM resolution is, however, not comparable to ERD.

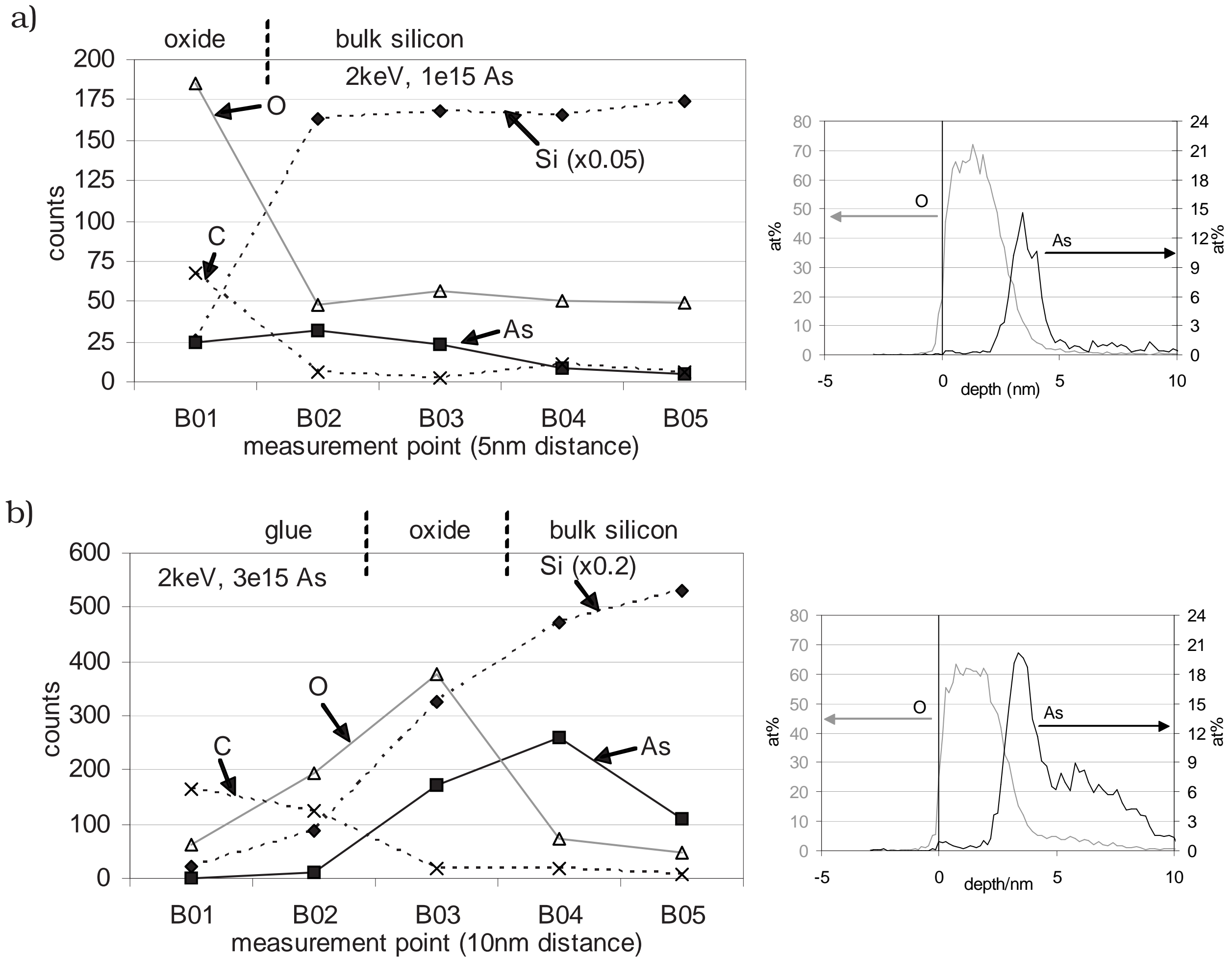

Figure 3.6: Dopant concentration measured by STEM-EDX across the Si/oxide interface of 2 keV, 1e15 and 3e15 As samples. The insets show RBS profiles of the same samples for comparison. (Samples from lot P020645 and P030122)

3.4 Dopant diffusion

Macroscopically, any diffusion follows the Fick equation ("normal" diffusion),

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2} \tag{3.3}$$

with the concentration C of dopant atoms. The diffusivity D is primarily a function of temperature.

However, for the processes that govern diffusion during a very short spike anneal, D becomes a very complex function of temperature, the binding conditions of each individual dopant atom, the presence of lattice damage and so on. Therefore, individual dopant atoms must be considered for a more precise description of the diffusion processes in a lattice. They are present on a (substitutional) lattice site, as interstitial between the lattice atoms, or in various types of combinations with lattice or other dopant atoms (clusters). Several direct (e.g. diffusion of an interstitial dopant atom) and indirect (diffusion mediated by a defect) diffusion modes exist (cf. [76, 77]). The two most important are:

- Vacancy: A substitutional dopant atom binds to a vacancy. For diffusion, they exchange place. This mechanism is preferred by large dopant atoms, such as As of Sb.

- Intersticialcy: A pair of one dopant and one Si atom occupy an substitutional lattice site and diffuse without dissociation. For boron, this is the preferred diffusion mechanism.

These two processes are, for the respective dopant species, largely predominant for high temperatures [78]. Most other dopant configurations (e.g. dopants bound in clusters) have a negligibly low diffusivity. In both cases, however, the mentioned main diffusion processes are suppressed if no lattice damage is present, which leads to a very low overall diffusivity.

3.4.1 Enhanced diffusion

Additionally, various enhancement processes increase the dopant diffusivity and lead to the so-called "anomalous" or "enhanced" diffusion. They describe processes in the sample that under certain circumstances generate additional defects (mainly silicon self-interstitials), which then tremendously enhance the diffusion. Enhanced diffusion typically is an important problem for boron, by far more than

for arsenic. The most important creation mechanisms for excess silicon interstitials are the following:

- TED – Transient Enhanced Diffusion: After implant, about one excess silicon atom per implanted atom is present in the lattice ("+1 model"). They can be present as interstitials or in the form of silicon-rich aggregates (mainly dislocation loops or 311 defects, which are formed during a thermal treatment above 600 °C). Starting at 750-800 °C, the clusters dissolve (e.g. by Ostwald ripening). The excess Si atoms diffuse and cause anomalous dopant diffusion [79–81]. Since the surface of the wafer is a sink for Si atoms, an anisotropic stream of interstitials towards the surface occurs.
- OED – Oxygen Enhanced Diffusion: Oxidation before or during an anneal causes silicon atoms to be injected into the crystal, leading to enhanced dopant diffusion.
- BED – Boron Enhanced Diffusion: In regions of high B concentration, the dopant atoms tend to precipitate and form clusters that may include silicide phases. The dissolution of the clusters during anneal injects Si interstitials into the lattice [82].

TED as well as the other types of enhanced diffusion set in at around 750 °C. They are mediated by the presence of excess silicon atoms, not by their movement. Therefore, these effects stop as soon as the excess silicon atoms have disappeared due to annihilation at the wafer's surface (hence the name "transient"). To reduce enhanced diffusion, it is necessary to reach as fast as possible a high temperature during anneal, so that the Si interstitials disappear as quickly as possible. This is one of the reasons why high ramp-up and -down rates are necessary [83].

In [84], Dunham et al. give an overview over the various rate and diffusion equations to model shallow dopant diffusion in silicon.

Agarwal et al. have shown that diffusion enhancement due to the TED mechanism is significantly reduced for ultra low implant energies below 1 keV [85, 86]. This is consistent with the expectation that at zero implant energy, no excess interstitials should be produced. However, for boron implants at low energies, a very shallow layer of high-concentration B is produced. Above a certain dose threshold ($3 \cdot 10^{14}$ to $10^{15}\,\mathrm{cm}^{-2}$ for a 0.5 keV implant), boron-enhanced diffusion (BED) is caused by this layer, resulting in a diffusion enhancement by a factor of up to four, similar to TED. Also for the samples prepared for this work, this effect is expected to produce significant amounts of excess silicon interstitials.

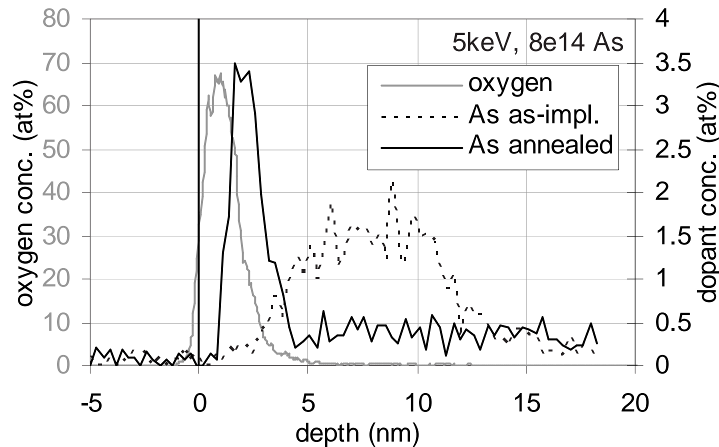

Figure 3.7: ERD/RBS profiles of an arsenic sample, as-implanted (thin line) and after anneal (thick line). (Sample from lot E010731)

3.4.2 Uphill diffusion

Fig. 3.7 shows a typical profile of an As implant. Most of the implanted dose is found in a depth of 5-10 nm, with a maximum concentration of around 1.5 at%. During spike anneal, obviously a large fraction of the dopant atoms move from this peak towards the interface, to create a pile-up with up to 3.5 at% of dopant atoms in the lattice.

This phenomenon of diffusion towards a region of high concentration is called uphill diffusion [87,88]. During implant, a region rich of excess silicon interstitials is created in the silicon, deeper than the implanted profile. In average, one silicon interstitial is created for each implanted dopant atom ("+1 model").

Since the wafer surface acts as an interstitial sink during anneal, a large anisotropic stream of silicon self-interstitials towards the surface occurs. By forming dopant-defect pairs, they carry dopant atoms along, and an anisotropic dopant diffusion towards the surface occurs. This appears as a diffusion towards regions of higher concentration, hence the name "uphill". Since the presence of Si self-interstitials is required for uphill diffusion, the effect is closely related to TED. Both effects appear in parallel.

Thus, two processes are necessary for the formation of a pile-up:

- An uphill diffusion of dopants is observed – dopants move from regions with low to medium concentration towards the high-concentration pile-up.
- The dopant atoms are trapped at the interface, either by building up dopant clusters, or by trapping to lattice sites or to dangling bonds. This trapping

prevents the diffusion back into the silicon. The interface is required to effectively act as a dopant sink.

If the second effect, the interfacial trapping of the dopants, is strong enough to retain all pile-up atoms, it prevents by itself any diffusion back into the bulk part of the junction. In this case, the uphill diffusion is actually not required to build up a huge non-equilibrium concentration in the thermodynamical sense. It only delivers enough dopants to the interface, where they are blocked from any further participation in the diffusion.

In some cases the as-implanted concentration around the interface is already so high that no additional creation of a pile-up is observed. For example, this is the case for the B sample in figure 3.3a. However, a comparison with other B samples (e.g. fig. 4.9b) shows that the uphill diffusion and trapping effects, as discussed here for arsenic, also hold for boron. In the specific case of fig. 3.3a, the normal diffusion from the implanted maximum towards the bulk is large enough to compensate the pile-up building effect.

It is also likely that the mechanism trapping the dopants at the interface has only a limited capacity and is not able to incorporate more dopants. This limits the dopant concentration in the pile-up size to a certain level.

3.5 Effects mentioned in the literature

3.5.1 Theoretical work

Sai-Halasz et al. [89] have shown that arsenic has essentially a unity sticking coefficient at a Si-SiO_2 interface during annealing. Their RBS studies indicated that the atoms were trapped in a single monolayer of oxide near the interface where they were immobile and electrically inactive.

Lau et al. [90] for the first time introduced the interface itself as a third phase next to the Si and SiO_2 phases. Their model, using rate equations for the transitions between the three phases, was able to well describe the experimentally found pile-up dose as a function of dopant concentrations on both sides of the interface for phosphorus-doped samples.

Baierle, Dabrowski et al. [91] have conducted ab initio simulations of phosphorus as a typical donor segregating to a Si/SiO_2 interface. They found that at least half a monolayer ($1\,\mathrm{ML} \approx 7\cdot10^{14}\,\mathrm{cm}^{-2}$ for (100) Si) of electrically inactive P atoms can be trapped at the interface. The group published evidence that the trapping of P at an Si/oxide interface occurs by (1) trapping at interfacial dangling bonds,

(2) trapping at vacancies and vacancy-oxygen complexes and (3) the formation of threefold-coordinated P pairs.

Refining their theory, they found that the dopant trapping mechanism at the interface depends on the dopant concentration C_P in proximity of the interface [92, 93]. For low C_P below $10^{17}\,\mathrm{cm}^{-3}$, the dopant atoms replace undercoordinated Si atoms (i.e. Si atoms with dangling bonds), while dopant pairs are formed at high concentrations above $10^{19}\,\mathrm{cm}^{-3}$. For intermediate C_P, the trapping mechanism has not yet been identified. Concerning the exact location of the trapped atoms, they found that the donor-oxygen bonds are energetically not favorable, which is consistent with the fact that P and As are expelled from the oxide during oxidation.

Vuong et al. [94, 95] presented a model for boron and arsenic pile-up formation. They assumed $2\cdot10^{14}\,\mathrm{cm}^{-2}$ electrically inactive trapping sites at the interface, together with certain activation energies for trapping and de-trapping. The model helped to significantly improve TCAD process simulations, to correctly predict the transistor's electrical behavior.

Shima et al. [96, 97] calculated the activation energies for boron for the various processes at and around the interface. They found that the barrier height of 0.4 eV for atoms getting trapped at the interface from the bulk silicon is smaller than for the reverse process (0.5 eV), which is again much smaller than the barrier for the diffusion of dopants from the interface region towards the oxide (0.8 eV). These results qualitatively explain why B atoms are trapped at the interface to form a pile-up. They conclude "that the main factor in this segregation is the existence of the Si surface".

3.5.2 Experimental Results

Also experimentally, the interfacial dopant pile-up has been studied. The interfacial segregation behavior was investigated mainly for phosphorus. Different groups have analyzed the phosphorus pile-up, with partly contradictory results. Chou et al. [98] indicated that the pile-up is at the SiO_2 side of the interface, while Schwarz et al. [99, 100] concluded that it is on the Si side.

Sato et al. [101, 102] used neutron activation analysis (NAA) and sheet resistance measurements to investigate the phosphorus pile-up. They found significant doses of phosphorus on the SiO_2 side of the interface, after oxidizing a P-doped sample or annealing a doped Si-SiO_2 composite structure. An observed P dose reduction is attributed to the loss of the pile-up during oxide removal. Based on their observation that, after oxide removal, an additional careful Si etching does

not further increase the sample's sheet resistance, they concluded that the pile-up is located on the oxide side of the interface.

The same group later also analyzed As implanted samples with sheet resistance measurements and RBS analysis [103] and came to analogous conclusions. However, the arsenic pile-up was observed to contain a smaller dose than the phosphorus pile-up. Also, an HF dip only remove "most of the arsenic pileup".

Griffin et al. [104] have experimentally observed that for phosphorus implants with doses of $5\cdot 10^{13}$ to $4\cdot 10^{14}\,\mathrm{cm}^{-2}$, up to half of the implanted dose was lost during anneal, amongst others due to segregation at the silicon-oxide interface. They found that stripping the oxide also removes the segregated atoms, and they consequently believed the phosphorus atoms to stick in the oxide near the interface, being electrically inactive. They further found that an interfacial phosphorus pile-up created during an anneal at 700 °C or 800 °C is mostly removed by a subsequent 1100 °C anneal, during which the dopant atoms diffuse back into the silicon.

Kasnavi et al. studied the behavior of arsenic during anneals on samples with different oxide thicknesses [105, 106]. Using XPS as main analysis method, they clearly observed a large As pile-up in the vicinity of the silicon-oxide interface. They found that an HF dip under inert atmosphere does not remove the As pile-up and concluded that the pile-up is located within 5 Å on the Si side of the interface. Using a very careful procedure of growing and etching single monolayers of oxide they were able to confirm that it is located in "the first few monolayers" on the silicon side of the interface.

Shima et al. [96, 97] also published experimental data. They investigated the pile-up in boron implanted samples, using XPS with high angles to reduce the photoelectron escape depth to around 2 nm. While they did not observe any B signal with a 2 nm surface oxide present, a clear signal was seen after the oxide had been etched away with HF. From this they conclude that the majority of the pile-up dopants is found at the Si side of the interface. With a sophisticated backside-SIMS analysis they found that the pile-up extends by not more than 0.6 nm into the silicon.

Koh et al. [107] quantitatively evaluated the dose loss in low-energy As implanted samples. After a 5 keV, 1e14 As implant they found that 43% of the dopants were located in the 5 nm thick screening oxide. From the remaining dopants, 50-70% segregated to the pile-up during a 850 °C, 30 min. furnace anneal.

Vuong et al. conducted SIMS measurements on B-doped SOI structures [87]. They observed a dopant pile-up at the interface to the screening oxide, and a second one at the interface to the buried oxide layer (BOX). They found that the pile-up dose depends significantly on the amount of excess silicon interstitials present. The dopants in an MBE-grown sample without any implant damage showed neither diffusion nor pile-up formation. On similar SOI structures, Crowder et al. [108] had showed already before that in the presence of the buried silicon-oxide interface with its pile-up, the dopant dose in the pile-up at the interface to the top oxide layer is reduced.

Topuria et al. [109, 110] used spatially resolved EELS (electron energy loss spectroscopy) and EDS (energy dispersive X-ray spectroscopy) on an STEM tool to investigate the arsenic segregation to the interface between silicon and a cobalt silicide, as it is used to form the contact to the transistor's source and drain junctions. They found that the segregated As atoms occupy substitutional lattice sites and are electrically active. The free carrier density around the interface is increased by 5-10%.

3.5.3 Contribution of this work

All experimental results published until now have in common that they suffer from the limited resolution of SIMS and other profiling techniques. Mostly, other analysis methods such as XPS or NAA are used together with a sophisticated sample preparation process to locate the pile-up.

This work presents direct evidence for the interfacial dopant pile-up. For the first time, ERD dopant profiles with sub-nm resolution are published, showing the pile-up and its location at the silicon-oxide interface in unprecedented detail. Together with the previously published experimental and theoretical results, a more complete picture of the pile-up and its formation process is given.

4 *Pile-up formation mechanisms*

4.1 Introduction

This chapter focuses on the mechanisms that lead to the creation of a pile-up at the silicon/oxide interface. Various samples were prepared with different implant and anneal conditions. A combination between ERD, SIMS and SRP results is used to extract the maximum possible information on the samples' dopant profiles.

ERD profiles of differently prepared samples are compared, in order to distinguish the influences of the various implant and anneal parameters. From the information obtained, a phenomenological model of pile-up formation is proposed.

For discussions on the depth resolution, the HWHM of the dopant pile-up will be compared to half the width of the right-hand oxygen slope, as described in the previous chapter (page 48). The resolution data given in this chapter can thus be considered as "$\pm$" values. The given resolution values are normally precise to around 0.05-0.1 nm.

4.2 Sample preparation

Various samples were implanted and annealed to investigate the near-surface dopant distribution after an anneal. Most of the samples were implanted with 5 keV BF_2, 0.5 keV B or 2 keV As through a 2nm screening oxide in an Applied XR80 ultra low energy implanter or an Eaton NV2800 medium energy implanter. The dose was varied between $4 \cdot 10^{14}$ and $3 \cdot 10^{15}\,\mathrm{cm}^{-2}$. The most commonly used anneal was a 1070 °C, 1s spike anneal in an AST SHS2800 lamp-based RTP tool. These parameters are close to the ones used for extension formation in a typical 90 nm process flow. A scheme of the process flow used for the experiments is given in fig. 4.1.

After anneal, the obtained junctions have a junction depth of around 35 nm (arsenic case) to 60 nm (boron case), measured at a concentration of $10^{18}\,\mathrm{cm}^{-3}$. The

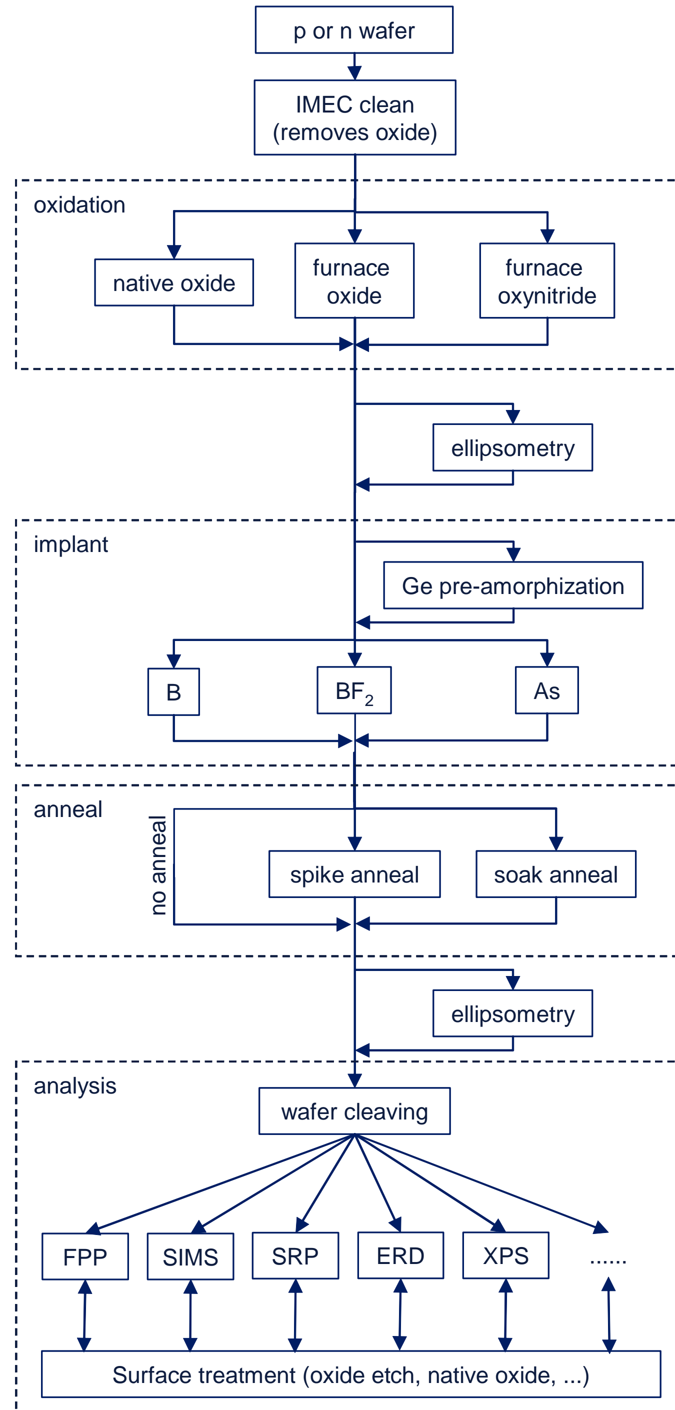

Figure 4.1: Process flow used for the experiments.

profile's maximum dopant concentration close to the surface is typically around $10^{20}\,\mathrm{cm}^{-3}$.

In a device, $10^{18}\,\mathrm{cm}^{-3}$ is the typical doping concentration of the well. Therefore the junction depth given here corresponds to the metallurgical junction depth that a certain junction would exhibit in a real device.

In our case, no special well doping was used, thus the background concentration is the one of a bulk wafer, in the order of $10^{15}\,\mathrm{cm}^{-3}$. The actual physical junction depth is in the range of 130-200 nm for all samples. This facilitates the sheet resistance measurement, since the FPP tool's probe tips never penetrate through the junction. The sheet resistance measured on the samples is considered identical to the one of a real junction with $10^{18}\,\mathrm{cm}^{-3}$ background doping, because the total number of carriers in the junction's tail is too low for any significant contribution to the total conductivity.

4.3 Dopant loss mechanisms

From the nominal implanted dopant dose, a significant amount is lost during the implant and anneal process. The dopants are subject to one or more of the following processes:

- Self-sputtering to the ambient during implant
- Outdiffusion to the ambient during anneal ("outgassing")
- Segregation to the interface
- Diffusion (and mostly activation) in the bulk part of the junction

Self-Sputtering

The implanted atoms cause with their collision cascade a rearrangement of the Si atoms close to the surface in the target wafer. During this process, surface atoms are sputtered away.[1] The sputtering also removes dopant atoms that are located at or close to the surface. The closer to the surface the implanted dopant atoms come to rest, the more they are affected by the sputtering process. The effect causes a dose limitation for ultra-low energy implants to around $1\text{-}1.5\cdot10^{15}\,\mathrm{cm}^{-2}$ for 0.2 keV B implants and $5\text{-}10\cdot10^{15}\,\mathrm{cm}^{-2}$ for 0.5 keV B implants. Above this

[1] This process is used on purpose in SIMS analysis, which uses beam energies similar to the ones of ultra-low energy implants.

dose, any additional implanted dose removes the same amount of already implanted dopant atoms. Self-sputtering becomes measurable for implanted doses above $10^{15}\,\text{cm}^{-2}$ for 0.5 keV B implants. At $3 \cdot 10^{15}\,\text{cm}^{-2}$, around 20-25% of the dose are lost [20, 111].

Similar to the self-sputtering effect, any implantation deteriorates the quality of the screening oxide, making amongst others ellipsometrical thickness measurements completely impossible. During the experiments done for this thesis it was found that a wet clean between implant and anneal slightly reduces the oxide thickness, even though the clean used is normally known not to remove any measurable amount of oxide. For the experiments described here, this clean was therefore omitted. It is believed that this implant-induced deterioration of the oxide quality also enhances oxidation during anneal.

Outdiffusion during anneal

A second dose loss effect is the dopant outdiffusion during anneal. Dopants that reside close to the surface or diffuse there during anneal can evaporate to the ambient. Similar to the self-sputtering, this effect becomes stronger for increased dose as well as for reduced energy, because for ultra-low energy implants, the dopants are located closer to the surface.

A thin 1-2 nm oxide on the silicon surface, such as a native oxide or a grown oxide (it can later be used as gate dielectric) reduces the dopant loss due to outdiffusion. However, the oxide itself can evaporate during anneal. To prevent this, standard spike anneals are run with 133 ppm oxygen in the N_2 ambient. The oxygen is believed to adsorb to the surface and thereby protect the oxide.

The diffusivity of boron in oxide increases anomalously for an oxide thickness below 6-8 nm [112]. This effect enhances the penetration of boron through ultrathin gate oxides or, as in this case, enhances the loss of boron through a thin screening oxide during anneal.

Since outdiffusion during anneal and self-sputtering have the same dependence on energy and dose, no differentiation between the two effects was made for this work. Only the total loss of dopants during implant and anneal was investigated.

Figure 4.2 shows the dependence of the retained dose on the implant parameters. A lower energy as well as a higher dose increase the fraction of dopants lost during anneal.

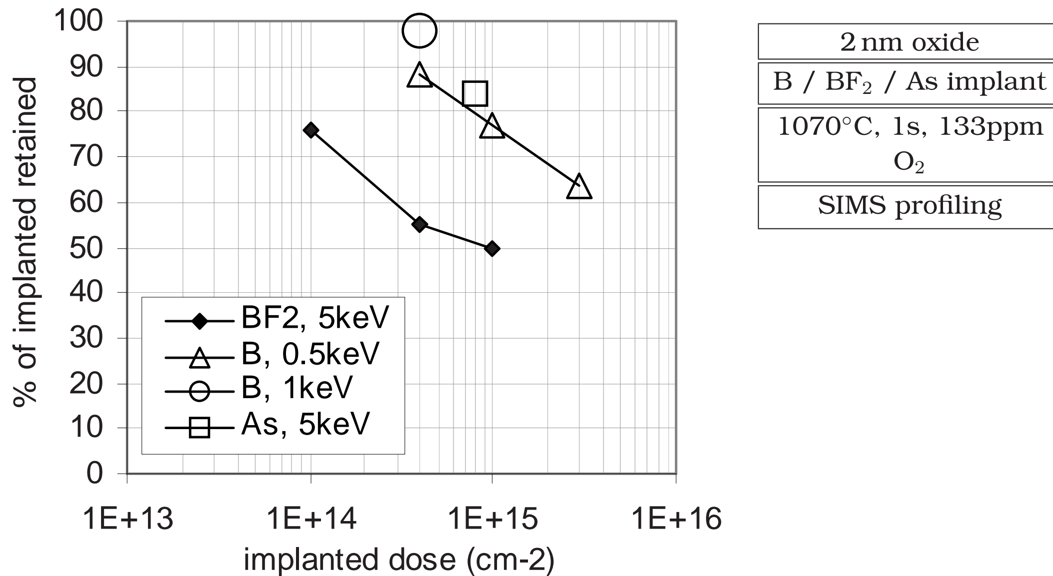

Figure 4.2: Percentile dopant retention after implant and anneal, as a function of the nominally implanted dose.

It is noteworthy that the co-implantation of fluorine (BF_2) tremendously increases the dopant loss. It is likely that this effect is caused by self-sputtering; the relatively heavy F atoms significantly increase sputtering during implant. It could be interesting to run an additional experiment where boron is implanted separately after a fluorine implant. For those samples, the self-sputtering would be identical to a single B implant, allowing a separate investigation of the effect of fluorine on self-sputtering and on outdiffusion during anneal.

Segregation to the interfacial pile-up

A large fraction of the dopants, between 25 and 50% of the total implanted dose, is observed to segregate during anneal to the interface between silicon and the screening oxide. Even though the results of this work suggest that a high active dopant concentration is found in the pile-up, most pile-up atoms are electrically inactive (cf. chapter 5). Therefore, segregation to the pile-up is a priori considered as an unwanted effect, since its dopants are lost for conductivity.

The numbers demonstrate that this effect has a large influence on the electrical properties of a transistor, and that its consideration is crucial for simulation and design of future CMOS devices. This work will focus on the details of the pile-up formation, its shape and dose, as well as its electrical conductivity.

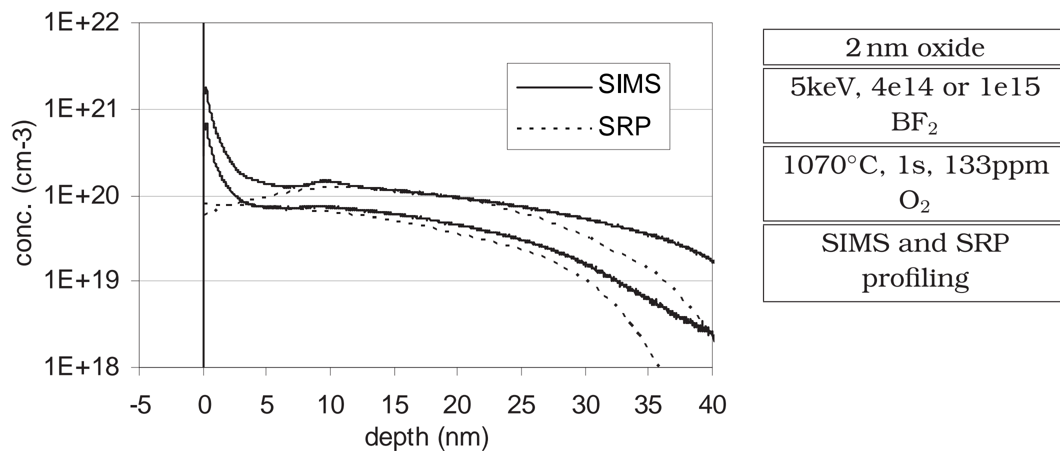

Figure 4.3: Comparison of SIMS and SRP profiles of a 4e14 and a 1e15 BF_2 implant. (Samples from lot E010731)

Dopants in the bulk part of the junction

Since the dopants that are subject to the first three mentioned effects are (mostly) lost for the junction conductivity, it is desirable to keep as many of the implanted dopants in the junction's bulk part and to activate them. Fig. 4.3 shows that for the typical doses that were investigated during this work, those dopant atoms in the bulk are activated at a level close to 100%, within the errors of SIMS and SRP (∼10%).

In the tail region of the profile, a difference is seen between SIMS and SRP. While SIMS gives quite precise information on the chemical concentration of the entire profile, the SRP results are subject to the carrier spilling effect, which influences regions with inhomogenous doping, such as the profile tail. In SRP, the electrical junction always appears shallower than the metallurgical junction seen in SIMS. (The phenomenon was explained in section 2.4, figures 2.2 and 2.4, page 26).

Figure 4.3 also shows that neither the SIMS surface peak nor the ERD interfacial pile-up are seen in SRP. This is explained by the limited resolution of SRP with a step distance of 1-2 nm and a tip imprint depth of 5 nm. It does not necessarily indicate that the pile-up is electrically inactive. A more detailed investigation of raw SRP data will be presented in section 5.5.5.

Sample	B	BF_2	As
Depth resolution from oxide slope	0.85 nm	0.75 nm	0.90 nm
HWHM pile-up	0.95 nm	0.90 nm	0.80 nm
Pile-up dose by ERD	3,8e14	2.1e14	4.2e14
Total implanted dose	1e15	1e15	8e14

Table 4.1: Comparison B-BF_2-As: Data extracted from the profiles in fig. 4.4. The values are accurate to about ± 0.05 nm

4.4 Dependence on implant conditions

From the ERD/RBS profiles (e.g. fig. 4.4) it is primarily concluded that there is a pile-up at the interface. To determine whether and how far it possibly extends into the oxide and/or silicon, the profiles are evaluated in more detail. In the following sections, ERD dopant profiles will be used to investigate the dependence of the pile-up shape on different implant and anneal parameters within the ranges useful for USJ formation.

4.4.1 Species dependence

Fig. 4.4 compares B, BF_2 and As implants. The dopant profile is given as implanted and after anneal. The oxygen profiles are used to determine the position of the oxide-silicon interface.

All samples exhibit a clear pile-up at the interface. Table 4.1 presents precise data extracted from the pile-up profiles. The depth resolution given in the first line is extracted from the right-hand slope of the oxygen profile.

For both B and BF_2 implants, the pile-up HWHM is by 0.1-0.15 nm bigger than the depth resolution. This indicates that the actual pile-up is very thin, spreading over not more than 2-3 Monolayers (1 ML $\approx$ 0.13 nm). The results demonstrate that the pile-up is located “at the interface”. However, due to the limited ERD resolution, it is difficult to judge whether and how much it spreads into the oxide or silicon side of the interface.

In both boron profiles, the pile-up seems to extend right until the surface. This is probably due to a part of the as-implanted profile being immobilized or trapped in the oxide and will be discussed below in more detail. Considering that this contributes to the peak’s width in the ERD profile, it seems possible that the actual pile-up itself is confined to a single monolayer.

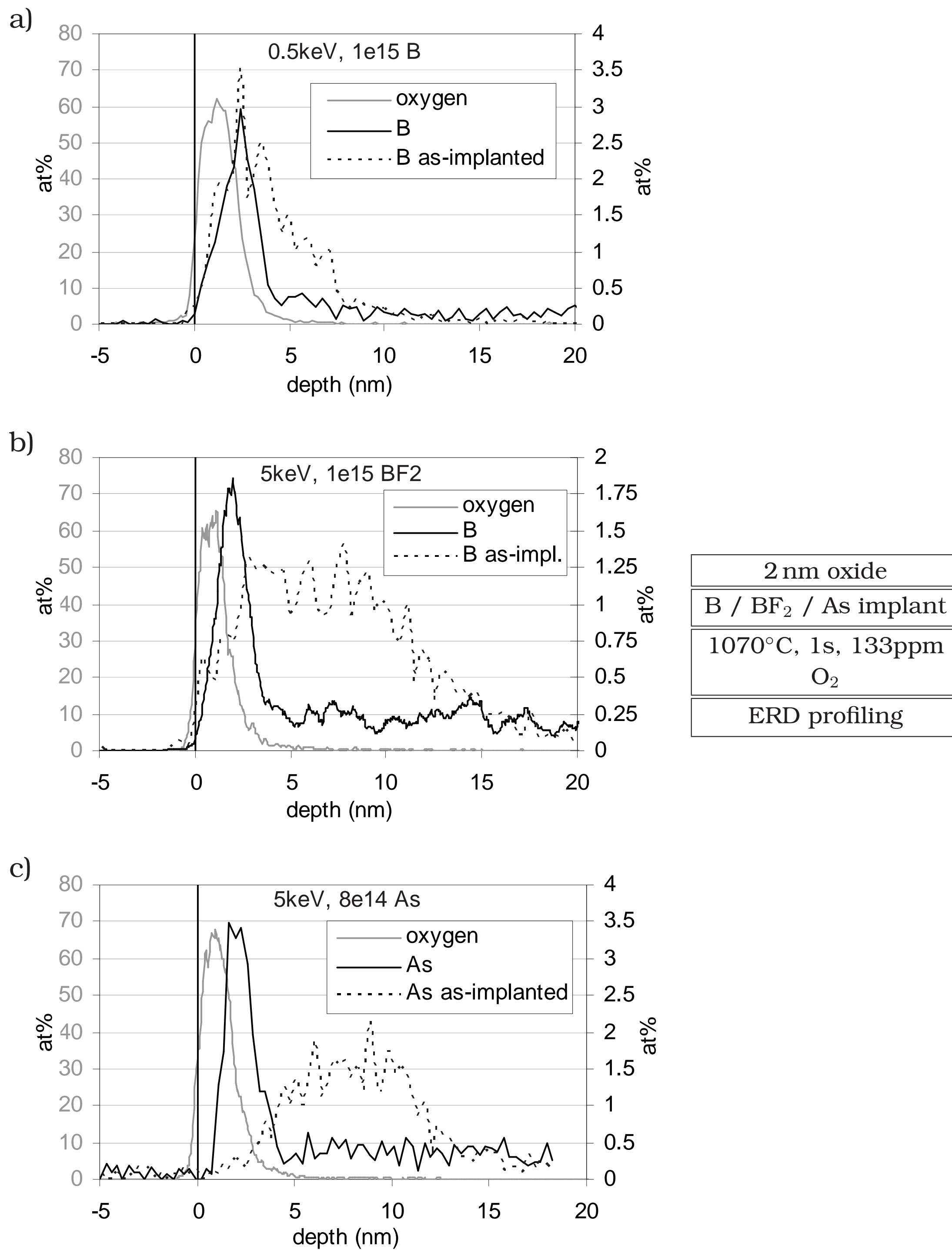

Figure 4.4: Comparison of B, BF_2 and As implants. As-implanted and annealed profiles are shown together with the oxygen profiles. SIMS profiles of similar samples were shown before in fig. 1.10. (Samples from lots T020208 and E010731)

For the As profile, an RBS measurement setup was used. As described already in section 2.8.2, the arsenic profile can be measured with slightly higher precision than boron or oxygen. As a result, the arsenic pile-up peak is measured even sharper (by 0.1 nm) than the oxide/silicon interface (determined from the oxygen ERD profile). From the profile, it seems likely that the As pile-up is confined to one monolayer, however a spreading into the oxide or silicon by a few monolayers cannot be entirely excluded.

In the bulk part of the junction below the pile-up, the concentration is observed quite constant in the order of 0.2 to 0.4 atomic percent. This corresponds well to the concentrations of around $10^{20}\,cm^{-3}$ that are typically obtained by SIMS in a depth of 5-10 nm for similar samples.

The differences between the B and the As pile-up will be discussed in more detail in the following sections.

4.4.2 Dose dependence

Discussion of the pile-up dose

In this section, the dependence of the pile-up size and shape on the implanted dose is evaluated. Where noted, SIMS results were used to calibrate the ERD/RBS profiles. SIMS results are in general more precise than ERD/RBS results concerning integral doses. The calibration, however, is difficult to perform: While SIMS does not resolve the pile-up close to the surface, ERD and RBS become very unreliable for great depths. The ERD/RBS profiles depth was limited to around 25 nm, in some cases to only 11 nm.

Consequently, only the intermediate region between around 5 nm and 20 nm can be used for a comparison of ERD/RBS and SIMS. The calibration procedure uses a scaling factor applied to the ERD profile to fit the SIMS results in this intermediate region. Where possible, the same correction factor is applied for all samples of one measurement series. This optimizes the comparability of results within one measurement series, assuming that they are all subject to the same disturbances during measurement.

However, ERD/RBS profiles exhibit significant noise levels in this intermediate region of the profile. Any calibration can therefore not always assure precise dose values; instead it optimizes their comparability.

For the following discussion, all arsenic profiles and the 3e15 B profile were calibrated using this method.

Boron For boron, the dopant dose in the pile-up is found to scale mainly linearly with the implanted dose (fig. 4.5a and 4.6a). Around 40% of the implanted dose segregate to the pile-up. In the presence of fluorine (BF_2 implant), this number is reduced to values around 20%.

Figure 4.6b displays the pile-up dose as a fraction of the retained dose after anneal. This graph analyzes the distribution of the dopants between the pile-up and the junction's bulk part, for a given retained dose. Any dopant atoms lost during implant and/or anneal are not considered anymore. It is interesting to note that here, no difference is seen between B and BF_2 implants. This suggests that the dopant distribution between pile-up and bulk junction depends only on the total amount of dopants present, but not on the presence of fluorine. It is likely that fluorine neither influences the diffusion nor the dopant trapping at the interface. This is consistent with the suggestion mentioned above, that the reduced dose in the BF_2 case is only due to the higher self-sputtering rate, and not to a possible change in outdiffusion during anneal.

When considering these values, one has to take into account that the dose loss during implant and anneal becomes more important with increasing dose (cf. fig. 4.2 and 4.5b). As a consequence, the relative distribution of the dopants between pile-up and the bulk part of the junction is shifted towards the pile-up for higher implanted doses. Fig. 4.6b shows that for a 3e15 boron implant (2e15 after anneal), two thirds of the remaining dopant atoms are confined to the pile-up, while only one third is activated in the bulk part of the junction. Since the dopants in the pile-up are mostly (perhaps entirely) electrically inactive, this behavior is detrimental to the goal of a low sheet resistance. This trend is also seen in fig. 4.5b as a divergence between the two graphs for high doses.

As discussed already before (section 4.3), dopant loss during implant and anneal causes a dose limitation of something between $5 \cdot 10^{15}$ and $10^{16}\,\mathrm{cm}^{-2}$, for the investigated implant and anneal parameters. Moreover, an extrapolation of the values in fig. 4.5 suggests that the pile-up effect limits the active dose in the junction's bulk part. The maximum achievable values appear to be in the region of 1-$2 \cdot 10^{15}\,\mathrm{cm}^{-2}$, for the investigated combination of a 0.5 keV implant and a 1070 °C, 1 s spike anneal.

It needs to be noted that the ERD results of the sample implanted with 3e15 boron (in all graphs the boron sample with the highest dose) were scaled to fit the SIMS results. The results shown for this sample are thus estimated.

Arsenic For arsenic, data acquisition with RBS was more difficult. Very high doses, sometimes above the nominally implanted dose, have been measured. Therefore, the 2 keV As values presented in figures 4.5 and 4.6 have been calibrated using SIMS profiles – A correction factor was applied to the entire profile to better fit to the concentration values seen in SIMS. However, since the RBS profile depth of 11 nm only goes little beyond the pile-up, the values presented here are considered accurate to only around $\pm 20\%$. All three 2 keV As profiles were measured together in one measurement series, therefore their relative trend and their qualitative evaluation is believed correct. The 5 keV sample was measured in a different RBS run and therefore calibrated separately.

Similar to the boron samples, no significant difference is expected between the pile-up results for the different implant energies. The fact that the 5 keV point lines well up with the other arsenic samples therefore indicates that the chosen calibration factors are indeed consistent.

Fig. 4.5a shows that the pile-up dose increases for increasing implanted dose, however less so than for boron. While for B, the percentage of the implanted dopants accumulated in the pile-up is observed quite independent from the implanted dose, this is not the case for arsenic. Instead, the relative dose of the pile-up drops for increasing implanted dose (fig. 4.6a,b).

The results suggest that the pile-up formation mechanism for arsenic is stronger than for boron. Already at low doses, a large pile-up is formed. Additionally implanted dopants contribute to a lesser extend to the pile-up, indicating a (weak) pile-up saturation effect. At an implanted dose of $4 \cdot 10^{14}\,\mathrm{cm}^{-2}$, around $(70 \pm 15)\%$ of all retained dopants are observed in the pile-up. Even though the error of this value is very large, the results demonstrate that contrary to the boron case, the arsenic pile-up is critical at the lower end of the dose range.

For doses below $4 \cdot 10^{14}\,\mathrm{cm}^{-2}$, the relative amount of dopants in the pile-up is – obviously – expected to reach a certain limit. This limit should be determined in further experiments, that are also recommended to quantitatively confirm the discussed effects.

Contrary to boron, the trend in fig. 4.5b does not reveal any indications about an upper limit for the activated dose in the junction.

It should be noted that all results discussed in this section, for boron as well as for arsenic samples, are also dependent on the other implant and anneal parameters, such as implant energy, type of screening oxide, thermal budget during anneal and so on. The influence of these parameters will be evaluated in the following sections.

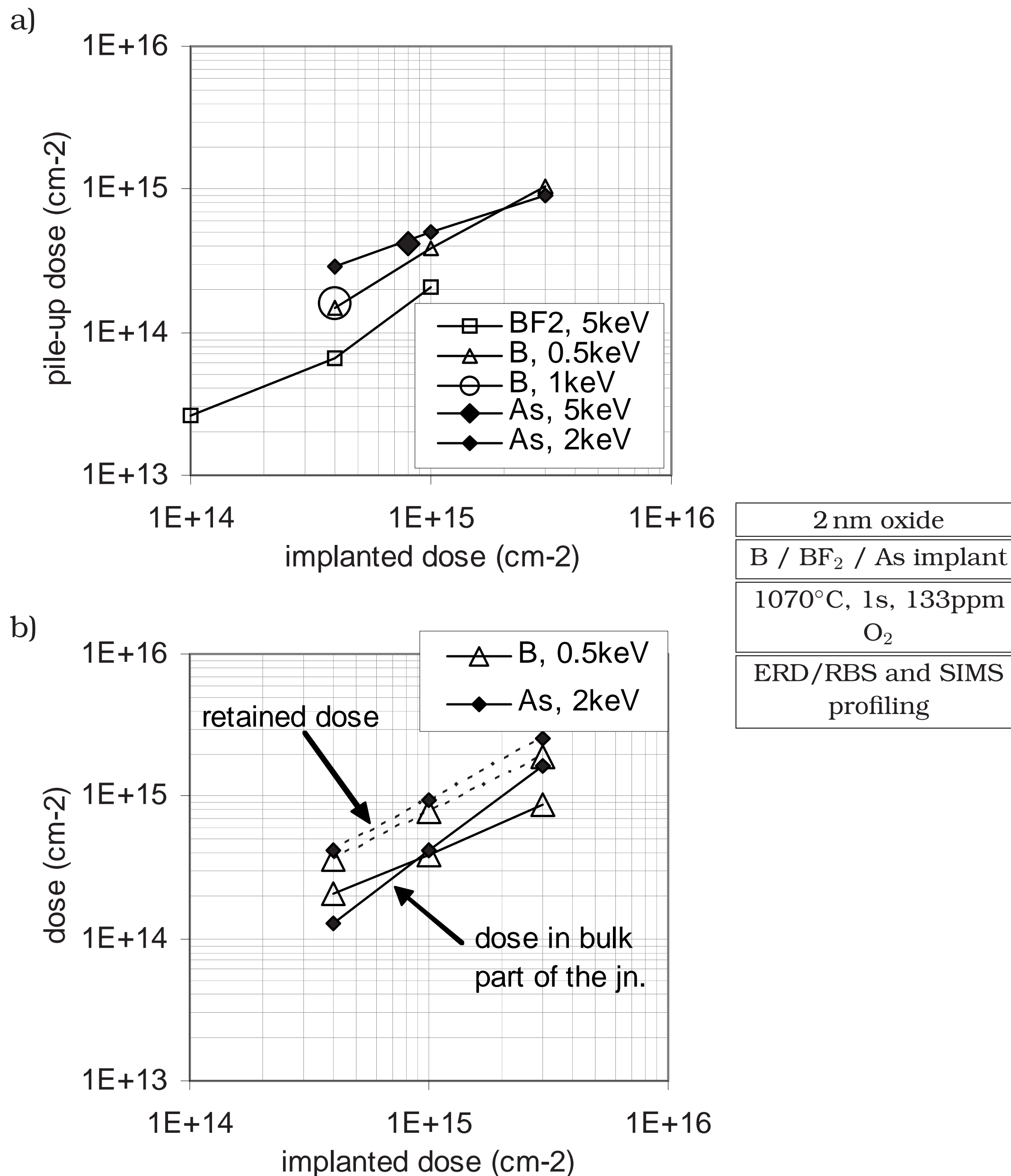

Figure 4.5: a) Pile-up dose as a function of implanted dose for different implant species.
b) Total retained dose and bulk dose after anneal. The difference between the two values is the pile-up dose, cf. graph (a)
These values were extracted from ERD/RBS and are partially corrected using SIMS results – see text for details, esp. for the As results. (Samples from lots E010731, T020208, P020465, P030122)

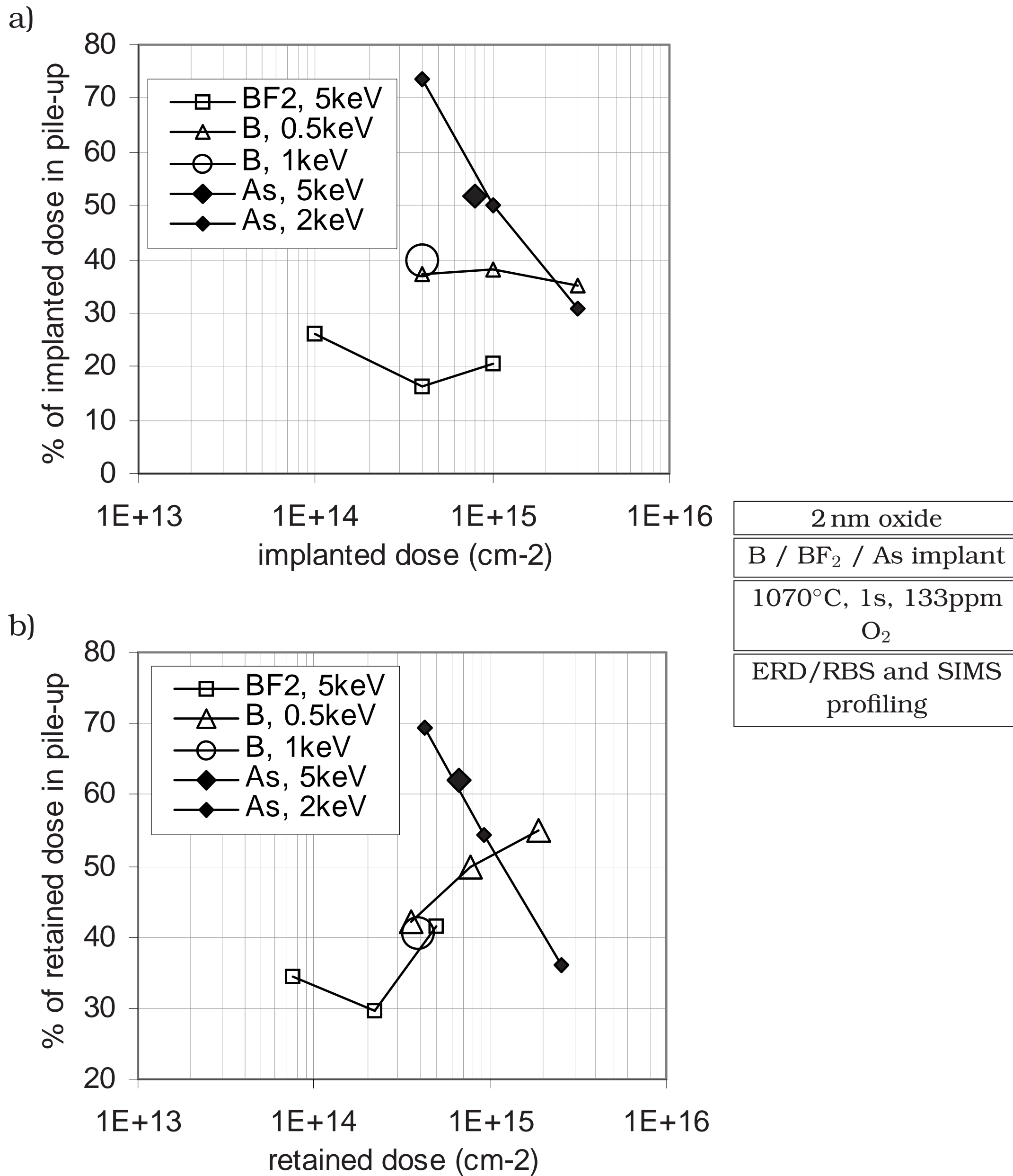

Figure 4.6: Pile-up dose given as percentile fraction of the total implanted dose (a) and of the dose retained after anneal (b). These values were extracted from ERD/RBS and are partially corrected using SIMS results – see text for details, esp. for the As results. (Samples from lots E010731, T020208, P020465, P030122)

Dopant	B			As		
implanted dose	4e14	1e15	3e15	4e14	1e15	3e15
after anneal:						
total dose by SIMS	3.5e14	7.7e14	1.9e15	4.2e14	9.2e14	2.6e15
total dose by ERD *	2.5e14	5.7e14	2.7e15	7.6e14	1.2e15	3.1e15
pile-up dose by ERD *	1.5e14	3.8e14	1.9e15	6.5e14	1.0e15	1.8e15
HWHM pile-up / nm	1.20	0.95	2.00	0.60	0.65	0.80
depth resolution / nm	0.95	0.65	0.85	1.00	0.85	0.85

Table 4.2: Influence of the implanted dose: Resolution and dose data extracted from the profiles in figs. 4.7 and 4.8. Please note that As was measured with RBS, therefore the As profiles have a better resolution than the oxygen ERD profile. (Samples from lots T020208, P020465 and P030122)
** Dose values are given as observed with ERD/RBS, no correction with SIMS data was made. Also, the ERD/RBS "total dose" was integrated only over the investigated depth (26 nm for B, 11 nm for As), not over the entire profile.*

Discussion of the profile shape

Figs. 4.7 and 4.8 show ERD/RBS profiles of samples implanted with B or As at doses between $4 \cdot 10^{14}$ and $3 \cdot 10^{15}\,\mathrm{cm}^{-2}$. Please note the changing scale for the dopant profiles on right side of each graph. A clear dependence of the pile-up height and thus also of the pile-up dose on the implanted dose is seen, as it was already discussed above.

Boron A significant broadening of the pile-up towards the bulk is seen for doses above $10^{15}\,\mathrm{cm}^{-2}$, more so for B than for As. For the 3e15 B sample (fig. 4.7c), the concentration of dopant atoms continuously diminishes until it reaches the background level in a depth of around 10 nm.

Strictly speaking, these dopants are not part of the actual pile-up, since they are not trapped at the interface. However, since they are part of the phenomena discussed here, they are included in the term "pile-up" in this work. Also, the dose values given in the tables and figures always include the entire dopant peak around the interface, independent of their microscopic location and trapping mechanism.

The ERD resolution in fig. 4.7c does not allow a clear differentiation between the dopants at the interface, and the ones on both sides next to the interface. Taking the depth resolution extracted from the oxygen profile and a dose correc-

tion using SIMS data into account, one can approximate the measured profile as a delta peak at the interface and a region of (relatively) low concentration around, convoluted with a Gaussian-shaped resolution function. The result of this investigation suggests that the dose trapped at the interface itself is in the order of $5\text{-}6 \cdot 10^{14}\,\mathrm{cm}^{-2}$.[2] This corresponds to almost one monolayer ($1\,\mathrm{ML} \approx 7 \cdot 10^{14}\,\mathrm{cm}^{-2}$).

This result suggests that the interface can trap almost as much as one monolayer of dopant atoms. However it is very probable that these dopants are not actually confined to one monolayer, but spread over 2-3 monolayers.

Consequently, it seems also plausible that for a 3e15 implanted dose, the dopant trapping capacity of the interface is exceeded. The excess dopant atoms are then pushed back below the interfacial region, where they form the observed extension of the pile-up into the silicon. It is likely that these dopants are clustered, which prevents them from further diffusion.

A possible additional effect is proposed, based on observations of Agarwal et al. [86]. They saw significant BED originating from highly doped layers, if their B concentration exceeds a threshold of “a few atomic percent”. In the BED effect, excess interstitials are injected into the silicon, where they cause enhanced diffusion. In the present case, the pile-up itself could inject silicon interstitials into the bulk. By transporting some dopant atoms along (similar to the uphill diffusion effect), this could lead to the observed broadening of the pile-up. Additional experiments should be run to further investigate this effect.

In all three boron samples, the pile up extends almost to the surface of the sample. This tail is explained by dopant atoms that are implanted into the oxide and stay there during anneal, because of the low diffusivity of B in oxide.

This low diffusivity of B atoms in the oxide is an apparent contradiction of a common effect in CMOS processing: Boron penetrates from the gate electrode (B doped poly-Si) through the gate oxide, which leads to a change in the cannel doping and thus in the transistor’s threshold voltage V_{th} [112]. In our case, however, the thermal budget of the spike anneal is much smaller than the one of a typical hour-long furnace anneal used in older CMOS generations. This greatly reduces the diffusion of B through the oxide [33].

Until now, no differentiation was made between dopant loss due to self-sputtering during implant, and dopant outdiffusion during anneal. For the samples investigated, it cannot be excluded that the dopants in the oxide are subject

[2] For this calculation, the ERD profile was calibrated using SIMS results.

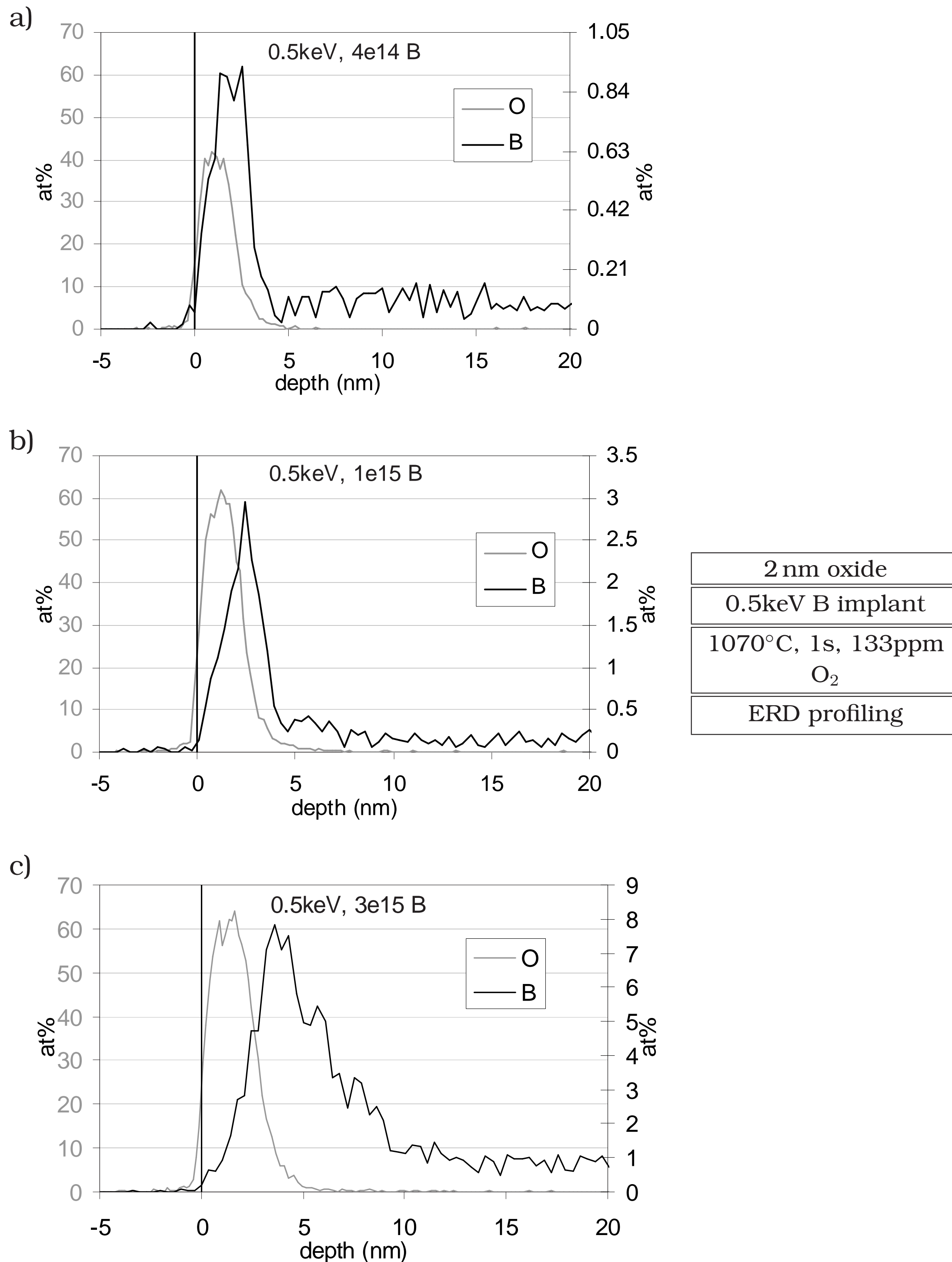

Figure 4.7: Influence of the implanted dose for boron implants. (Samples from lot T020208)

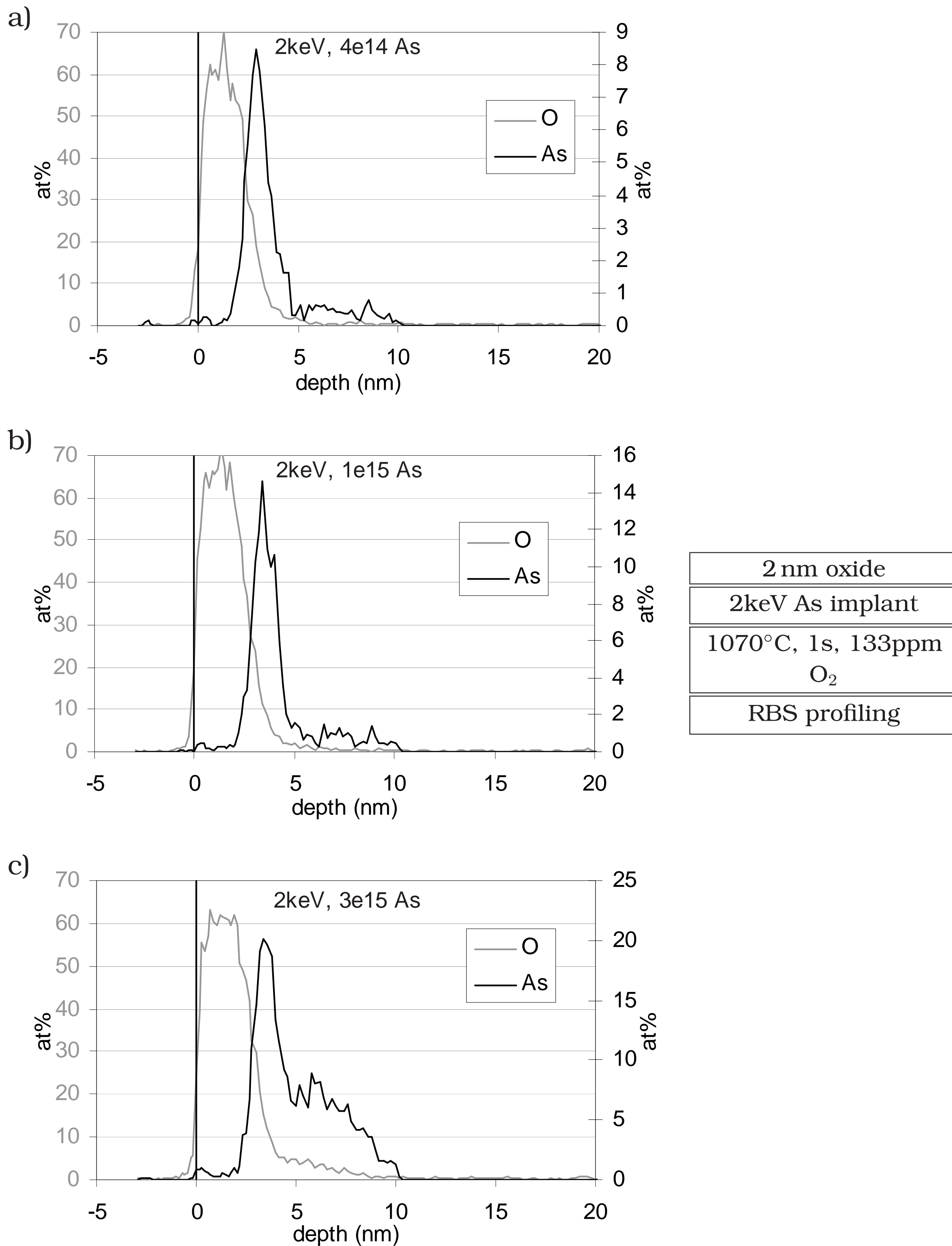

Figure 4.8: Influence of the implanted dose for arsenic implants. (Samples from lots P020465 and P030122)

to diffusion trough the oxide and loss to the ambient during anneal. Close to the surface, this would lead to a profile shape similar to the ones observed.

However, if this diffusion was large, it would either lead to a homogenous dopant distribution in the oxide, or transport a large fraction of the dopants to the surface from where they are desorbed to the ambient. In both cases, a step between the pile-up and the dopant profile in the oxide would be created, since the pile-up dopants at the interface are not taking part in the general diffusion, as already shown before. This step would look similar to the right-hand slope of the pile-up towards the silicon, resulting in a quite symmetrical shape, similar to the arsenic pile-up. Since on the investigated samples, no such step is seen, the dopant diffusion through the oxide is believed to have only a minor effect on the dopants. It is believed that the boron atoms can be considered as mainly immobile in the oxide for the short spike anneals used.

Arsenic Also for the high-dose As sample, a significant broadening of the pile-up into the silicon is observed. It appears as a shoulder to the pile-up, with a concentration of around 7 at%, instead of the continuously decreasing dopant concentration that was observed for the boron case. It is likely that again, the pile-up exceeds some dose limit, and dopants are pushed into the bulk, where further diffusion is prohibited due to clustering. Further experiments comparing ERD/RBS profiles of even higher doses should be done to better understand this pile-up broadening phenomenon.

For the doses of 10^{15} and below, no such broadening is seen. The right-hand slope of the dopant pile-up is quite parallel to the one of the oxygen profile. This suggests that within the limits of ERD resolution, the pile-up does not extend into the silicon, at least by no more than the pile-up width (estimated to less than 2-3 ML in section 4.4.1). The same statement holds for the oxide side of the interface, since the pile-up profile is highly symmetrical.

Apart from the shoulder in the 3e15 case, all three shown As profiles exhibit a very similar pile-up shape. The measured pile-up width increases slightly with higher doses. The ERD/RBS profile resolution, however, is not sufficient to prove or disprove such small variations in the order of 0.1 nm. Considering these profiles as well as the other results presented in the following sections, it is likely that the arsenic pile-up is actually confined to one monolayer.

[4] A BF_2 implant with energy E is equivalent to a B implant with energy $0.22E$, and a F implant with double dose at the energy $0.39E$.

Species	B	B	BF_2	As	As
Energy (keV) (Equivalent B energy)	0.5	1	5 (1.1)	2	5
Projected Range R_p (nm)	2.6	4.8	5.3	4.3	7.4

Table 4.3: Projected range R_p for B, BF_2 and As implants in a Si sample with a 2 nm surface oxide layer.[4] The values were obtained from Monte-Carlo simulations with SRIM [21].

Dopant	B		As	
implanted dose implant energy	4e14 0.5 keV	4e14 1 keV	1e15 2 keV	8e14 5 keV
after anneal:				
total dose by SIMS	3.5e14		9.2e14	6.0e14
total dose by ERD *	2.5e14	3.9e14	1.2e15	6.7e14
pile-up dose by ERD *	1.5e14	1.6e14	1.0e15	4.2e14
HWHM pile-up / nm	1.20	1.20	0.65	0.80
depth resolution / nm	0.90	0.90	0.85	0.80

Table 4.4: Influence of the implant energy: Resolution and dose data extracted from the profiles in figs. 4.9 and 4.10.
** Dose values as observed by ERD/RBS, no SIMS correction was done. For the 1e15 As sample it is believed that the values given are by a factor of 2 too high (see text).*

4.4.3 Energy dependence

The following section will focus on the influence of the implant energy (i.e. the depth of the implant) on the pile-up shape after anneal. Table 4.3 gives an overview over the projected ranges for the implant energies used during this work.

Figures 4.9 and 4.10 compare boron or arsenic implants for the different energies described in table 4.3. It should be noted that the two arsenic samples (fig. 4.10) were processed in two different lots. In the 5 keV case, a clean was done between implant and anneal. It is likely that this clean is responsible for the fact that the sample's oxide layer is thinner than in the 2 keV case.

For arsenic, a clear uphill diffusion during anneal is observed. The same is the case for the deeper B profile (4.9b), but the effect is less pronounced.

For the shallow B (4.9a), no obvious uphill diffusion is observed any more, the left-hand side of the pile-up looks very similar the the as-implanted profile. For

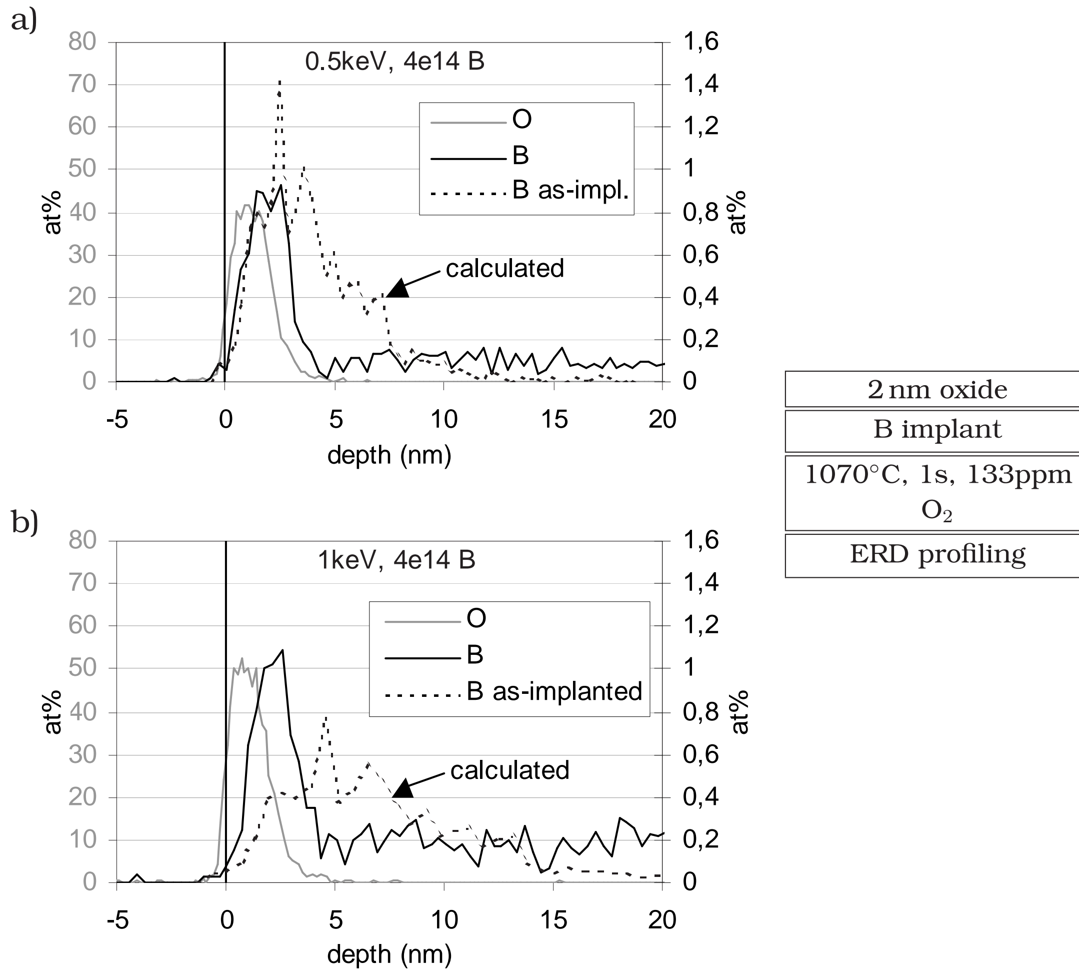

Figure 4.9: Influence of the implant energy for boron. The dashed lines indicate the as-implanted dopant profiles. The as-implanted profiles in subfigures a) and b) were calculated (scaled) from a 0.5 keV, 1e15 B sample. (Samples from lot T020208)

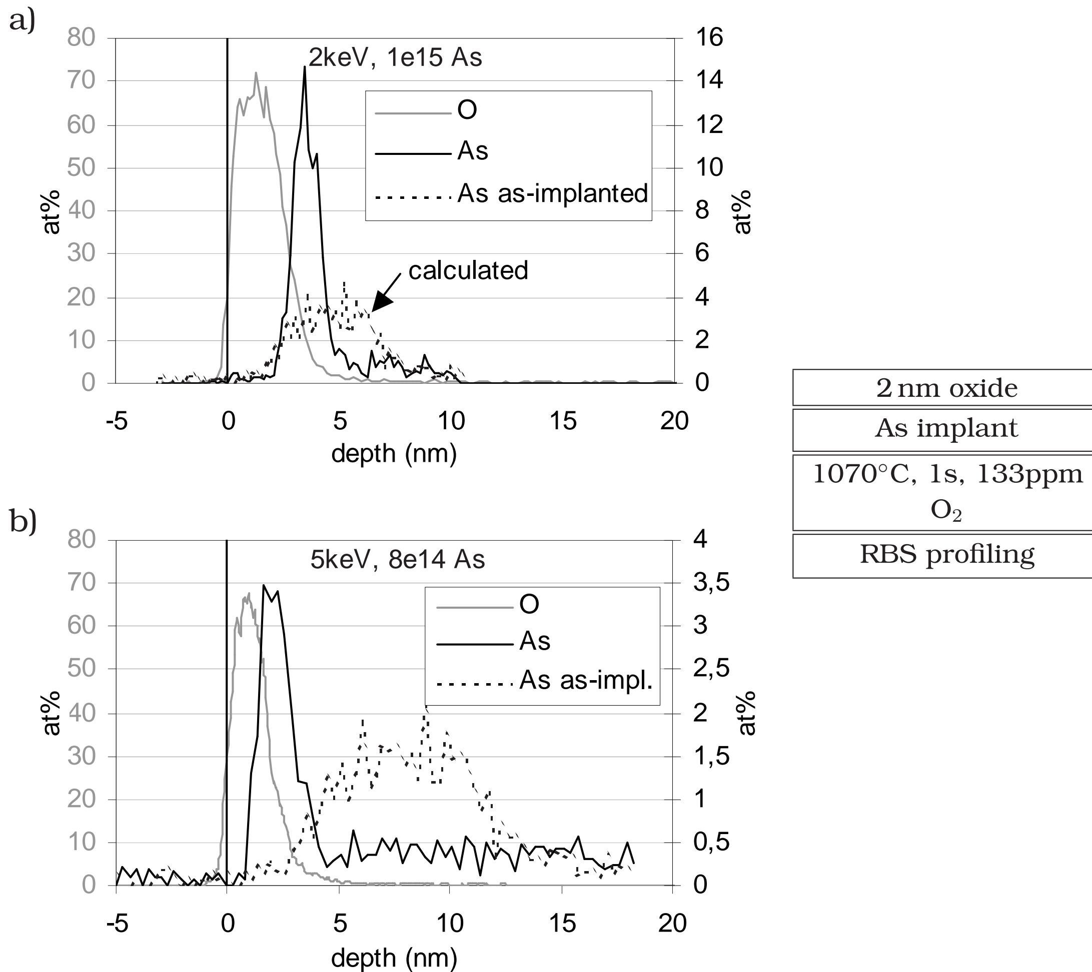

Figure 4.10: Influence of the implant energy for arsenic. The dashed lines indicate the as-implanted dopant profiles. The as-implanted profile of subfigure a) was calculated (scaled) from the 5keV, 8e14 As sample in b). Note the slightly different doses in a) and b). In fig. a, it is believed that, for an unknown reason, the RBS data overestimate the real arsenic concentration profile by a factor of around 2 – see the description of the calibration procedure, p. 69. The graph displayed here shows the original data as recorded in RBS, no calibration is applied. (Samples from lots E010731 and P020465)

this sample, an alternative pile-up formation process has to be considered: Since boron has a low diffusivity in oxide, the dopant atoms might just stay where they are placed during implant, whereas the dopant concentration on the Si side of the interface is reduced due to diffusion.

However, since there is no fundamental physical difference between the samples if figs. 4.9a and b, it is believed that also in the shallow, high dose boron case (fig. 4.9a), the same uphill diffusion and trapping mechanism occurs as seen in the deeper case. Additionally, normal diffusion from the as-implanted profile's peak to regions of lower concentration occurs, compensating the uphill diffusion flux. Therefore, the pile-up formation process is not obvious any more.

The two B implants with different energy result in a pile-up of almost identical shape and dose. The results suggest that neither the initial distribution of the dopants in the bulk silicon nor the presence of as-implanted dopants in the oxide change the pile-up formation mechanisms, for the investigated energy range of 0.5-1 keV. The uphill diffusion seems to be strong enough to level out any energy-dependent differences. It is therefore suggested that the dopant atoms implanted into the oxide participate in the pile-up formation just as the other dopant atoms do.

This reasoning also holds for arsenic. In fig. 4.10, No significant influence of the energy is observed. Considering that the concentration profiles in fig. 4.10a are believed to be by a factor of 2 too high for an unknown reason (see the description of the calibration procedure, section 4.4.2), the pile-up of the 2 keV sample shows a very similar dose to the one of the 5 keV sample. This is also seen in the discussion of samples with different doses, figs. 4.5 and 4.6.

It has to be noted that the calibration factor is considered to be correct to only ±20%, as discussed in section 4.4.2. Since the two samples under discussion here were processed in two different RBS measurement runs, this error cannot be avoided.

4.4.4 Different screening oxides

For sub-100 nm CMOS generations, nitrided oxides and ultimately high-k materials replace the pure oxide as a gate dielectric. It is therefore important to investigate the influence of nitrogen in the surface oxide on pile-up dose and shape.

Three different types of oxides were used:

- Pure oxide: A standard gate-quality dry oxide, grown in a furnace in a 35 min process at 750 °C. The total processing time is 2 hrs at 750 °C.

Dopant	B			As		
implanted dose oxide (2nm)	1e15 O	1e15 NO	1e15 DPN	1e15 O	1e15 NO	1e15 DPN
after anneal:						
total dose by SIMS	7.2e14			9.2e14		
total dose by ERD *	8.7e14	1.1e15	1.2e15	1.2e15	1.2e15	1.3e15
pile-up dose by ERD *	5.6e14	7.5e14	9.1e14	1.0e15	1.0e15	1.1e15
HWHM pile-up / nm	0.80	0.95	1.30	0.65	0.65	0.65
depth resolution / nm	0.85	0.70	0.85	0.85	0.70	0.85

Table 4.5: Different screening oxides: Data extracted from the profiles in figs. 4.11 and 4.12.
** Values as observed by ERD/RBS, no SIMS correction was done.*

- Oxynitride : This nitrided oxide process starts with the same process as described above for the pure oxide, only in the end of the oxidation process, NO gas is added to the furnace ambient. During oxidation, O or N atoms diffuse through the oxide to the oxide/silicon interface, where they bind to Si atoms. Therefore, the nitrogen is included in the oxide very close to the oxide/silicon interface for this specific process. The resulting total dose of nitrogen is around 8% of the oxygen dose ($\sim 7 \cdot 10^{14}\,\mathrm{cm}^{-2}$).

- Heavily nitrided oxide: After growing an oxide in a furnace, a nitrogen plasma is ignited. Nitrogen atoms from the plasma are implanted into the wafer (Decoupled Plasma Nitridation, DPN). A nitrogen dose of around 14% of the oxygen dose is measured ($\sim 1.3 \cdot 10^{15}\,\mathrm{cm}^{-2}$).

For convenience, the samples with these three oxide types are labelled “O”, “NO” and “DPN”, respectively. Samples with the three oxide types were investigated with ERD/RBS. The influence of the nitrogen on the profile shape is fundamentally different for boron (fig. 4.11) and arsenic (fig. 4.12).

For As, no significant difference is seen between the samples with and without nitrogen. The pile-up doses vary by only 7%, less than the expected statistical error. With an HWHM of 0.65 nm in all cases, the As pile-up is sharper than expected from the oxygen profile shape (depth resolution 0.70-0.85 nm). The pile-up atoms are thus confined to not more than a few monolayers in all cases.

Furthermore, the As dopants are observed to leave the oxide during anneal, as it will be discussed later in more detail (section 4.5.3). The presence of nitrogen is

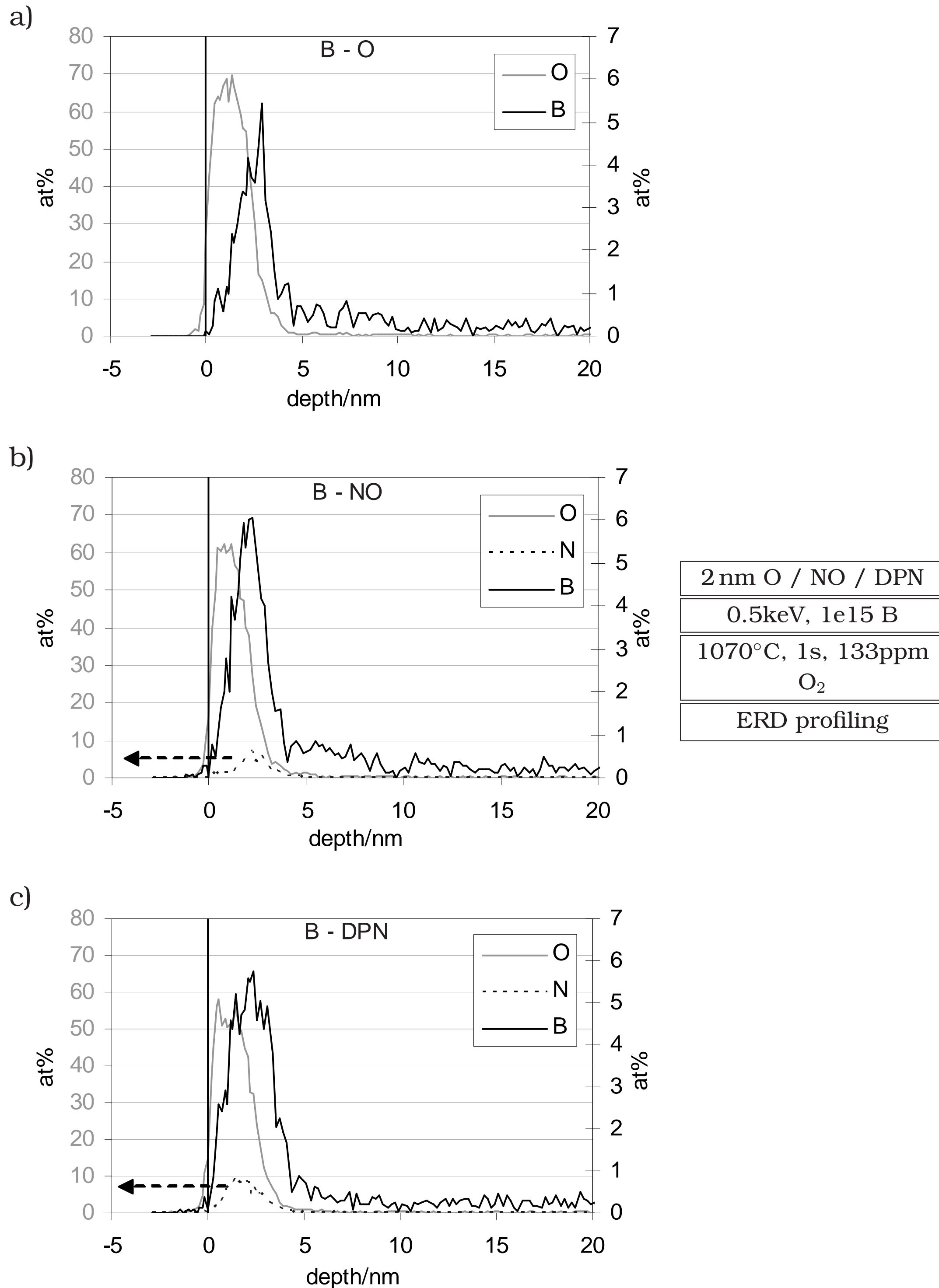

Figure 4.11: Comparison of normal oxide (O), oxynitride (NO) and a heavily nitrided oxide (DPN) for boron implants. The N profiles are in scale to the oxygen profiles, i.e. they are aligned to the left concentration scale. (Samples from lots P020465 and P030122)

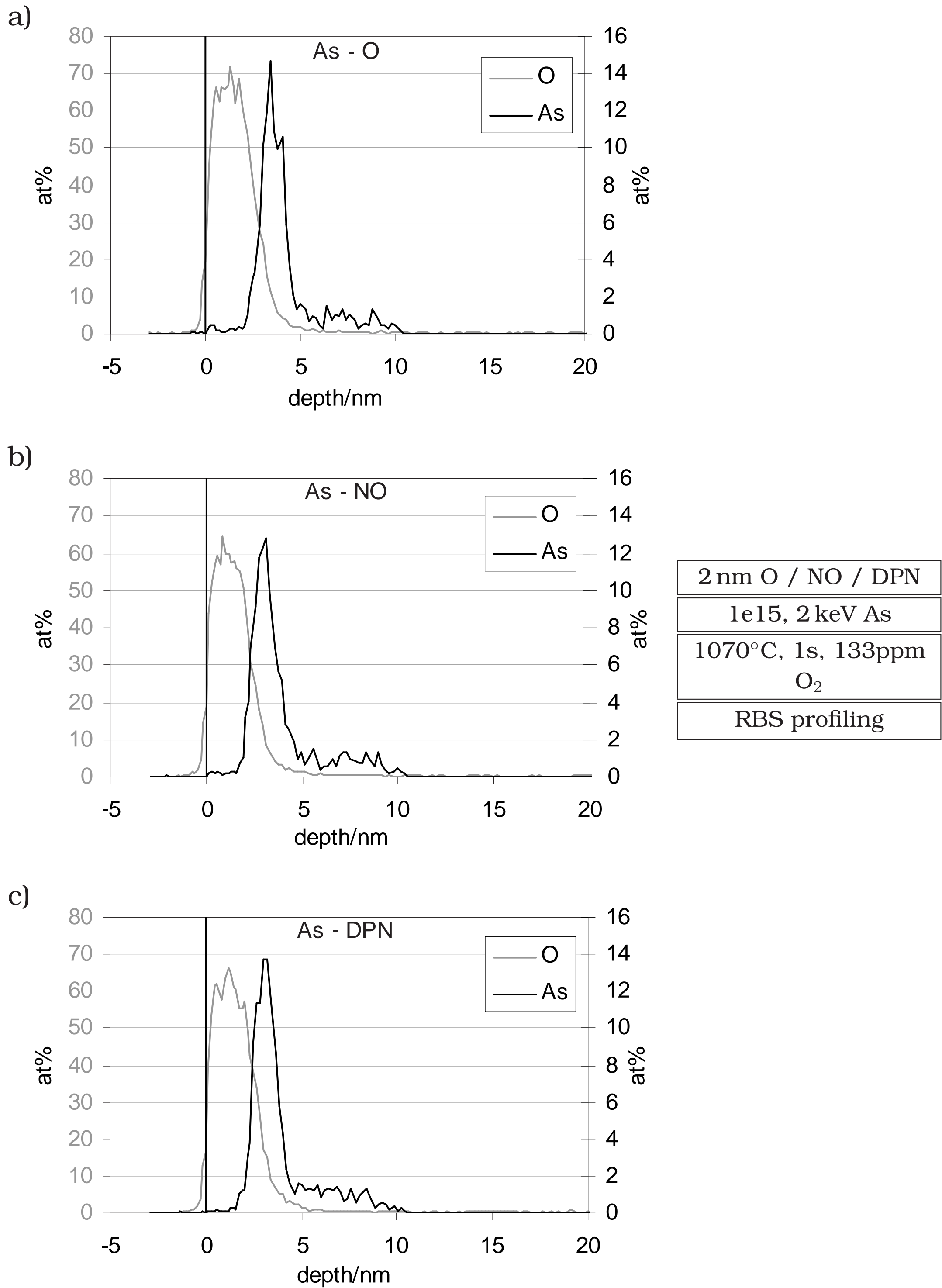

Figure 4.12: Comparison of O, NO and DPN for arsenic implants. Nitrogen profiles are not available for these samples, but are believed to be identical to the ones seen on the B samples in fig. 4.11. (Samples from lots P020465 and P030122)

not changing this effect, nor does it influence the dopant's piling-up behavior at the interface.

In the case of the B implanted samples, a clear difference is seen between the profiles. While the pile-up is observed at the interface in the pure oxygen (O) case, it extends more into the oxide for the NO, and even more for the DPN samples. The HWHM varies between 0.80 (O), 0.95 (NO) and 1.30 nm (DPN). The pile-up doses measured by ERD (uncorrected) are observed between 5.6e14 (O), 7.5e14 (NO) and 9.1e14 (DPN), whereas the dose seen in the bulk part of the junction (determined for the layer between 5 and 15 nm of depth) is very similar for all samples (3.0e14 to 3.4e14). These numbers show that the presence of nitrogen in the oxide prevents the loss of boron to the ambient; instead, the dopants are retained in the oxide.

The results suggest that the presence of N reduces the mobility of boron in the oxide. Boron atoms that would otherwise be lost due to outdiffusion through the oxide are blocked in the oxide, at least to a greater extend than in the case of a pure oxide without nitrogen. The presence of nitrogen thus reduces outdiffusion during anneal. Instead, the dopant atoms accumulate in areas where N is present. This is supported by the fact that in fig. 4.11, the part of the pile-up in the oxide follows the N distribution quite well.

The effect that N reduces the mobility of boron in oxide has been extensively described before (e.g. [113–115]). In a transistor, boron penetration from a B-doped poly-Si gate through the gate oxide can change the threshold voltage V_{th}. To reduce this effect, nitrided gate oxides are used in commercial CMOS manufacturing.

4.5 The role of oxidation during anneal

Having discussed the influence of the implant parameters on pile-up formation, this section will focus on the anneal. The influence of oxygen in the annealing ambient and its importance for pile-up formation will be discussed.

4.5.1 Oxygen in the annealing ambient

As described already above, most of the samples in this work were annealed in a nitrogen ambient with 133 ppm oxygen, to avoid oxide loss during anneal. In order to investigate the influence of this concentration on the further oxidation of the wafer surface during anneal, a series of samples were annealed with various oxygen content in the annealing ambient.

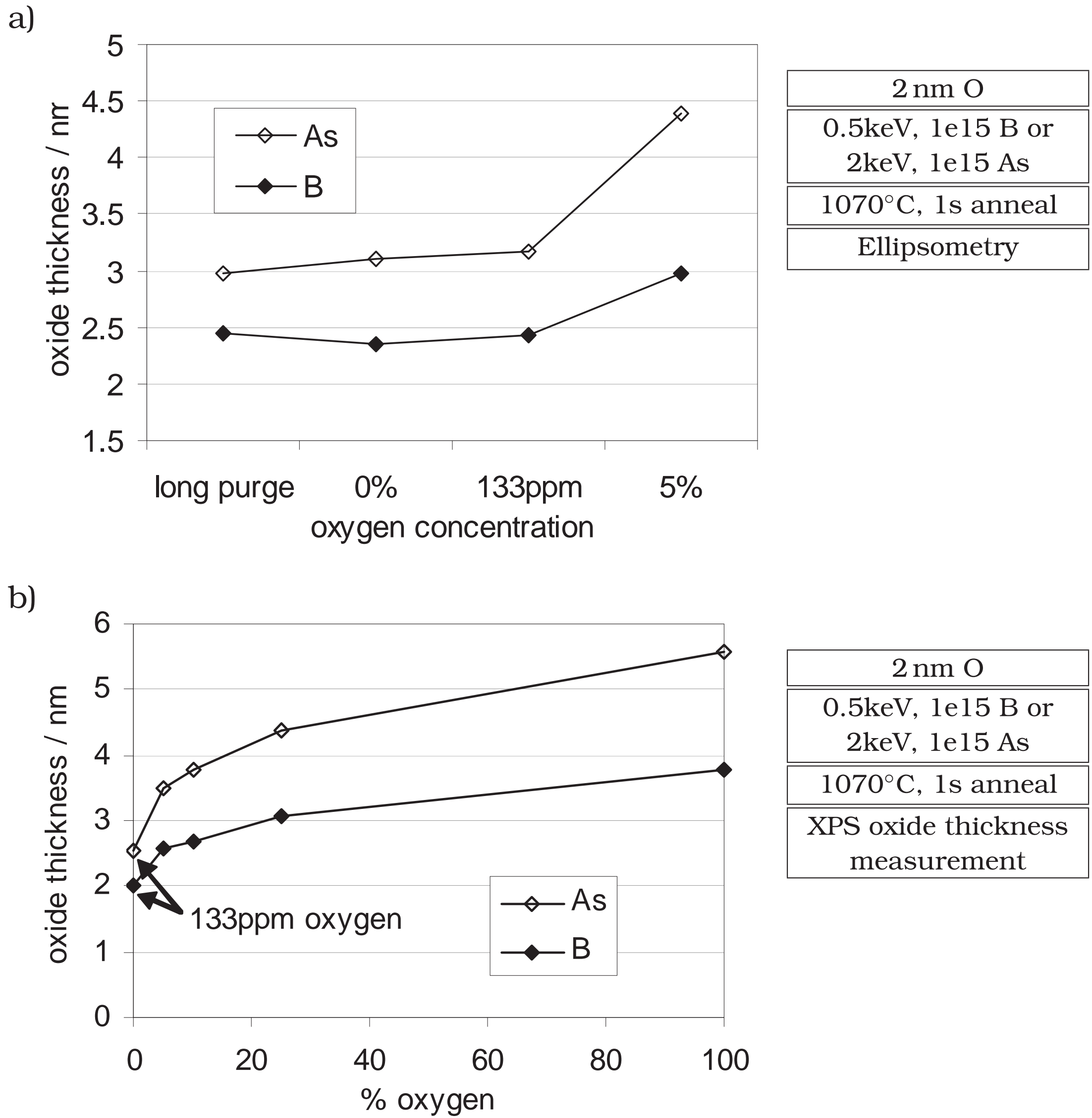

Figure 4.13: Oxide thickness after anneal, as a function of oxygen content in the N_2 annealing ambient. The initial oxide thickness was around 2 nm. Ellipsometry typically overestimates the oxide thickness, compared to XPS. This causes the slight difference between (a) and (b) in the results seen for the 133 ppm and 5% cases. (Samples from lots P020465, P020712 and P030131)

Fiory et al. [116] observed an oxide growth of 1-1.5 nm on wafer without a native oxide, during a spike anneal in pure oxygen ambient. For these thin oxides, the thickness depends linearly on oxidation time, and is modelled by

$$d_{Ox} = \frac{a\, t}{T}\, e^{-E_A/kT} \tag{4.1}$$

with a proportionality factor a and an activation energy E_A that are found experimentally.

For our purposes, neither XPS nor ellipsometry allow a precise direct observation of the oxide growth during anneal, since both methods do not offer a satisfactory possibility to measure the oxide thickness before the anneal: The ellipsometry values are influenced by the implant damage. This damage is (partly) healed during anneal, therefore measurements before implant and after anneal are not easily comparable. On the other hand, for the XPS system available, the wafer needs to be broken in small samples. After this, no further processing is possible, because the lamp-based RTP tool available is not able to handle such small samples. Furthermore, cleaving of wafers is always creating particles, therefore cleaved wafers are not allowed in the cleanroom anymore. As a best possible solution, the experiment described in the following compares ellipsometry results measured before implant and after anneal.

1070 °C, 1 s spike anneals with 0%, 133 ppm and 5% oxygen in an N_2 ambient were compared. Additionally, a 0% oxygen anneal with a 5 min. N_2 purge of the annealing chamber ("long purge") before temperature ramp-up was done. The results are displayed in fig. 4.13a. A comparison with XPS measurements on the 133 ppm samples showed that the ellipsometry values are between 0.4 and 0.6 nm too high for the measurements after anneal for all thicknesses. The oxide thickness before implant and anneal was around 2.2 nm, which is correct to around 0.2 nm, due to the good quality of the original oxide.

Fig. 4.13a shows that for the As samples, oxidation occurs in all cases, even after a 5 min purge. It is likely that oxygen adsorbs to the annealing chamber walls during wafer loading and unloading. Even if after loading, the chamber is purged to a very low oxygen concentration, desorption during the heating sequence increases the oxygen concentration to a significant level. This was also observed by the gas sensors in the annealing chamber.

For B, the long purge sample shows even a higher oxidation than the 0% sample. This indicates that the repeatability of the ellipsometry measurement and/or the experimental setup is not better than 0.2 nm. Therefore, it is not possible to

Dopant	B		As	
anneal ambient	133ppm	5%	133ppm	5%
implanted dose	1e15	1e15	1e15	1e15
total dose by SIMS	7.2e14	7.6e14	9.2e14	8.8e14
total dose by ERD *	8.7e14	9.0e14	1.2e15	1.1e15
pile-up dose by ERD *	5.6e14	6.5e14	1.0e15	9.4e14
HWHM pile-up / nm	0.80	1.25	0.65	0.85
depth resolution / nm	0.85	0.95	0.85	1.25

Table 4.6: Resolution and dose data extracted from the profiles in figs. 4.7 and 4.8. Please note that As was measured with RBS, therefore the As profiles have a better resolution than the oxygen ERD profile. (Samples from lots T020208, P020465 and P030122)
** Values as observed by ERD/RBS, no SIMS correction was done.*

distinguish in detail the amount of oxide grown on the “long purge”, the 0% and the 133 ppm sample for B. However, since the chamber is proven not to be entirely oxygen-free during anneal, a small amount of oxidation of less than 0.2 nm is expected also for the B samples – this corresponds to less than one Monolayer of silicon being oxidized.

In the 5% case, clearly around 0.5 nm (B) to 1 nm (As) of oxide has grown during anneal. As displayed in fig. 4.13b, a further increase in the oxygen concentration in the ambient leads to a thicker oxide. However, the oxidation enhancement is not linear with the oxygen concentration.

In general, As implanted samples grow more oxide than the B implanted samples do. The enhanced oxidation in the presence of arsenic has also been observed before [75]. It seems to be promoted by the negative doping. A similar effect on the growth rate of a native oxide will be discussed below. It is suggested that in this case, also the stronger implant damage caused by the heavier As atoms facilitates the diffusion of oxygen atoms towards the interface and hence increases the oxidation rate.

4.5.2 Influence of oxidation on the pile-up

ERD was used to investigate the influence of oxidation during anneal on the height and shape of the pile-up. Figures 4.14 and 4.15 compare profiles of samples annealed with 133 ppm and 5%. The pre-implant oxide thickness was again 2 nm.

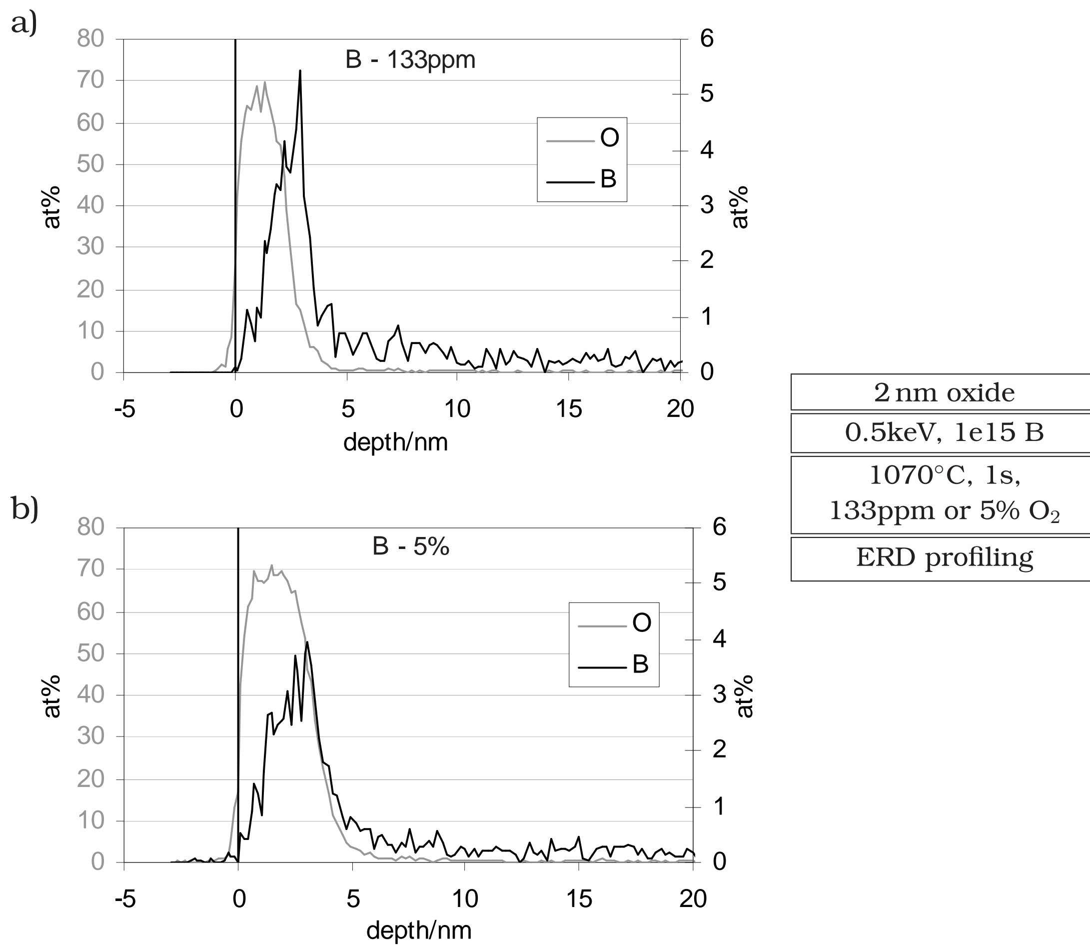

Figure 4.14: ERD profiles of B implanted samples annealed with 133 ppm (top) and 5% (bottom) oxygen concentration in the ambient. (Lot P020465)

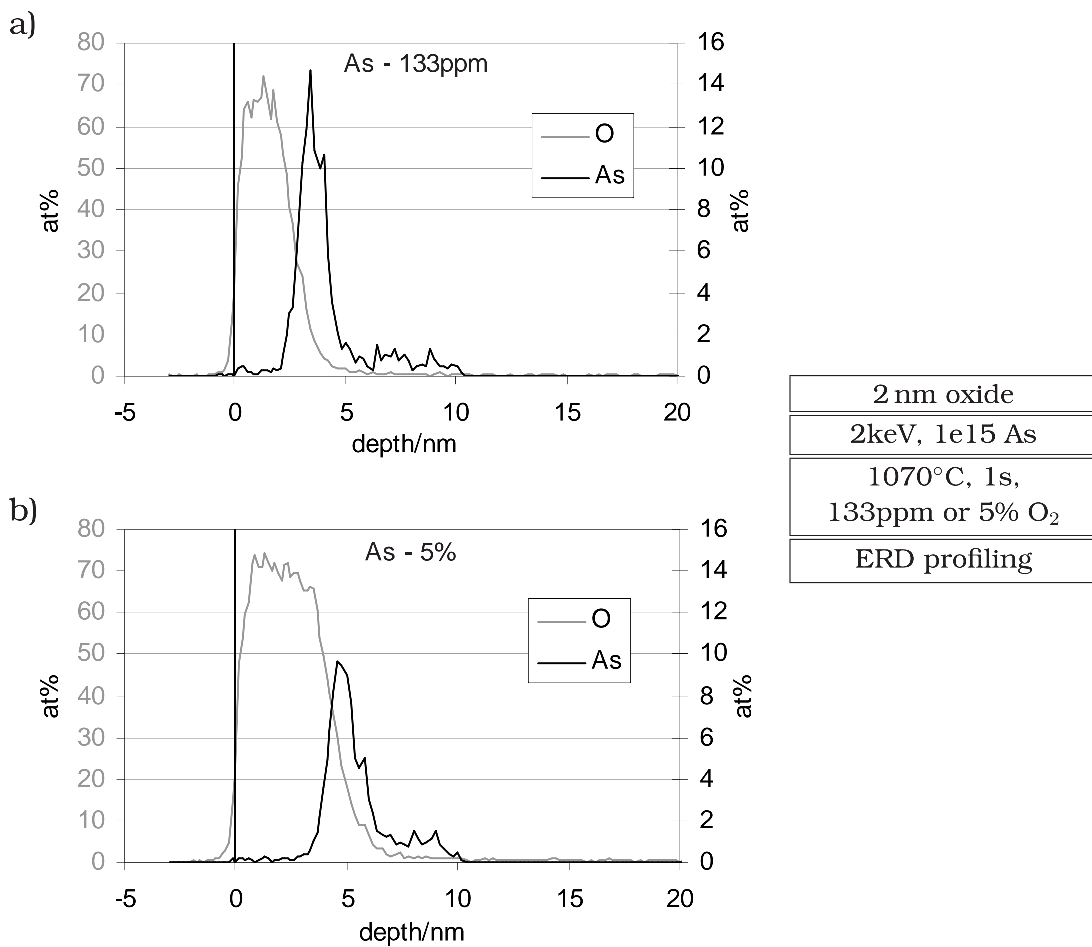

Figure 4.15: RBS profiles of As implanted samples annealed with 133 ppm (top) and 5% (bottom) oxygen concentration in the ambient. (Lot P020465)

The oxide thicknesses, determined from the middle of the right-hand oxygen profile slope, are 2.4 and 3.4 nm for the B samples and 2.7 and 4.4 for the As samples. This shows again that the oxidation during anneal is enhanced for arsenic implanted samples.

In the As case, only a small difference is seen between the 133 pmm and 5% anneals. The doses (measured by RBS) are 1.00e15 and 1.07e15, respectively. In the 5% case, the depth resolution at the interface is worse (1.25 nm instead of 0.85 nm resolution, extracted from the oxygen ERD profile), due to the greater depth as well as to increased oxide thickness variations or interface roughness. Therefore, the As pile-up is measured wider and less high in the 5% case.

The enhanced oxidation is pushing the As pile-up deeper into the sample, following the silicon-oxide interface. Again, no dopant atoms are seen in the oxide after anneal, and the pile-up is confined to a very narrow region around the interface in both cases.

During the oxidation process, arsenic atoms present in the consumed silicon layer contribute to the pile-up formation. This could explain the 7% difference in dose observed between the two samples. However, this difference is too small to be significant; the mentioned effect should be regarded with caution.

No other significant effects are observed from the oxidation during anneal. If oxidation is necessary to form a pile-up, already the oxidation in a the 133 ppm oxygen ambient is largely sufficient in the As case.

For B, however, the increased oxidation causes the pile-up to be wider, while reducing its height. The dose increases very slightly from 8.7e14 to 9.0e14 (measured by ERD), while the HWHM increases from 0.80 nm to 1.25 nm. The additional pile-up width in the 5% case is entirely located on the oxide side of the interface.

As discussed for arsenic, the boron pile-up dose is only influenced marginally by the increased oxidation. However, the increased oxidation does not push the pile-up deeper into the oxide as for arsenic, but instead spreads it over a larger volume. The ERD profiles suggest that this volume is mainly identical to the amount of oxide grown during anneal. As a consequence, the dopant concentration in the pile-up is reduced in the case of stronger oxidation.

Looking at the silicon side of the interface, no significant difference is observed between the two samples, behalf the pile-up's peak concentration. This demonstrates that the pile-up formation process at the interface is only changed quantitatively, but no qualitatively new features are observed.

It is concluded that the arsenic pile-up is always located at the interface, while fore boron, an important fraction of the pile-up is incorporated into the oxide in the 5% case. Since also during a 133 ppm anneal, some oxide is growing, it is suggested that the B pile-up does extend slightly into the oxide in all cases. Its width depends on the amount of oxide grown. Contrary to arsenic, the boron pile-up is thus not confined to a monolayer at the interface.

4.5.3 Anneals without oxidation

In order to further investigate the influence of oxidation on the pile-up formation, two samples were prepared with a 5 nm instead of the usual 2 nm screening oxide. The samples were implanted with higher energy (7 keV, 4e14 BF_2 or 5.5 keV, 1e15 As) to compensate for the thicker oxide. The spike anneal was done under standard conditions (1070 °C, 1s, 133 ppm oxygen). ERD profiles are displayed in fig. 4.16.

The oxide thickness decreased from 5.2 to 4.8 nm during implant and anneal in the BF_2 case and increased from 5.3 to 6.0 nm in the As case. The pre-implant measurements were done by ellipsometry, whereas the post-anneal values were extracted from the ERD oxygen profile. The values must therefore be compared with caution.

For the BF_2 implant, the values suggest that some oxide was lost, probably due to sputtering during implant, and that the oxidation during anneal was very small, if any. Also the basic clean done for the BF_2 after implant might have removed some oxide. However, since a 2 nm oxide is observed to grow by not more than 0.2 nm during a spike anneal, it is believed that for the sample in fig. 4.16a, no oxidation occurred at all.

The B profile, as measured by ERD, exhibits no special pile-up at the interface. Instead, a concentration step from 0.5 at% in the oxide down to around 0.1 at% in the silicon is seen at the interface. This could be interpreted as "classical" segregation behavior; however, the step is higher than the expected concentration ratio of around 2. A comparison to the as-implanted profile, simulated by SRIM [21, 22], shows that the dopant atoms in the oxide have a very similar shape as the as-implanted profile. Thus, these dopants are probably trapped in the oxide. It is concluded that the spike anneal does not provide sufficient thermal budget to significantly diffuse dopants through the thick oxide.

The fact that no pile-up is seen in the thick oxide case demonstrates that oxidation is necessary for the pile-up creation. The oxidation process, seen on an

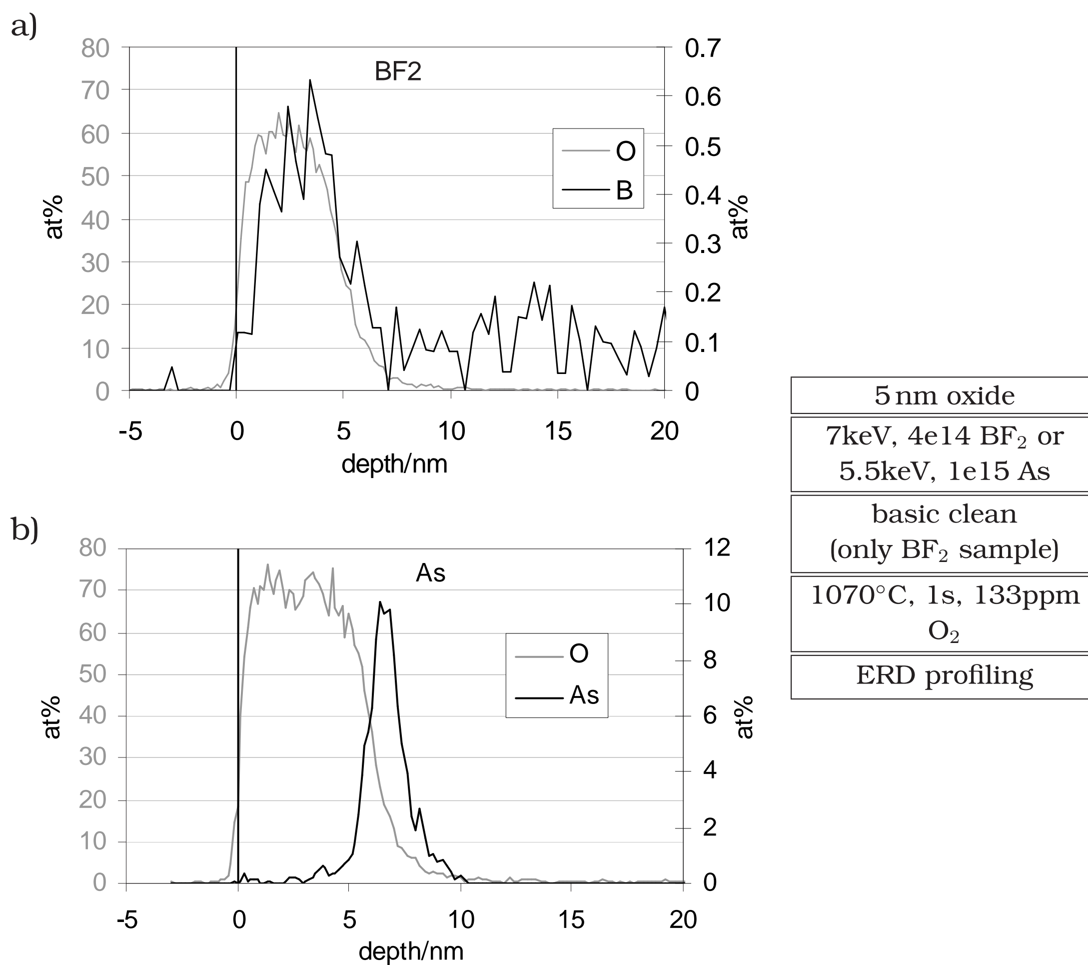

Figure 4.16: ERD/RBS profiles of samples with 5 nm thick oxides. (Lot P010672 and P020465)

atomistic level, is basically a rearrangement of the silicon atoms and an inclusion of additional oxygen atoms. Dangling bonds and other trapping sites are created during that rearrangement process. These sites can trap dopant atoms that are brought to the interface by uphill diffusion. This mechanism, that will be discussed in more detail in section 4.8, is likely to be responsible for the formation of the pile-up.

For the arsenic sample, at least 0.7 nm of oxide have grown during anneal. This shows that the 5 nm oxide is not thick enough to prevent oxidation of As-implanted samples. A pile-up has been created at the oxide-silicon interface that is similar in dose to the ones observed on samples with thin oxide.

Due to the non-avoidable oxide growth, no final conclusions can be drawn for samples implanted with arsenic. However, it is likely that the same atomistic processes as described above for boron are also responsible for the formation of the arsenic pile-up. It is believed that in a sample annealed without any oxidation, no pile-up is formed. Instead, the classical segregation behavior is expected, leading to a concentration step at the interface, if the thermal budget of the anneal is high enough for the sample to reach a thermodynamical equilibrium. Further experiments with anneals in a better controlled ambient are suggested to help clarifying this point.

It is interesting to note that all As atoms leave the oxide during anneal. This is observed on all other samples shown before (e.g. figs. 4.8, 4.10, 4.12), and becomes especially apparent in figure 4.16. The effect shows that the diffusivity of As in oxide is much higher than the one of boron, and that it is sufficient to remove all dopants from the oxide during the very short spike anneal time. It also demonstrates that the presence of arsenic in the oxide is energetically not favorable, as it is expressed classically by a segregation coefficient >1 for arsenic at a silicon-oxide interface. The solubility of arsenic in oxide is thus much smaller than the one of boron.

The effect described also limits the dose loss during anneal. The oxide layer acts as a barrier for dopants that would otherwise be lost due to outdiffusion. This explains why the percentile retained dose is always higher for arsenic than for boron, for identical process conditions.

4.6 Influence of the annealing temperature

The discussion in section 3.4.2 claims that not only normal diffusion, but a significant uphill diffusion is necessary for pile-up creation. On the other hand, if dif-

Dopant	BF_2	
	(a)	(b)
implanted dose	4e14	4e14
after anneal:	(no pile-up)	
total dose by ERD *		2.8e14
pile-up dose by ERD *		9.7e13
HWHM pile-up / nm		1.40
depth resolution / nm		0.80

Table 4.7: Influence of the annealing temperature: Data extracted from the profiles in fig. 4.17.
** Values as observed by ERD/RBS, no SIMS correction was done.*

fusion is entirely prevented (e.g. by using a too low thermal budget), the dopants stay at their respective places, and hence no pile-up formation is possible.

One example for such a process is a deposition of amorphous silicon (a-Si). Figure 4.17a shows an ERD profile of a sample implanted with BF_2, on which the surface oxide was removed in a clean that included an HF dip. After an exposure to the cleanroom air of around 20 h, nominally 11 nm of a-Si was deposited[5] on the sample in a furnace. This process includes a thermal budget of 500 °C for around 2 h.

The ERD profile shows two small peaks of native oxide, one at the surface an one at the interface between a-Si and c-Si. No dopants are seen in the a-Si, indicating that the temperature was too low for any diffusion through the a-Si. Obviously, this also prevents the formation of a dopant pile-up at the top oxide interface.

Also around a-Si/c-Si interface in 16 nm depth, no pile-up is observed. The pile-up was removed during the HF dip, and no new pile-up was created during the a-Si deposition. This is explained by the fact that no oxidation occurred at the a-Si/c-Si interface and hence no pile-up formation is possible.

As a contrast to that, fig. 4.17b shows a sample, on which BF_2 was implanted through a thin native oxide (estimated to around 0.5 nm). Then a 2 nm oxide was grown. The oxidation process in a furnace includes a 750 °C, 2 h thermal budget. No further spike anneal was carried out. During such an thermal step, interstitial clusters that form already at lower temperatures are dissolved again, caus-

[5] The a-Si layer is observed to be around 15 nm instead of 11 nm thick, which is, however, irrelevant for this discussion.

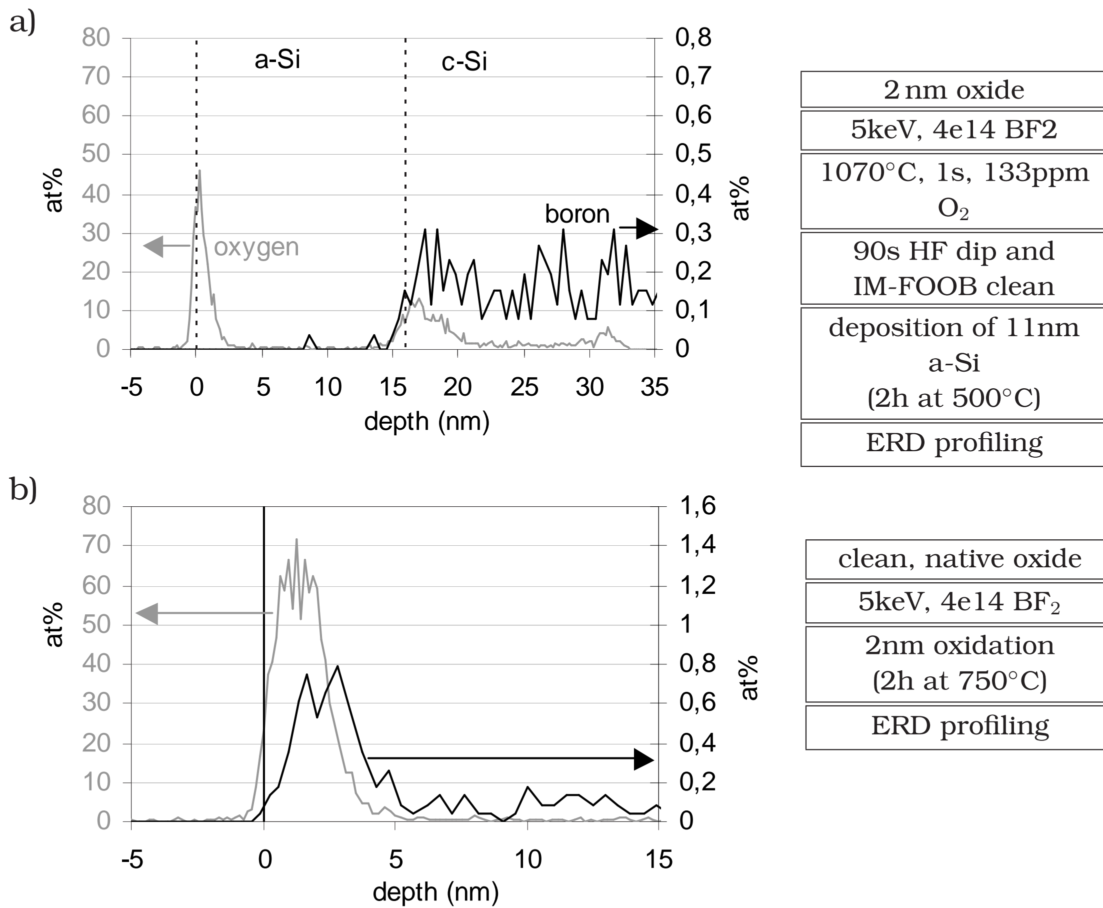

Figure 4.17: ERD profiles of two samples implanted with 5 keV, 4e14 BF_2.
a) After the spike anneal, the oxide was removed in a clean and an a-Si layer was grown (500°C, 2 h).
b) instead of a spike anneal, the sample was oxidized at 750 for 2 h.
(Samples from lot P000575 and E010353)

ing TED. At 750 °C, enhanced diffusion lasts for several hours [79], thus TED is present during the entire 2 h process time. The thermal budget is, however, not sufficient for good activation and an entire removal of the implant damage.

The ERD profile shows a clear pile-up at the interface. The result demonstrates that during an oxidation step at 750 °C, enough diffusion occurs to make the pile-up formation possible. Similar to the case of the spike anneal with strong oxidation (5% oxygen in the ambient) described above, the pile-up is not only located at the interface, but extends quite far into the oxide.

Comparing the 750 °C, 2 hrs. anneal to the 1070 °C, 1s spike anneal (e.g. fig. 4.18a), a very similar pile-up maximum concentration of 0.6% is observed. Due to the oxidation, the pile-up becomes thicker in the 750 °C case. Its dose is increased from 6.6e13 cm^{-2} (spike anneal) to 9.7e13 cm^{-2} (oxidation). This suggests that diffusion and oxidation are already sufficient for pile-up formation. A long annealing time seems to enhance pile-up formation more than a high temperature. This is analogous to the diffusion of dopants in the junction's bulk part, where a high temperature and short anneal time are helpful to optimize activation while reducing diffusion.

To investigate the temperature range of 950 °C to 1070 °C, which is used for thermal spike annealing, a series of electrical measurements was done. In the pile-up's electrical properties, no significant influence of the spike anneal temperature was observed. The experiment will be described in more detail in section 5.6.3.

4.7 Surface treatment to locate the pile-up

As discussed above, the interfacial dopant pile-up is seen to extend into the silicon for high doses of $3 \cdot 10^{15}$ cm^{-2} and more. In order to investigate more precisely the case of lower implanted doses, the surface oxide was removed in an HF dip. Entire wafers were etched in an automatic wet bench in a 2% HF bath for 30s. This etch step removes all of the surface oxide, but exhibits an good selectivity to silicon. More details on oxide removal in an HF dip will be discussed in the following chapter (cf. page 109).

After the oxide removal, the samples were stored on air for several weeks before analysis. During this time, a native oxide of around 1 nm grows, consuming around 0.5 nm of silicon[6]. According to the previously described results, it

[6] For 1 nm of oxide grown, 0.44 nm of silicon is consumed.

is expected that this oxide incorporates B atoms present in the consumed layer, whereas As atoms are expected to be pushed away by the oxidation front.

Figure 4.18 shows samples implanted with B, BF_2 and As. In each graph, the profile of an as-annealed sample is compared to the one of a sample after the described 30 s dip in 2% HF. The comparison shows that the HF dip removes the pile-up.

For the B samples, a very small portion of the pile-up seems to remain after the dip. This could be explained by a broadening of the pile-up into the silicon. However, also if the pile-up does not extend into the silicon at all, it is possible that the HF dip does not remove it entirely. It is known that an HF dip tends not to remove any B atoms at a bare silicon surface, probably because a threefold coordinated boron atom is more resistive to etching than a threefold coordinated silicon atom at the sample surface [117].

Taking the above-mentioned formation process into account, it seems however more reasonable that the pile-up is confined to the thickness of the interface itself, and that the observed pile-up remains are actually located at the very surface of the freshly etched sample. In any case, the pile-up dose remaining after the HF dip is very small ($<$10% of the pile-up dose for the B sample, $<$1% for the BF_2 sample), and the discussed effect thus applies only to a very small dose.

For As, the pile-up is removed entirely. This is an apparent contradiction to the results of Kasnavi et al. [106] who found that an HF dip in an inert ambient does not remove the pile-up. For the experiment discussed here, an HF dip under normal atmosphere was used, since no appropriate glove box was available. It is possible that diluted HF in the presence of atmospheric oxygen can remove a very thin layer of As implanted silicon (discussed in more detail in the next chapter), which then leads to the observed loss of the pile-up. Under this assumption, the results once more confirm that the arsenic pile-up is confined to the first few monolayers of silicon.

The findings are, however, only correct for the low doses of $4 \cdot 10^{14}$ or $8 \cdot 10^{14}\,cm^{-2}$ investigated here. When investigating much higher doses of $3 \cdot 10^{15}\,cm^{-2}$and above, the pile-up starts to extend into the silicon by several nm, as shown and discussed before. These dopants are expected to stay in the sample also after an HF dip.

4.8 A model for the pile-up formation

The results mentioned above are used to propose a phenomenological model of the pile-up creation mechanisms.

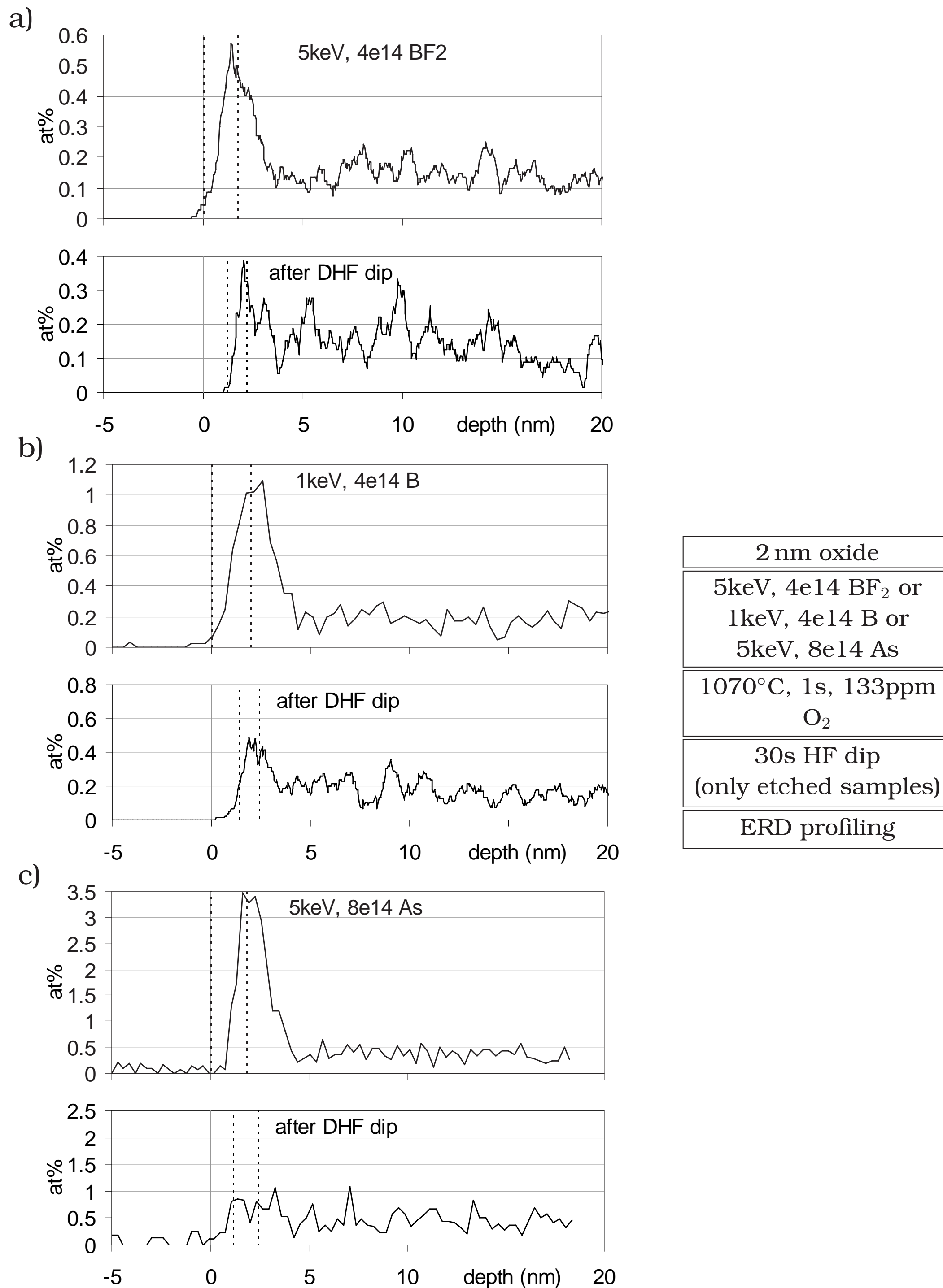

Figure 4.18: ERD/RBS profile comparisons between samples before and after a 30 s dip in 2% diluted HF and native oxide regrowth. The depth scale of the etched profiles is shifted to compensate for the material lost. The dashed lines indicate the surface oxide. (Samples from lots P010403, E010731, T020208)

At first, dopants diffuse during the spike anneal, with a preferential direction towards the surface (uphill diffusion). When arriving at the silicon-oxide interface, a large fraction of the dopants is trapped. The oxidation mechanism, which is necessary for pile-up formation (at least in the case of boron), can microscopically be regarded as a rearrangement of atoms on the Si side of the interface, while incorporating oxygen atoms that arrive through the oxide. During this rearrangement process, numerous trapping sites are created (e.g. dangling bonds) that can trap dopant atoms. The interface thus acts as a dopant sink.

Boron

Dopants trapped at the beginning of the oxidation process stay at their position and are incorporated into the oxide, while the interface moves away from them. Only the dopants that are incorporated at the end of the oxidation are close to the interface's final position. Due to this effect, the pile-up is not confined to the interface, but extends into the oxide. The width of the pile-up is equal to the thickness of the oxide grown during anneal.

In principle, the profile of the pile-up should reveal information about the course of the anneal: The dopant concentration at a certain depth is a function of oxidation speed and concentration of available dopant atoms for that moment, when the oxide-silicon interface was located at that specific depth. However, a quantitative evaluation is very difficult, since the dopant incorporation is influenced by numerous parameters that change over the course of the spike anneal, like temperature, dopant and interstitial concentration, diffusivities etc. It seems also possible that the moving interface drags a part of the pile-up dopants along. Even with models for all these effects available, the limited ERD resolution does not allow a good extraction of the relevant parameters from the profile shape alone.

The spreading of the pile-up into the oxide is not determined by the initial pile-up formation process, but by the incorporation into the growing oxide. On the silicon side of the interface, no such spreading mechanism exists. Therefore, the pile-up extends into the silicon by not more than the thickness of the layer in which the lattice rearrangement during oxidation is taking place, thus by 1-2 monolayers.

Arsenic

Since arsenic is not soluble in SiO_2, the incorporation of dopants into the oxide is not applicable. Instead, the As dopant atoms are observed to leave the oxide,

either to the ambient or towards the bulk. The spike anneal used (1070 °C, 1 s) is sufficient to remove all dopants from the oxide.

Again, dopant atoms diffusing to the interface, either by leaving the oxide or by an uphill diffusion from the bulk, are blocked by interface traps. The higher pile-up dose indicates that the interfacial trapping mechanism is more effective for As than for B.

The trapping only occurs at the interface itself. Since arsenic is not soluble in oxide, the moving interface (oxidation during anneal) shovels the dopants along (cf. [93]). The spreading of the pile-up in both directions is limited by the depth of the oxide-induced lattice distortion. The experimental results suggest that the pile-up is not more than 1-2 Monolayers wide.

4.9 Diffusion mechanisms

In the previous chapter (section 3.4.2), uphill diffusion was mentioned as one main effect necessary for pile-up formation, next to the trapping of dopants at the silicon/oxide interface.

It is known (e.g. [88]) that silicon self-interstitials, one form of implant damage, can cause significant diffusion of dopants towards the surface. This effect is closely linked to TED, and it is certainly present in the samples investigated.

However, the diffusion observed here could also be due to normal diffusion, in combination with a strong dopant trapping mechanism at the interface. This trapping acts as a strong dopant sink and reduces the number of dopants available for diffusion close to the interface to very low levels. As a consequence, normal diffusion would bring more dopants to the interface.

Considering the dose results shown on the previous pages, it is unlikely that normal diffusion alone is able to bring enough dopant atoms to the interface, and to create the observed pile-up doses of up to two thirds of the entire retained dose. It is therefore believed that the enhanced uphill diffusion, mediated by interstitials, plays the predominant role.

From the experiments discussed it is not yet entirely clear how much each mechanism contributes to the observed diffusion. This should be clarified in further experiments. For example, one could use an additional deep Si implant to enhance TED and uphill diffusion effects, or use a deposited doped junction to avoid TED and to observe only normal diffusion.

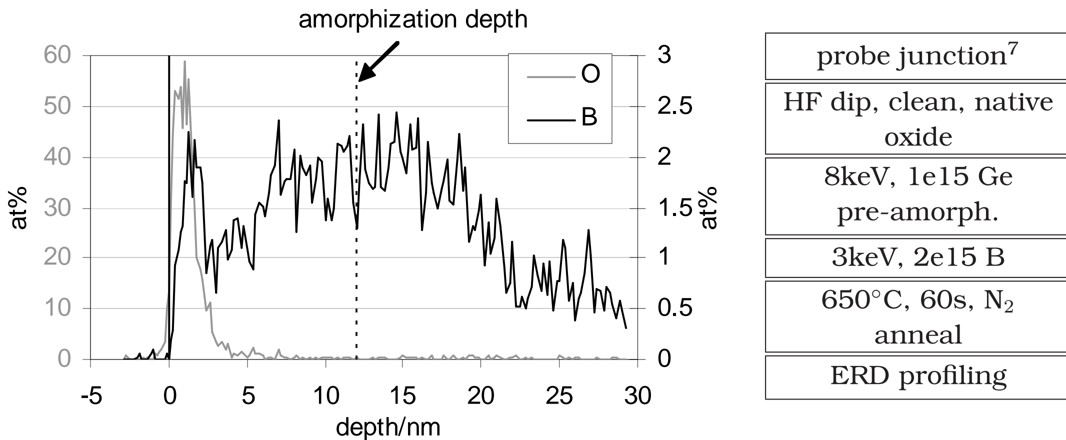

Figure 4.19: ERD oxygen and boron profile of an SPER annealed sample. (Sample from lot P030024)

4.10 Pile-up on SPER annealed samples

SPER is a novel junction formation technique, that uses a low-temperature anneal to crystallize a pre-amorphized layer. During the recrystallization, dopants are quenched into the lattice, without being given the opportunity to reach a thermodynamical equilibrium between substitutional and interstitial lattice sites. Under optimized conditions, very high active dopant concentrations above the solid solubility limit can be reached. The temperature is chosen low enough to avoid diffusion (cf. section 1.4.4).

Important differences between SPER and a spike anneal are the tremendously reduced anomalous diffusion (TED) due to the low processing temperature, and the a-Si/c-Si interface that moves towards the surface during the anneal. Its speed is in the order a few nm/sec, depending on the temperature.

Figure 4.19 shows the ERD profile of an SPER annealed sample. The sample was pre-amorphized with 8 keV, 1e15 Ge to a depth of 12 nm. After a relatively deep 3 keV, 2e15 B doping implant[8], the sample was annealed in a pure N_2 ambient at 650 °C for 1 minute. As described before, a small amount of oxygen is always present in the RTP tool.

[7] The probe junction is formed by a clean, then a 20 keV, 5e12 B implant and a 1100 °C, 1s spike anneal. It is used to facilitate the sheet resistance measurements and has no significant influence on dopant profile or sheet resistance (cf. section 2.3.1, page 24)

[8] The projected range is 12-13 nm.

The anneal temperature is too low to remove Si self-interstitials or other deep damage. Consequently, only very little TED or uphill diffusion is observed, and also the normal diffusion is mostly suppressed. The profile is almost preserved in its as-implanted state. The dose observed with ERD of 2.05e15 is (within the ERD errors) similar to the implanted dose, thus no dopant loss is observed during the low-temperature anneal.

After regrowth, only a small pile-up is seen at the silicon-oxide interface, reaching a maximum concentration of 2 at%. This is much less than the 7-10 at% expected for a comparable spike annealed sample with the same dose. As discussed above, a strong uphill diffusion is necessary for pile-up formation. In the SPER case, almost no Si uphill diffusion occurs due to the low temperature used, and consequently only a little pile-up is created.

Alternative to the uphill diffusion model, another pile-up formation mechanism has to be considered for SPER: The observed (small) transport of dopants towards the surface could be caused by the moving amorphous/crystalline interface, that carries a small fraction of the dopants along towards the surface. When the recrystallization process stops at the oxide interface, the dopants stay at their position, forming a small pile-up. To distinguish the two transport mechanisms, a future experiment with reduced annealing time is suggested, allowing only a partly regrowth of the amorphous layer. In that case, any dopants carried along with the interface would form a small peak at the position of the interface.

Another experiment with elongated annealing time, on the contrary, would allow more time for TED. If uphill diffusion by self-interstitials is the main factor for dopant transport to the interface, this would lead to increased pile-up formation. For the sample discussed here, a 1 minute anneal was used. This is longer than the typical time needed for recrystallization (in the order of 10s for an anneal temperature of 650 °C). For additional research, one could deposit doped amorphous silicon on a wafer. Without any implant damage, neither TED nor uphill diffusion should occur during the regrowth of such a sample.

4.11 Summary

Various samples were implanted and annealed with different processing parameters within the range that is useful for USJ formation. After spike anneal, all B, BF_2 and As samples exhibited a pile-up of dopants at the interface between the silicon and the screening oxide layer. For As, the pile-up is probably confined to the thickness of the interface itself, i.e. to not more than 2-3 monolayers. The

boron pile-up is seen to extend a few monolayers into the oxide. For both dopants, the pile-up does not extend significantly into the silicon. Only for high implanted doses of $3 \cdot 10^{15}\,\mathrm{cm}^{-2}$ and above, a shoulder of clustered dopant atoms is seen next to the interface on the silicon side. The dose trapped at the interface itself was seen to reach almost one monolayer ($7 \cdot 10^{14}\,\mathrm{cm}^{-2}$).

For boron, the pile-up becomes stronger for higher doses. At $3 \cdot 10^{15}\,\mathrm{cm}^{-2}$, two thirds of the dopants retained after anneal of a boron sample were located in the pile-up. Furthermore, self-sputtering during implant and outdiffusion during anneal cause increased dopant loss for high implanted doses. Pile-up formation and dopant loss set an upper limit to the amount of dopants that is retained and activated in the bulk part of the junction after anneal. The results indicate that this limit is in the order of $1\text{-}2 \cdot 10^{15}\,\mathrm{cm}^{-2}$, for the implant energies and the type of spike anneal that were investigated.

For arsenic, on the contrary, the relative importance of the pile-up was seen to increase for lower doses. At an implanted dose of $4 \cdot 10^{14}\,\mathrm{cm}^{-2}$, around $(70 \pm 15)\%$ of all retained dopants are observed in the pile-up. For higher implanted doses, this percentage is reduced. The additionally implanted dopant atoms preferentially stay in the bulk part of the junction.

Oxidation during anneal is observed to significantly enhance pile-up formation. At least for boron, no pile-up is built during an anneal without oxidation. In the RTP tool used during this work, the oxygen contamination of the chamber unavoidably causes some oxidation. This is already sufficient for pile-up formation, if the sample is only covered by a thin oxide of 2 nm or less. For arsenic implanted samples, oxidation and pile-up formation are observed during anneal even if the sample is covered by a 5 nm thick oxide.

A model for pile-up formation is proposed that includes uphill diffusion of dopant atoms towards the surface and trapping of these dopants at the silicon/oxide interface. The dopant uphill diffusion is mediated by an anisotropic diffusion of silicon self-interstitials towards the surface. Dopant atoms are carried along and get trapped at the Si/SiO_2 interface. The trapping sites are generated by the oxidation process.

Uphill diffusion is closely related to TED, since the presence of Si self-interstitials is required for both effects. Any method that suppresses TED by optimizing the different implant and anneal parameters is therefore expected to suppress as well the pile-up formation process.

If oxidation occurs during anneal, the silicon-oxide interface progresses. In the case of boron, dopant atoms are immobilized at the place where they were initially

trapped, while the oxide interface continues to move deeper. Therefore, the boron pile-up can extend into the oxide. The width of this extension is determined by the thickness of the oxide grown during anneal. In some cases, also a fraction of the as-implanted profile is observed in the oxide, because the spike anneal's thermal budget is too low to allow a significant diffusion of these dopants.

Arsenic, on the contrary, is not soluble in oxide. The dopant atoms residing in the oxide after implant diffuse towards the silicon or to the ambient during anneal. The pile-up dopants trapped at the interface are carried along by the moving interface. As a result, the arsenic pile-up is located in a very narrow region, not more than a few monolayers wide. This region seems to be identical to the physical width of the interface itself.

For SPER samples, almost no diffusion is observed, the majority of the dopant profile is preserved in its as-implanted shape. However, also here, a small pile-up is observed at the silicon-oxide interface. The main factor limiting the pile-up dose seems to be the fact that only very little TED occurs during regrowth.

5 *Electrical effects in the pile-up*

5.1 Introduction

In the previous chapter, only the chemical characteristics of the pile-up were investigated. However, the junction's electrical properties are even more interesting, since they influence directly the performance of an actual device. It is therefore desirable to understand not only the physical background of pile-up formation, but also its impact on the dopant activation and hence on the junction's sheet resistance. Also, the effect of surface (or other) treatment during processing on the pile-up's conductivity should be understood, in order to prevent an increase in sheet resistance.

In this chapter, electrical measurements will be described that complement the information from the chemical dopant profiles discussed before. To separate the pile-up's contribution to the conductivity from that of the bulk junction, a surface treatment was used, consisting of HF dips removing the surface oxide and the subsequent native oxide regrowth. Comparisons of differently prepared samples give direct information about the influence of the processing parameters on the electrical performance of the junction. As the activation results are based on only small changes in the measured sheet resistances, a lot of care has to be taken to evaluate the reliability and accuracy of the measurements.

5.2 Accuracy of sheet resistance measurements

In this section, the sheet resistance measurement technique, already presented in section 2.3.1, is assessed for accuracy and repeatability on the samples used. Values are given only for the predominantly used SSM-240 tool. The tool's sample stage is designed for wafers with diameters between 2 and 6 inches, thus the 8-inch wafers used for this work had to be broken into samples. This also allowed to investigate a large variety of conditions while reducing the amount of processed wafers needed.

The manually cleaved samples were typically sized between 3x3 and 5x7 cm^2. This is largely sufficient to avoid edge influence on the measurement. The four-point probe tool SSM-240 was used in in-line mode, with the measurement spot aligned manually to the center of each sample.

For each measurement, R_s was measured five or ten times with a distance of 10 μm. The variation between the individual measured values was never above $\pm$0.25%[1] for boron samples ($\pm$0.75% for arsenic), and typically below $\pm$0.1% ($\pm$0.2% for arsenic). This error is due to R_s variation within the 0.5-1 mm measurement distance as well as to the tool's precision.

To investigate the repeatability, one sample was measured consecutively several times. Before every measurement, the sample was newly placed on the holder. The accuracy of the manual placing on the measurement stage is around 1 mm. Together with the tool repeatability, a variation of less than $\pm$0.15% for ten consecutive measurements was found.

If much more than ten measurements are done on the same spot, a deterioration of the samples is observed. After 25 measurements, the measured R_s increases by around 0.5%. However, the experiments discussed in the following sections included not more than 10 measurements on the same sample, thus no significant sample deterioration is expected.

Bigger variations were observed between the absolute sheet resistance of different samples. The samples investigated exhibited a within-wafer non-uniformity of around $\pm$1.5%, wafer-to wafer variations of also around $\pm$1.5%, and lot-to-lot variations of $\pm$4-6%. The last two numbers, however, are only based on very few wafers (2-4). For this comparison, only samples from wafers with nominally identical processing were used.

5.3 Surface treatment methods

A surface treatment was used to investigate the shape of the pile-up as well as its electrical properties. The treatment consisted of a hydrofluoric acid (HF) dip to remove the screening oxide, and the subsequent native oxide regrowth in the cleanroom ambient. Before the presentation of the electrical results, the next sections will focus on the HF dipping process and the quality of the results

[1] All percentages here are given as standard deviation (one sigma) in percent of the average sheet resistance.

5.3.1 Oxide removal by HF dips

Automatic wet bench

The samples were etched in various HF baths. HF removes oxide according to the reaction $SiO_2 + 6HF \rightleftharpoons H_2SiF_6 + 2H_2O$. The selectivity of HF between oxide and silicon is in the order of 100:1, because silicon is only etched after being oxidized in the aqueous solution [118].

For entire wafers, a Steag automatic wet bench was used to etch the surface oxide in a 2% HF solution for 30-60 s, after which the wafers are transferred to a DI water bath for a 5 min rinse. After the rinse, the wafers are dried in a Marangoni process lasting 10 minutes, which included water and alcohol application. After etching, the sheet resistance was measured with the Tencor RS75 tool, and a Tencor ASET F5 was used to measure the thickness of the regrowing native oxide.

The disadvantages of this combination lie in the long rinsing and drying time. The results indicate that the surface passivation after the etching dip is not perfect, and a small native oxide is probably growing during rinse and drying. Due to limited wafer transport speed, the automatic wet bench was not able to precisely adjust etch times below 30 s. Furthermore, ellipsometry proved to be not sufficiently precise for oxide thicknesses below 1 nm. Therefore, the results need to be interpreted with care.

The etch rate of the 2% HF bath of the automatic wet bench was determined at around 9 nm in 30 s. Since the investigated samples all had surface oxides of not more than 7 nm, a dip of 60 s is clearly long enough to remove all oxide from the sample.

Manual HF dips

In order to optimize rinsing and drying times, to allow a fast sample handling, and above all to improve etching and measurement accuracy, a manual etching process was used for most experiments. The wafers were broken into samples with sizes ranging between 3x3 and 5x7 cm^2, as described above. The samples were manually etched in a beaker with buffered HF (BHF) or 2% diluted HF (DHF), then rinsed for around 30 s in two separate beakers with de-ionized water (DIW) and an overflow bath, to allow a fast, but sufficient rinse.

The BHF solution is composed of 6% HF, 35% NH_4F and 59% H_2O. (It is also called BHF 7:1, since it contains 1 part of 49% HF solution for every 7 parts of 40% NH_4F solution.) An oxide etch rate of around 100 nm per minute is expected [119],

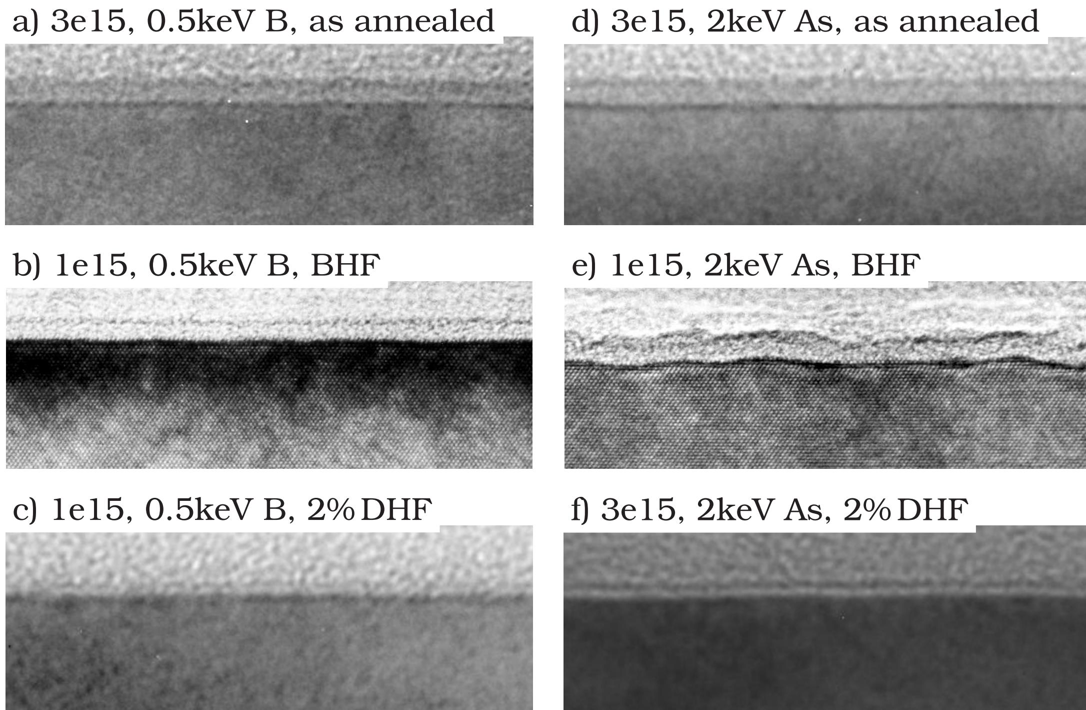

Figure 5.1: TEM cross-section images of samples implanted with B or As after anneal (top) and after a 60 s HF dip and native oxide regrowth. BHF (middle) and 2% DHF dips (bottom) are compared. (Samples from lots T020208, P020334, P030122, P030275)

the etching goes thus much faster than with the 2% DHF. The buffer reaction $NH_4F + H_2O \rightleftharpoons NH_4OH + HF$ keeps the pH value constant. This BHF solution is optimized for a constant and uniform etch process and a better surface passivation than a pure aqueous HF solution [120]. The etch selectivity of BHF and DHF will be assessed later.

Immediately after the dip, the sheet resistance of the samples was measured with the SSM-240 four point probe tool. On other samples, the oxide thickness was measured with XPS.

This manual method of etching results in samples with a well passivated surface, i.e. all dangling bonds at the surface are saturated by hydrogen atoms. The quality of the passivation is at least good enough to prevent any oxidation for the first few hours. Sometimes, an oxide-free surface was observed even after exposing the sample for more than one day to the cleanroom ambient. The etch step in the automatic wet bench seems to leave the samples with a much worse passivation.

5.3.2 Surface quality after oxide removal

TEM images were used to assess the quality of the surface after an HF dip. Figure 5.1 compares TEM bright field images of B and As implanted samples. In the top row, two samples after anneal are shown with an oxide thickness of 2.6 to 3 nm. The oxide quality is very good, the interface is smooth.

The two following rows show samples on which the oxide was etched in a 60 s BHF (middle) or a 60 s 2% DHF dip (bottom). The oxide seen is a native oxide that grew over several weeks' time. The oxide and interface quality is still very high. Only in the case of the As sample etched with BHF, an increase in surface roughness is seen – this indicates that BHF is slowly etching As-doped silicon.

On a larger scale of several 100 nm, Iacona et al [121] have observed a similar increase in interface roughness and correlated it with the formation of As precipitates. Only for a high oxidation temperature of 1100 °C, the roughness was similar to other dopant species.

One possible explanation as to why BHF attacks silicon, based on a statistical variation in the As concentration at the interface, is as follows. As discussed before, the presence of arsenic at the interface enhances oxidation during anneal, compared to B implanted samples. This was also observed for native oxide growth (discussed below). A local maximum in arsenic concentration could therefore locally enhance oxidation during the BHF dip. The grown oxide is immediately removed by the etchant. Since the oxidation process pushes the As dopants away, a small pile-up (below ERD detection limit) is created below the oxide, which increases the local interfacial concentration of arsenic. This positive feedback loop could be sufficient to create an interfacial roughness visible in TEM.

5.3.3 Etching selectivity

In general, an HF dip, in DHF as well as in BHF, is considered to have a very high selectivity between oxide and silicon. To test this, a range of samples were etched in the automatic wet bench and in a manual procedure.

Automatic wet bench

Samples covered with a 2 nm oxide and implanted with B, BF_2 and As were etched for 30 s, 90 s, 180 s or 360 s.

The process flow in the automatic wet bench starts with an 2% DHF dip for a chosen time. Then the wafers are rinsed for 5 minutes in DI water and dried

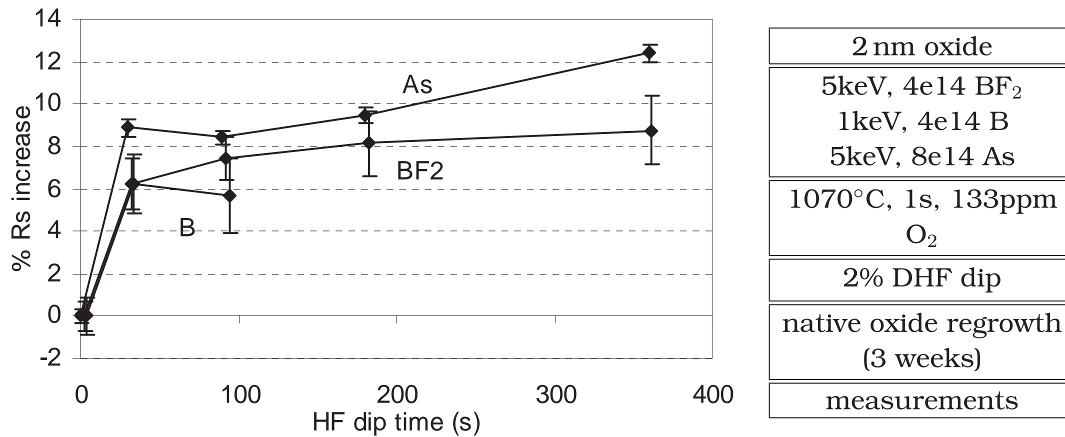

Figure 5.2: Percentile sheet resistance increase after HF dip (in an automatic wet bench) of wafers implanted with 5 keV, 4e14 BF_2, 1 keV, 4e14 B and 5 keV, 8e14 As, as a function of dip time. (Samples from lot P010403)

using a Marangoni[2] dry (around 10 minutes). Due to this lengthy procedure, a small amount of oxide grows already on the wafers before they leave the tool, hence no measurement without oxide was possible. In order to provide stable and repeatable measurement conditions, a native oxide was given time to grow. The values discussed here were measured several weeks after the HF dip.

Their relative sheet resistance increase, compared to non-etched samples, is shown in fig. 5.2. For all investigated dopants, the increase in sheet resistance is almost independent of the HF dip time. The slight average increase over time observed for the BF_2 and As samples corresponds to an estimated silicon removal of 0.07 nm/min (BF_2) or 0.3 nm/min (As). This indicates that for As, a small amount of 1-2 monolayers might be removed per minute. For the purpose of this work, this is not problematic, but should be kept in mind. For the B implanted samples, the available data show no significant etching of silicon at all.

This material removal might be caused by cleanroom light creating electron-hole pairs in the wafer, enabling HF to dissolve silicon. Since there is no possibility to change the cleanroom light or to add an opaque cover to the wet bench in our case, this effect cannot be avoided or investigated in further detail. For the manual HF dips, experiments using opaque beakers will be described further below.

[2] During a Marangoni dry, the wafers are slowly lifted from a DI water bath, while gaseous alcohol (IPA) is blown over them. The IPA displaces the water drops on the wafer, and then evaporates without leaving any stains behind.

The values displayed in fig. 5.2 are measured at 49 points on entire wafers. The relatively large error bars in the boron case are due to non-uniformity issues across the entire wafer.

Manual HF dips

Another series of tests was carried out manually to investigate the etch selectivity of the DHF and BHF dips. Samples implanted with 0.5 keV, 1e15 B or 2 keV, 1e15 As were etched in BHF or 2% DHF with various dip times between 20 s and 300 s. Most of the samples were etched in an open beaker, in the light of cleanroom and fume hood. For other samples, opaque beakers with cover were used. These samples were simultaneously exposed to HF and light only during insertion into the HF bath and during transfer to the rinse beaker (i.e. 1-2 seconds in total). The sheet resistance was again measured immediately after the HF dips and compared to the value before the dip. The results are given in figs. 5.3 and 5.4.

Contrary to the discussion above (section 5.2), the error values presented here and in the following sections will be given in absolute percentage points (e.g., a resistance increase of (5±2)% stands for a sheet resistance increase between 3% and 7%), instead of a standard deviation given in percent of the average value.

Manual etching with BHF. In fig. 5.3a, an average percentile R_s increase of (2.1±0.8)% is observed for B etched in BHF with normal light. No significant time dependence is seen. This indicates that a BHF dip on a B implanted sample is sufficiently selective, however the relatively low signal-to-noise ratio of 3:1 has to be kept in mind. The large error of these measurements indicates that the reproducibility of the results is not better than ±0.8% (percentage points). For an etch time of 40 s, the sheet resistance increase is very low (observed similarly for a DHF dip, fig. 5.4), however this is not considered as a significant effect.

The dark curve suggests also a good selectivity, with a result of (1.3±0.6)% R_s increase. These numbers appear to be slightly smaller than for the etch with light, however they are within the repeatability limits of the experiment (see next section).

The values observed here are lower than the ones of the automatic wet bench, because for this experiment, the sheet resistance was measured immediately after the BHF dip, instead of after several weeks (cf. p. 111). As it will be shown later, the growing native oxide also contributes to an increasing sheet resistance.

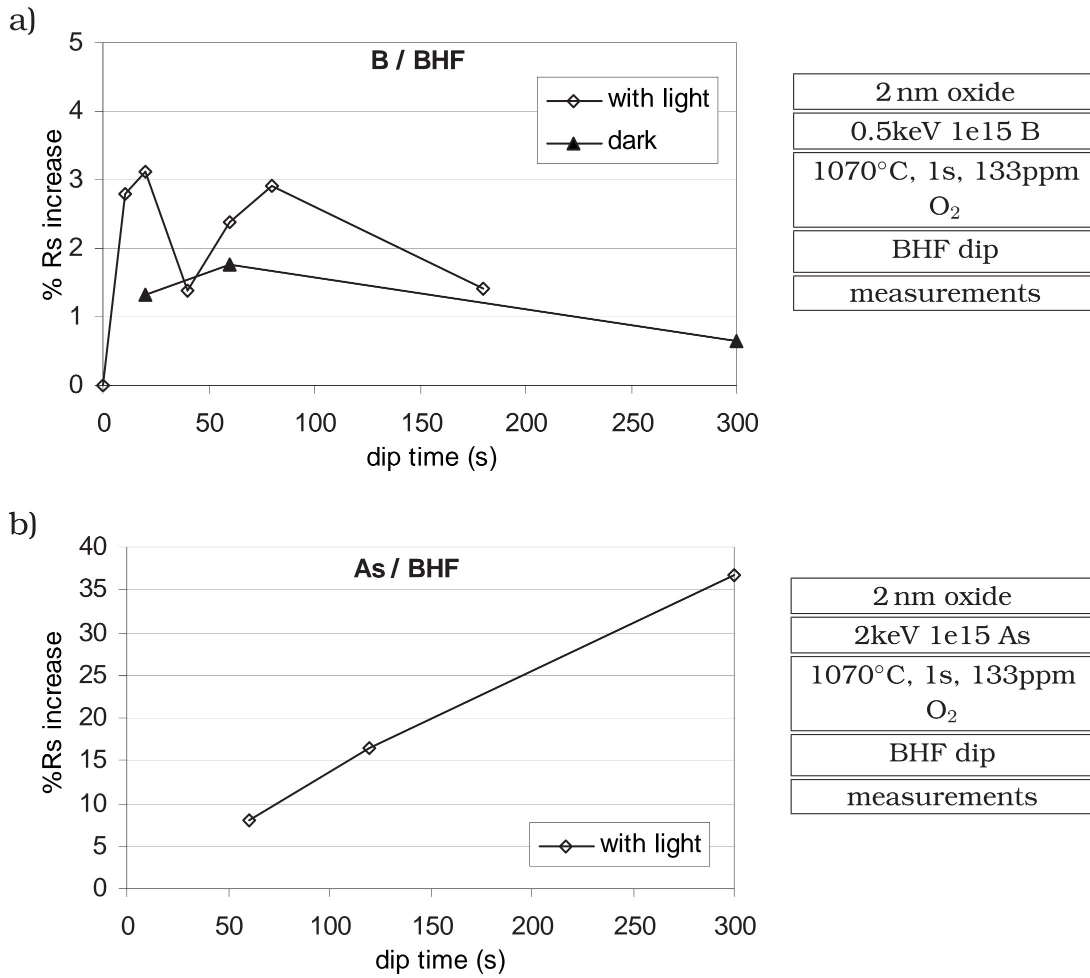

Figure 5.3: Percentile sheet resistance increase after manual buffered HF (BHF) dip of wafers implanted with 0.5 keV, 1e15 B or 2 keV, 1e15 As, as a function of dip time. (Samples from lot P030275)

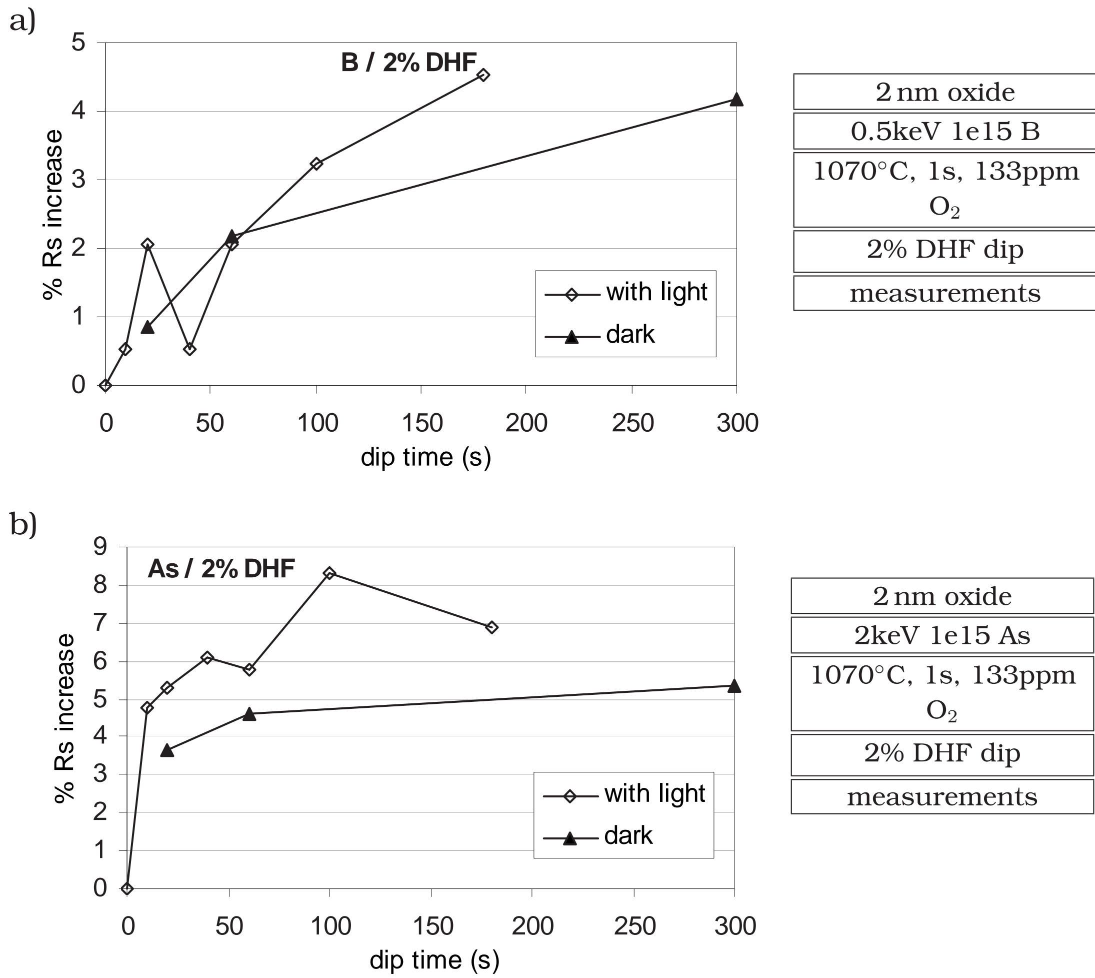

Figure 5.4: Percentile sheet resistance increase after manual 2% diluted HF (DHF) dip of wafers implanted with 0.5 keV, 1e15 B or 2 keV, 1e15 As, as a function of dip time. (Samples from lot P030275)

Contrary to this result, a clear etch time dependence of the sheet resistance is seen on the arsenic sample etched in BHF (fig. 5.3b). After 5 minutes, the sheet resistance has increased by almost 40%. This suggests that buffered HF actually does attack silicon. A loss of the top 3-5 nm of the doped silicon would explain such an R_s increase. The precise value depends on the level of activation of the pile-up dopants, the interface roughness after native oxide growth and some more details. The results suggest once more that the presence of As at the sample's surface promotes oxidation during etching, enabling the HF to remove silicon. Light probably enhances this effect by providing more electron-hole pairs.

The results thus suggest that during a 60 s BHF dip in cleanroom light, around 0.6-1 nm of the arsenic implanted silicon is removed, while for boron samples, no measurable silicon loss is observed.

Manual etching with 2% DHF. The usage of diluted HF results in almost opposite effects than BHF: For B implanted samples (fig. 5.4a), the results suggest now a time dependence, however on a very small scale.

This result is contradictory to the one described above, where a DHF dip in an automatic wet bench was seen to be sufficiently selective. It is not clear where this difference comes from. Since oxidation during the etching process is known to be the major cause of a bad selectivity, it seems possible that the aqueous HF solution in the beaker collected more oxygen from the ambient than the one in the automatic wet bench, where the etchant was continuously replaced in an overflow bath. However, for the experiments described in the following chapters, all B implanted samples were etched in BHF, thus any uncertainty on the selectivity of DHF was circumvented.

For the As sample (fig. 5.4b), DHF has a quite high selectivity. The results show only a slight increase in R_s with etch time, much smaller than seen for BHF. A step-like R_s increase of (5±0.5)% is seen in the beginning; then R_s increases by around (1±0.3)%/min. The dark etch results in an R_s increase of (4.5±0.7)% with almost no time dependence. An extrapolation of the observed lines to a zero etch time suggests that for the given wafer, the pile-up removal accounts for around (4.5±0.5)% R_s increase, the rest is caused by silicon removal during the HF dip.

The results indicate clearly that for B, BHF is the better etchant. Therefore the following discussion will mainly focus on this combination. Only for the samples etched in the automatic wet bench, DHF results will be discussed, since no BHF bath was available.

For As, on the contrary, the decision is less clear. While BHF has the disadvantage of significantly attacking the silicon, it is known to produce a better surface passivation after the etch, with more stable surface conditions. The following discussions will therefore cover both BHF and DHF dips.

In all cases, and especially the As/BHF case, care needs to be taken to keep the HF dips at exactly the same time for all samples. The results below therefore only discuss dips of 60 s. With this specification, the comparability of all samples is optimized. However, the possible silicon etching needs to be taken into account, if results are given as absolute values.

5.3.4 Repeatability

While the previously discussed experiments gave more fundamental information on the HF dipping properties, this section will focus on the actual repeatability of a standard 60 s etch.

At first, differences between samples etched together in one experiment are discussed. They are limited in number and dipped immediately after each other in the same etching and rinsing baths. Possible variations might result from a non-perfect time control (variations of ± 1 s are possible), or slight differences in the rinsing procedure.

Then, variations between different experiments (i.e. on different days) are evaluated. Here, a possible contamination of the beakers, a differently composed etching solution (e.g. for the diluted HF dips), or changing light conditions (e.g. when using a different working place) might influence the results.

Repeatability within one experiment. Four B and four As samples, from the same two wafers, were etched for 60 s. BHF was used for the B samples, 2% DHF for the As samples. The (separate) samples were prepared from the same As and B wafer and etched immediately after each other in the same bath. The beaker contained around 700-800 cm^3 of BHF or DHF. This quantity is considered sufficient to etch the four samples without any deterioration[3].

Fig. 5.5a shows again the percentile sheet resistance increase between two measurements immediately before and after the HF dips. For the B samples, an

[3] A typical sample contains 3 cm*4 cm*2 nm=$5 \cdot 10^{-6}$ cm^3 of oxide. The beaker contains around 15 cm^3 of (pure) HF, thus six to seven orders of magnitude more than the oxide volume. To avoid any risk of etchant deterioration, the etchant and the rinsing baths were replaced after at maximum seven samples in all experiments.

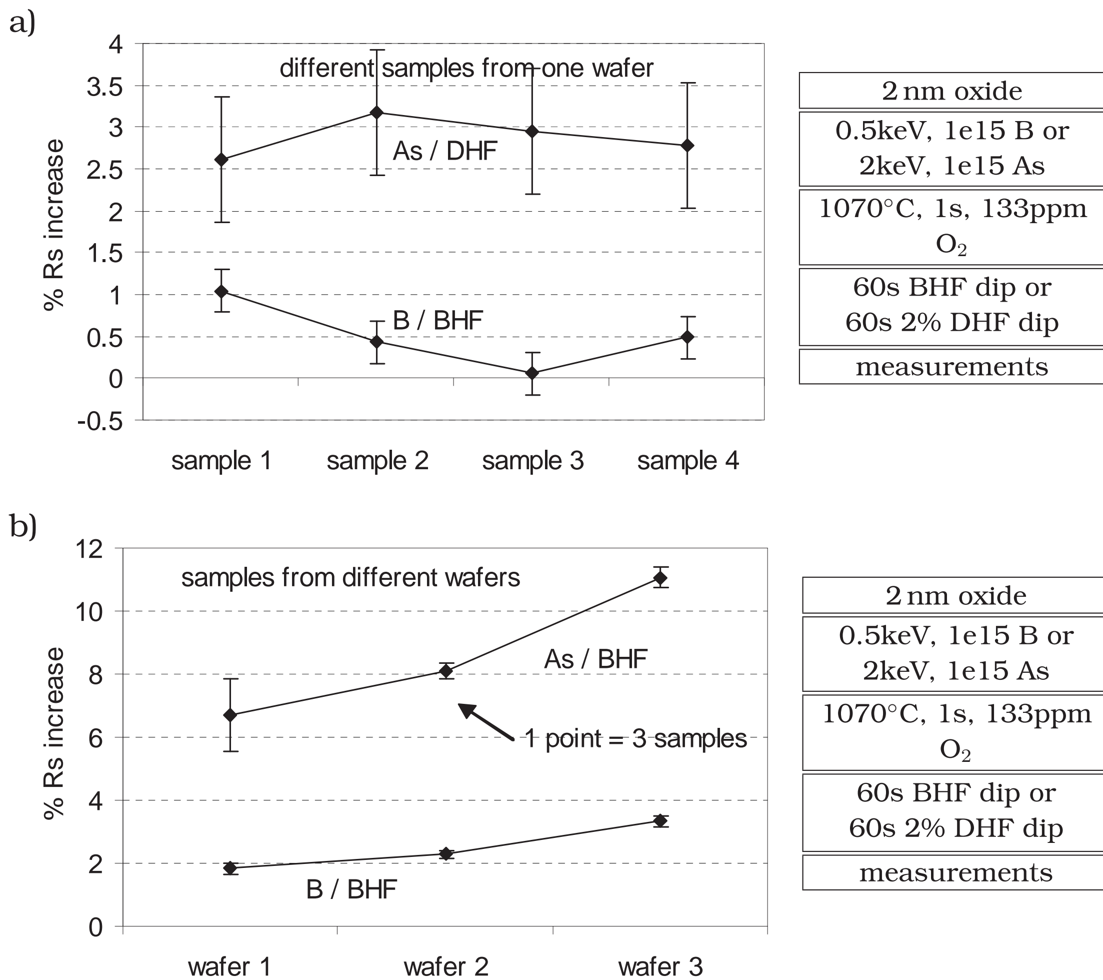

Figure 5.5: Percentile sheet resistance increase after manual HF dip of wafers implanted with 0.5 keV, 1e15 B and 2 keV, 1e15 As, for 60 s dips in BHF or 2% DHF. Contrary to the other figures, the error bars in subfigure b) do not indicate the measurement error of the FPP tool, but show the variation between the samples. Each point is an average of three samples from the same wafer. (Samples from lot P030275(a) and P020334/P020465 (b))

R_s increase of $0.5 \pm 0.4\%$ was observed, and $2.9 \pm 0.3\%$ for As. In both cases, the repeatability of ±0.3 or ±0.4 percentage points is quite good.

For further investigation, a new experiment was done to also compare samples from different wafers and from different lots. The results are shown in fig. 5.5b. Three boron and three arsenic implanted wafers were investigated. From each wafer, three samples were etched and measured. The variations between the three samples of each wafer are given as error bars.

However, contrary to the previously described results, the samples were etched together in one dip. This avoids dip-to-dip variations, therefore the error bars in fig. 5.5b are smaller than the variations described above for fig. 5.5a. Except the three samples from the first As wafer, these variations of 0.1-0.2% are explained by FPP measurement repeatability as described above in section 5.2. This demonstrates that the variations seen before between identical samples, etched separately but within one experiment, were due to variations in dipping, rinsing and handling.

For each B and As, wafer 1 and 2 were processed together in lot P020334; wafer 3 was processed in lot P020465. The results show a quite large difference between the different wafers for As of $\pm 2\%$ (percentage points), and smaller, but not negligible variations of $\pm 1\%$ for B.

Repeatability between experiments. When comparing the results of fig. 5.5a to the ones presented before (the 60 s points in fig. 5.3a and 5.4b), a difference in the average value is observed – instead of an R_s increase of 2.5% or 5.7%, fig. 5.5a shows only values of 0.5% and 2.9%, respectively, even though the samples were broken from the same wafers.

These differences are higher than what is seen for different samples within one experiment. It is thus concluded that the result variations between different experiments are in the order of $\pm 1\%$ (percentage points) for B and $\pm 2\%$ for As. These values are estimated from only two experiments.

5.3.5 Conclusions on accuracy and repeatability

Table 5.1 gives an overview on all different accuracy issues. The errors are given as percentile errors of the absolute resistance values. Since most of the electrical results of this work are presented as percentile sheet resistance increase, the errors can be considered as absolute errors of the increase, given in percentage points.

	B	As
FPP measurement issues or sample non-uniformity		
individual measurement error	0.25%	0.75%
FPP repeatability	0.15%	
within-wafer variations	1.5%*	
wafer-to-wafer variations	1.5%*	
lot-to-lot variations	4-6%*	
HF dip variations or pile-up electrical activity		
sample-to-sample variations	0.2%	0.2% DHF
dipping and rinsing between samples	0.4%	0.3%
wafer-to-wafer variations	1%	2%
experiment-to-experiment variations	1%	2%

Table 5.1: Overview on all accuracy issues. These values are estimated one sigma standard deviation values.
** The marked errors can be avoided by only considering the sheet resistance increase after an HF dip, instead of the absolute sheet resistance.*

The observation of a percentile sheet resistance increase also helps avoiding problems with variations in dopant dose or activation. The errors marked with a * are therefore irrelevant to the results discussed in this work.

The table shows that the largest variations are seen between wafers and/or between experiments. The results shown below in this work are therefore always extracted from comparisons of values within one experiment (i.e. all samples etched immediately after each other), using samples from the same wafer. Like this, the error is not higher than 0.3 (B) or 0.4 (As) percentage points, respectively.

The error bars on the following figures represent, however, always the measurement accuracy of the individual R_s measurement, ±0.25% for B or ±0.75% for As. The issues on repeatability and sample-to-sample variations are considered in the discussion.

5.4 Observation of native oxide growth

After an HF dip, a native oxide grows on samples on a timescale of several days to weeks. If the sample surface is well passivated after the dip, the initial phase of the growth process can be observed.

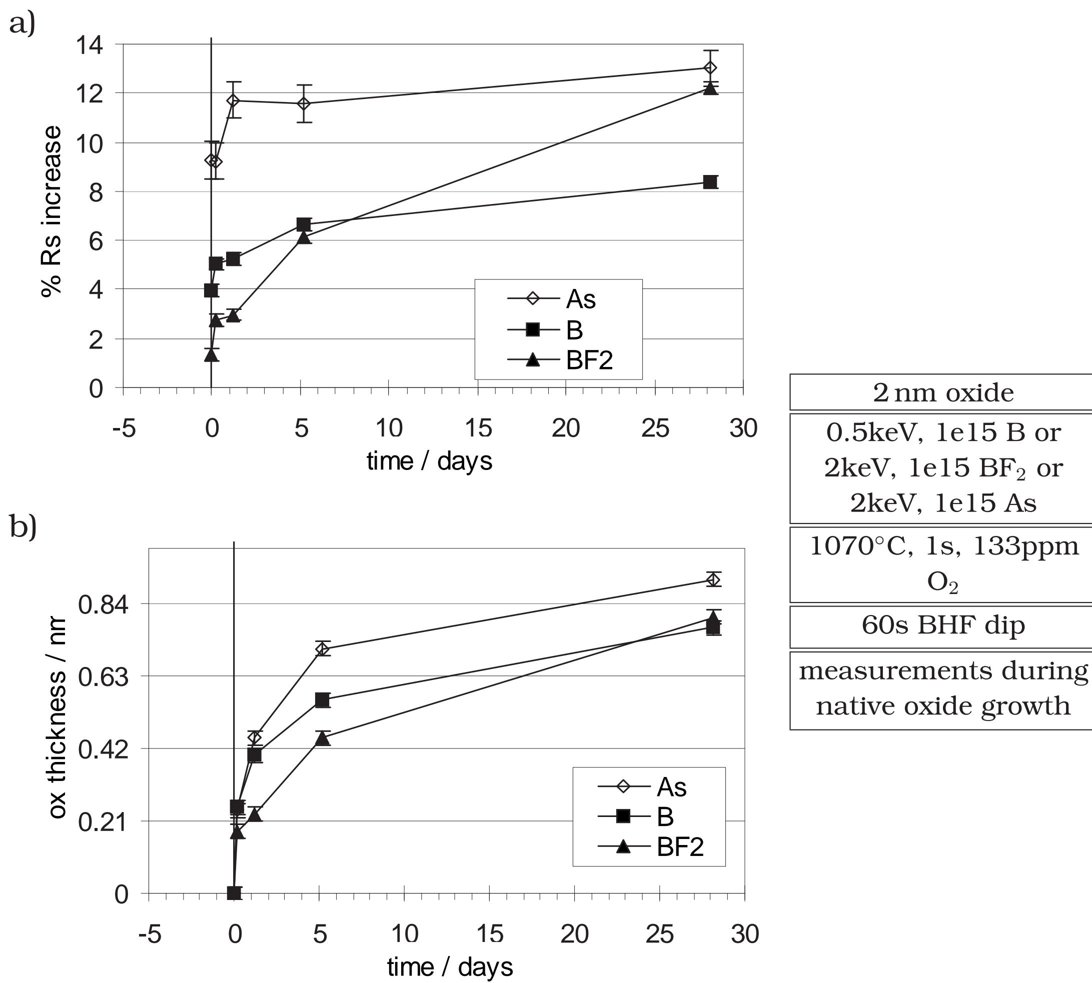

Figure 5.6: Observation of R_s increase (a) and oxide thickness (b) after a BHF dip, as a function of time. (Samples from lot P020334)

This section will describe a series of experiments to investigate the change of sheet resistance during the regrowth process of a native oxide. From the results, information on the concentration of active dopants in the near-surface region of a sample can be extracted.

5.4.1 Experimental conditions

For the measurement series, the following procedure was established: For each wafer, two samples are prepared, labelled “XPS” and “FPP” sample. They are etched simultaneously (in reality immediately after each other) to ensure identical processing. Immediately after the HF dip, a series of measurements is started. At well defined moments after the dip (e.g. immediately, after 6 h, one day, two days, one week, four weeks), sheet resistance (FPP) was measured on one sample, and oxide thickness (XPS) on the other. Care is taken to continue the measurements simultaneously on both samples. Typically, a precision of ± 10 minutes in the beginning and a few hours at the end of the series (after several weeks) are considered sufficient. Between each two measurements, the samples are stored together or under identical conditions to ensure an identical oxidation on both samples.

Since the FPP measurement might be sensitive to sample shape and the precise measurement spot, care is taken to measure the sheet resistance always on the same spot in the middle of the FPP sample. Using this technique, the error of the R_s measurement is $\ll 1\%$, as described above in section 5.2.

XPS measurements are done in vacuum. This contradicts the requirement that both samples should be kept in the same ambient during the entire experiment time. Therefore, only a small piece is broken off the XPS sample for every measurement in the series. While the large sample is stored in the cleanroom ambient, the small piece is placed in the XPS tool for the oxide thickness measurement. This technique allows also to load pieces of several samples into the machine at the same time. The vacuum ensures that during the measurement process, which takes more than one hour per sample, no additional oxide is growing. After the measurement, the measured pieces are discarded.

The results of a typical series of measurements for a B, a BF_2, and an As sample are shown in fig. 5.6.

The experiment took about one month to complete. Initially, R_s was measured before the HF dip as a reference basis. The data points were measured immediately after the HF dip, then 5 hours, 1 day, 5 days and 28 days after the dip. The oxide

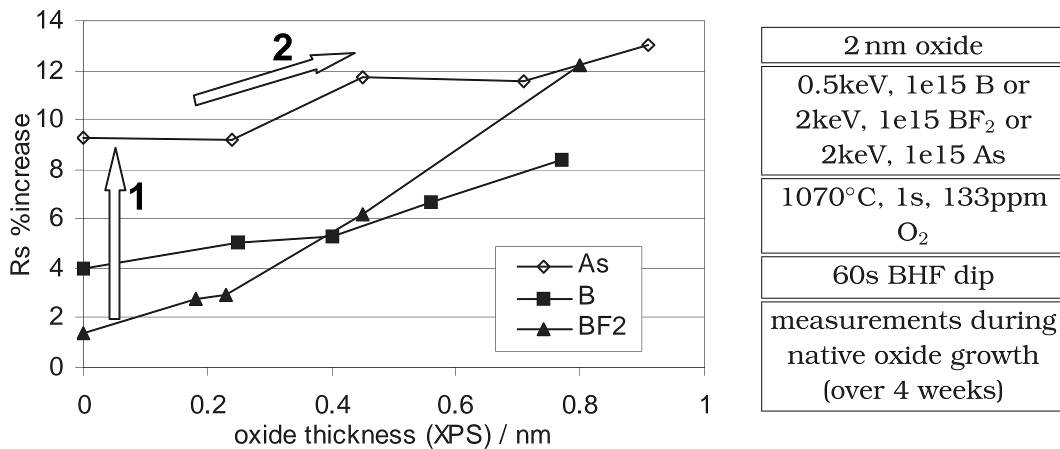

Figure 5.7: Oxide thickness vs. sheet resistance increase. The values are taken from the experiment given in fig. 5.6. (Samples from lot P020334)

thickness values show clearly that immediately after the dip (i.e. within 15 to 30 minutes), no oxide at all is seen on the wafer (this is precise to around 0.1 monolayers of oxide). Then, the oxide initially grows fast to reach between 0.3 to 0.5 nm after one day. After one month, the oxide thickness is around 0.8 to 0.9 nm. At this moment, oxidation hasn't yet stopped entirely, but slowed down to levels that are negligible for this work.

5.4.2 Measurable electrical effects

As discussed before (section 4.7), the pile-up is removed together with the oxide during an HF dip. During the described experiment, there are therefore two processes that might change the sheet resistance:

1. The pile-up is removed together with the oxide, but the silicon is not etched. Any active dopant atoms in the pile-up are lost.

2. During the subsequent native oxide regrowth, silicon is consumed. Dopant atoms located in the consumed layer of silicon are deactivated. The deactivation mechanism can be either incorporation into the oxide (most probably the case for B), or the creation of a very small, mostly inactive interfacial pile-up (possibly the case for As). Due to the small number of affected dopants, ERD resolution is not sufficient to distinguish the two cases.

Fig. 5.6 shows the typical evolution of sheet resistance and oxide thickness after an HF dip, with R_s given as percentile increase compared to the last mea-

surement before the HF dip. In fig. 5.7, these data are combined to show the sheet resistance increase as a function of the oxide thickness.

The results given in figure 5.7 show two distinctive features:

1. Immediately after the HF dip, a jump in sheet resistance is observed of around 4% for B and 10% for As implanted samples.

2. After that, an increase in sheet resistance is observed, which is mainly linearly dependent on the native oxide thickness.

It stands to reason that the initial jump in sheet resistance is due to the pile-up removal during the HF dip, and that the following linear increase in sheet resistance is caused by a continuous incorporation and deactivation of dopants into the growing native oxide. Therefore, the initial 4-10% R_s jump seen in fig. 5.7 indicates that the pile-up contributes significantly to the junction conductivity.

At first sight, the results suggest that the pile-up in the As case has a much higher contribution to the total conductivity than the one in the boron case. In the region below the pile-up, both samples seem to exhibit the same active dopant concentration, since the gradients of the curves are very similar. For BF_2, on the contrary, the pile-up seems to have a relatively little influence, whereas the region below the pile-up presumably contains a much higher active concentration than in the case of the B and As samples.

However, various side effects influence the measured sheet resistance values. They will be discussed in the next two sections. It will be shown that the value of the initial jump in sheet resistance is not very accurate, whereas the R_s increase after a sufficient native oxide growth (i.e. at the right end of the lines in fig. 5.7) can be considered as a quite reliable result.

5.4.3 Band bending at the surface

The HF dip is generally known to leave the surface in a well passivated state. This is consistent with the result described above, where no oxide was seen on the samples immediately after the HF dip. The time the sample needs before oxidation starts, depends strongly on the sample. It was observed to vary between one hour and more than one day for the different experiments.

However, even a good surface passivation can include charged surface states. If the number of surface states is high enough to pin the fermi level to a certain value, band bending occurs, which leads to a depletion or accumulation in the

junction's top layer, for both p- and n-doped material. On some B doped samples, even a small sheet resistance decrease was observed immediately after the HF dip.

Typical surface states can reach concentrations of around $10^{13}\,\mathrm{cm}^{-2}$ [93, 117] for a native oxide. This corresponds to around one percent of the implanted dose (around $10^{15}\,\mathrm{cm}^{-2}$). In the extreme case, all surface states are charged. Depletion or, respectively, accumulation of an equal number of carriers then leads to a resistance change in the order of one percent.

It is extremely difficult to quantify these surface states. Their number and their charge state depends strongly on the sample type, small variations in the oxide etching process, the storage conditions after etch, the cleanroom light, contaminations in the etching or rinsing bath and so on.

These effects alter the height of the initial R_s jump after the HF dip, and make it very difficult to analyze for absolute results. Most likely, the result variations between different experiments described above are caused by such non-controllable electrical surface effects. Therefore, the immediate measurement after the HF dip (i.e. the initial R_s jump) will not be used for a quantitative analysis during this work. However, it is possible to compare samples that were processed together in one experiment, in order to extract qualitative information on the pile-up.

The surface states disappear when the native oxide has grown to a stable layer of 0.5 nm of thickness or more. Then, the oxide-silicon interface has a sufficiently good quality, and any initial charging effects have disappeared [122]. This is the case for the results measured at the end of each measurement series, i.e. the rightmost points in fig. 5.7. From these data, it is possible to obtain reliable quantitative information on the pile-up's electrical properties.

After native oxide growth, the density of dangling bonds at the oxide-silicon interface is in the order of $10^{13}\,\mathrm{cm}^{-2}$ instead of $10^{12}\,\mathrm{cm}^{-2}$ for a thermal oxide, according to [92]. This number is two orders of magnitude lower than a typical total junction dose. Any carrier trapping or charging effects will therefore have small effects on the sheet resistance. For precise results, this should however be kept in mind.

5.4.4 Side effects reducing mobility

Various other effects can reduce the mobility of carriers close to the surface, and thereby additionally increase the sheet resistance of the entire junction. The most important are:

- Carrier mobility reduction in the proximity of the oxide

- High concentration of substitutional or clustered dopants, or other defects in the pile-up
- Increasing surface roughness during native oxide growth

These effects increase the junction's sheet resistance and can lead to an under- as well as an overestimation of the pile-up's contribution to conductivity. In section 5.5.3, an experiment will be presented to investigate the influence of these effects.

5.5 Conductivity of the pile-up

5.5.1 Quantitative results

On several samples, the surface oxide was removed in a 60 s BHF dip. Sheet resistance increase and oxide thickness were observed during regrowth, analogous to the data presented in figure 5.7. In order to avoid the influence of any possible electrical surface effects during the initial stage of oxidation, only the last measured values of each sample (right-hand end of fig. 5.7) were used to extract electrical information.

The following procedure is used to estimate the active dose in the pile-up: From SIMS, information on the total dose is obtained. It is assumed that the entire dose in the bulk part of the junction is active, as it was proven by SRP (cf. page 66). Together with the sheet resistance before the HF dip and after native oxide regrowth, the active dose lost during the entire process is calculated. From the XPS measurements, the native oxide thickness and therefore the amount of silicon consumed is known with quite good precision. By dividing the lost dose by the layer thickness, an average doping concentration in this layer is calculated. For As, additionally the amount of silicon lost due to the BHF dip needs to be considered. Finally, an estimation of the various errors is done.

Fig. 5.8b shows results for the samples that were presented in fig. 5.7. Three values are given for each sample. First, the active dopant concentration is given as a reference, as it was measured by SIMS in a depth of 5-10 nm. Secondly, the estimated concentration of active dopants is given, as described above. This value is an averaged concentration over the entire layer of silicon removed during the process. Thirdly, the highest value for each sample indicates the expected active concentration in a pile-up, assuming that the active dose is not homogeneously distributed over the removed layer, but concentrated in a pile-up that is only one

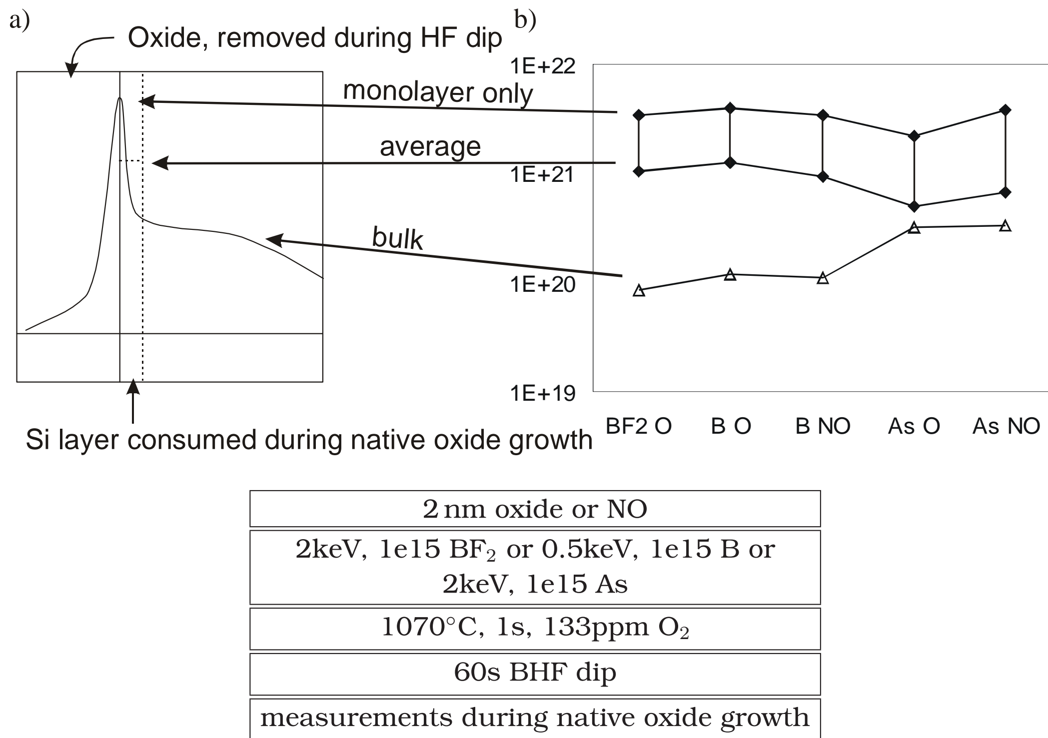

Figure 5.8: a) Scheme of the material and dopants removed during an HF dip and native oxide regrowth. b) Estimated active concentration in the investigated surface layer. The active concentration in the bulk silicon is given for comparison. The samples discussed here are the same as in figures 5.6 and 5.7. Two additional samples with oxynitride (NO) are included in this evaluation. (Samples from lot P020334)

monolayer thick. The schematic drawing in fig. 5.8a explains the meaning of each of the values.

The two top lines indicate a range, in which the real value is expected. An assessment on how precise the two limits are will be given in the next sections.

5.5.2 Interpretation of the measured concentration

The calculation as described and demonstrated above includes a variety of assumptions. For a quantitative interpretation of the results, several major issues need to be considered. On the one hand, the distribution of the active dopants over the entire investigated layer is not known and can only be estimated from the ERD profiles:

- As it was demonstrated with the ERD profiles, the pile-up is expected to be not more than 2-3 monolayers thick, thus much thinner than the entire layer removed during the measurement procedure. Therefore, the upper value of the range given represents the more realistic calculation.
- The mobility of the pile-up's carriers is reduced due to the immediate vicinity of the oxide interface as well as to the high concentration of active or clustered dopant atoms in the pile-up. Assuming a constant mobility (as it was done for the calculation above) therefore leads to an underestimation of the actual active concentration. This indicates again that the actual values are more likely to be found at the high end of the given range.

On the other hand, the measured sheet resistance increase might not only be due to the removal of active dopants, but also to other effects:

- During HF dip and native oxide regrowth, the surface roughness might increase microscopically, which additionally increases the sheet resistance. At least for the As samples etched with BHF, a rougher surface was observed with TEM. Since the calculation above assumes dopant loss as the only cause for the R_s increase, this effect leads to an overestimation of the active dose, indicating that the actual active concentration might be much lower than estimated above.
- As discussed above, charged interface states can accumulate or deplete the carriers close to the interface and thereby influence the total conductivity. If the number and kind of these states is different for the annealed oxide and

the regrown native oxide, this difference disturbs the measurement of the sheet resistance increase in either direction.

- Various other detrimental effects during etching or during the R_s measurement might influence the measured R_s increase in both directions.

These issues are very difficult to measure directly or to estimate quantitatively. For a better understanding of the importance of these effects, an experiment was carried out that directly investigated he influence of most of the measurement artifacts to the results. It will be presented in the following section.

5.5.3 Repeated HF dips

A set of four wafers was each etched four times in the automatic wet bench. After every dip, the wafers were given several day's time to develop a native oxide of 0.6 to 1 nm. After the first and the last dip, sheet resistance and oxide thickness were investigated as described above. The two intermediate HF dips were done only to remove more material from the wafer surface and were not investigated in detail. Since this experiment was carried out on full wafers, using DHF dips in the automatic wet bench, and ellipsometry to determine the oxide thickness, the results are less precise than the ones presented above. However, silicon etching was avoided on the As wafers.

The first HF dip removes the pile-up and gives information about the top Si layer just below the oxide, including the pile-up. Since between each two dips, a native oxide was allowed to regrow, more and more material was removed from the surface. Therefore, the results of the fourth HF dip allow the extraction of information on a layer of silicon in 1 to 2.5 nm depth. The depth of this layer can be estimated from the native oxide thickness after each HF dip; the results for the different samples are given in fig. 5.9b.

Figure 5.9a shows the results of the first and last HF dip. The bulk doping concentration (measured by SIMS) is given as a reference. Two main conclusions are drawn from these results.

Firstly, the investigation of the first HF dip results in a higher concentration value than the one of the last HF dip. This indicates that the active dopant concentration close to the surface is definitely higher than in a depth of 1-2 nm, for these samples by a factor of 2 to 5.

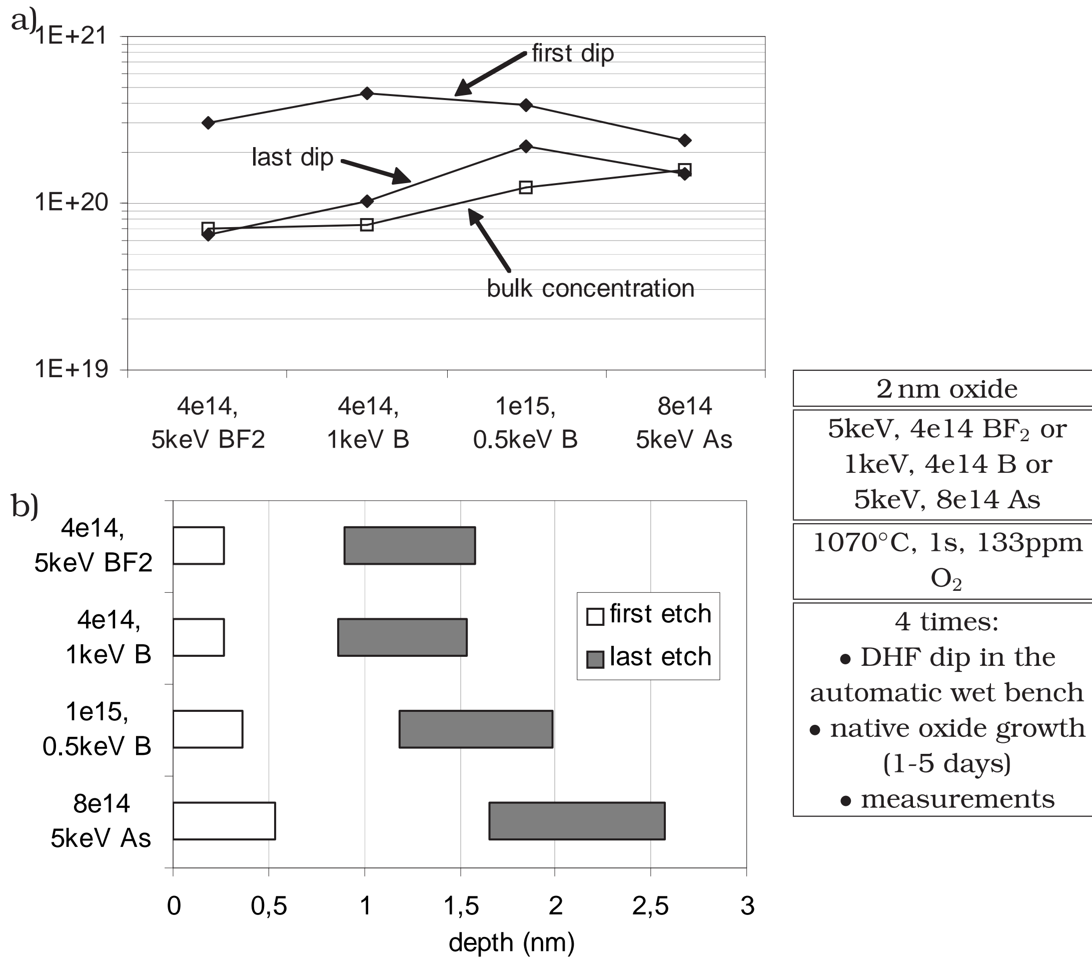

Figure 5.9: a) Active pile-up concentration, averaged over the removed layer. From a total of 4 HF dips, data from the first and last dip are displayed. The active concentration in the bulk part of the junction is given for comparison.
b) Depth of the investigated layer of the first and the last HF dip. The calculation of these values is based on the amount of native oxide grown after each dip, as measured by ellipsometry. (Samples from lot T010941)

Secondly, the results of the last HF dip exhibit a quite good agreement with the SIMS data. While for two samples, almost identical concentration values are seen, the other two samples exhibit a difference of only a factor of two.

It has to be noted that for the evaluation of this experiment it is assumed that the side effects influencing the sheet resistance measurements are similar for each of the consecutive HF dips. This is plausible for most of the expected effects. E.g. the interfacial roughness is expected to increase by the same amount after each dip, at least in a first approximation.

A priori, this result is obtained only for the automatic wet bench dips. However, any differences between the automatic DHF and the manual BHF or DHF dips are believed to only influence the initial phase of the native oxide growth. After several hours or days, when an oxide of more than 0.5 nm has grown, no effects of the initial surface conditions are seen any more [122, 123]. Therefore the results gained here are also applicable to manual DHF or BHF dips.

From this experiment, it is concluded that the high active concentration found close the the oxide interface is not a measurement artifact, but an actual effect of a highly active interfacial pile-up. The calculation accuracy is subject to various detrimental effects and seems to slightly overestimate the active concentration by up to a factor of 2. For a calculated value of x, the actual active dopant concentration is therefore expected between $0.5\,x$ and x.

However, it should be kept in mind that this experiment only compares the averaged concentration values of each removed layer with the SIMS data. As discussed above, the pile-up is much narrower than the removed layer of the first HF dip, and consequently the active concentration in the pile-up itself is expected much higher than the values given in fig. 5.9.

5.5.4 Overview and comparison

The results show that very high active concentrations are seen in the pile-up. For boron, an active pile-up concentration of 1-3 $\cdot 10^{21}$ cm^{-3} is observed, exceeding the bulk solid solubility limit of $2.5 \cdot 10^{20}$cm^{-3} [26] by around one order of magnitude.

For arsenic, the bulk solid solubility limit of $1.5 \cdot 10^{21}$cm^{-3} is already very high, the active pile-up concentration probably exceeds it by little. Also here, the pile-up is seen to contain a much higher active dopant concentration than the junction's bulk part.

These values correspond to only around 10-20% of the pile-up dopants being active for all samples. This estimation demonstrates that while most of the

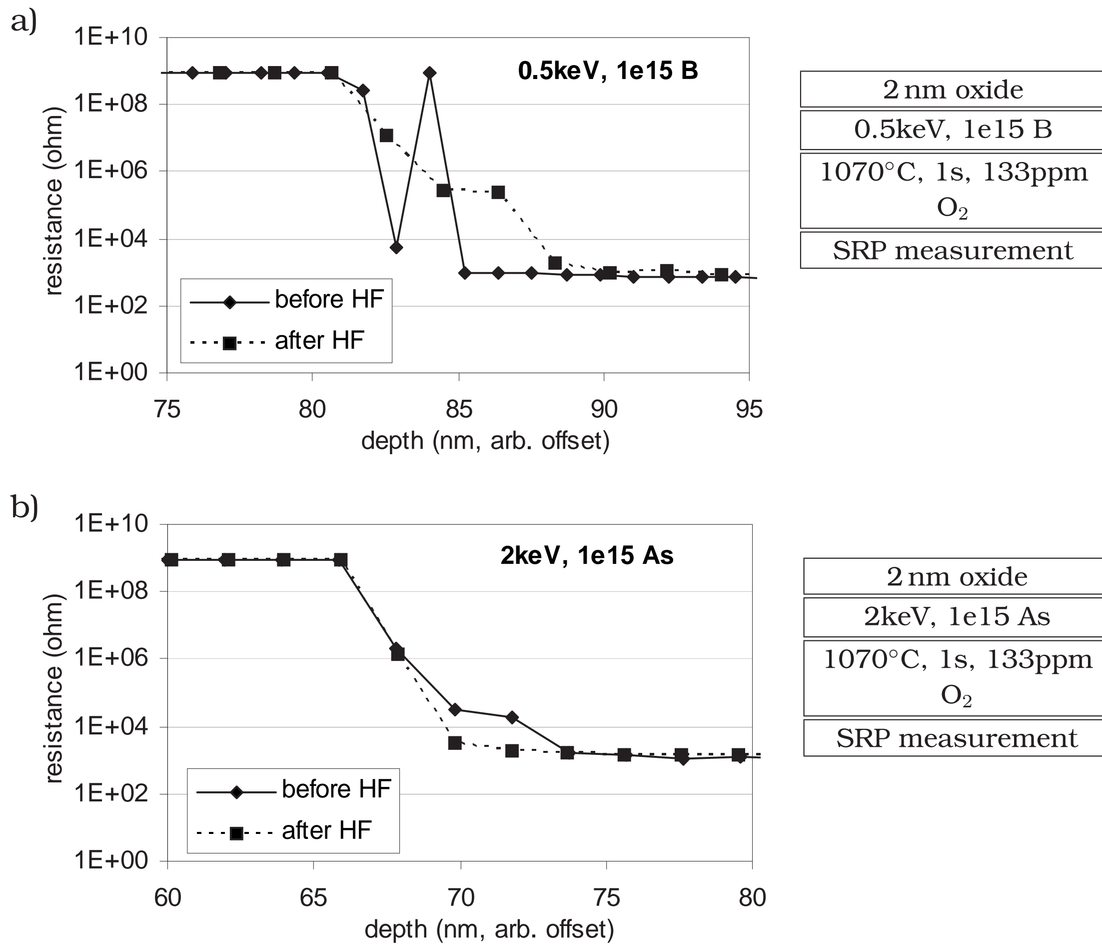

Figure 5.10: Raw resistance values from SRP measurements. (Samples from lot P030275)

pile-up dopants are actually deactivated, still very high active concentrations are observed.

5.5.5 SRP measurements

SRP measurements with the best available resolution were made to investigate the conductivity of the silicon directly below the oxide. As discussed above (section 2.4), SRP depth resolution is not sufficient to resolve the pile-up. This is due to the tip geometry (tip radius, imprint depth), but also to the models integrating over the whole junction depth.

By investigating directly the raw resistance values measured between the two tips of the SRP tool, all model-related errors can be avoided. Figure 5.10 shows

raw resistance profiles measured by SRP around the interface. A vertical distance of 1.2 to 1.9 nm between two measurement points was achieved[4].

The electrical resolution of these profiles is in the order of 3 to 5 nm. It is mainly determined by the diameter of the probe tip. The first contact between probe tip and a junction is made when the probe tip just penetrates the oxide and touches the conducting layer. For good, low-resistive conduction, the tip has to penetrate several nm into the junction to make a good contact.

Consequently, under optimal conditions a highly active pile-up should lead to a steeper transition from the high resistance region of the oxide to the low resistive doped junction. Such an effect is visible for the B profiles (fig. 5.10a) – the as-annealed sample (solid line) exhibits a steeper transition than the etched sample (dotted line). However, an opposite effect is observed for arsenic.

Even for the raw SRP data, the resolution is too low to observe the very narrow pile-up. Furthermore, several effects are likely to adversely affect the traceability of the pile-up: It is very difficult to understand and model the impact of the probe penetration to the immediate surroundings, such as material redistribution and high local pressure [75]. Also, geometrical inhomogeneities due to surface adsorbates or bevelling damage are expected to be bigger than the pile-up width; they are probably responsible for the features seen in fig. 5.10.

It is concluded that SRP is not able to confirm the electrical activation of the pile-up atoms, but cannot disprove the theory either.

5.6 Influence of implant and anneal parameters

The previous sections quantitatively discussed the pile-up conductivity by evaluating the R_s increase after around 1 nm of native oxide growth, in order to avoid surface charging effects during the initial stages of oxidation. This section will evaluate R_s data measured immediately after an HF dip. No quantitative conclusions are possible, but comparisons between different samples allow the extraction of qualitative information on the pile-up.

Before the HF dip, the total junction conductivity is composed of the bulk and the pile-up conductivity. After the HF dip, the pile-up has disappeared. For all samples of one species, the doping concentration in the bulk part of the junction is almost identical, therefore identical surface conditions are expected after the

[4] The precise value depends on the bevelling angle, which is always slightly different for the samples, depending on the preparation.

HF dip. Since the HF dipping conditions can be kept to a very constant level for samples dipped immediately after each other (cf. the repeatability discussion above), it is possible to compare the different samples to each other, as long as they are etched and analyzed together (i.e. within one experiment).

In the following, data about the initial R_s jump after the HF dip will be presented and discussed. As discussed above, this jump is mainly due to the removal of the active dopants in the pile-up. Its height gives an idea about the pile-up conductivity, relative to the conductivity of the bulk part of the junction.

Since large variations between experiments are expected, only samples from within one experiment should be compared in the following discussion, as they are shown within each figure. Values from different figures should not be compared.

All samples discussed in the following sections were implanted with 1e15, 0.5 keV B or 1e15, 2 keV As and annealed in a 1070 °C, 1s spike anneal with 133 ppm oxygen in the ambient, except where noted.

5.6.1 Different screening oxides

Figure 5.11a compares a sample with a pure oxide to one with a heavily nitrided oxide. For nitridation, a RPN (Remote Plasma Nitridation) process was used that results in a similar nitridation as the DPN process discussed in section 4.4.4. The results show a slight pile-up increase for As and a slight decrease for B due to the presence of nitrogen at the interface. While the effect for B (-0.3 percentage points) is within the repeatability error, the change in the As case (+0.75 percentage points) is considered just significant.

The ERD results presented in section 4.4.4 showed that the total size and shape of the pile-up for As is not influenced by the presence of nitrogen. It is therefore suggested that nitrogen slightly enhances the activation of the dopants in the pile-up. For B, nitrogen was seen to increase the dose in the oxide side of the pile-up, whereas the silicon side showed no difference. This is consistent with the results presented here.

5.6.2 Influence of oxidation

Figure 5.11b shows the influence of oxidation during anneal. Three As and three B samples were annealed in a 1070 °C, 1s spike anneal, with an oxygen concentration in the ambient between 0% and 5%. As discussed before (section 4.5.1), a slight oxide growth is also seen during the “0%” anneal.

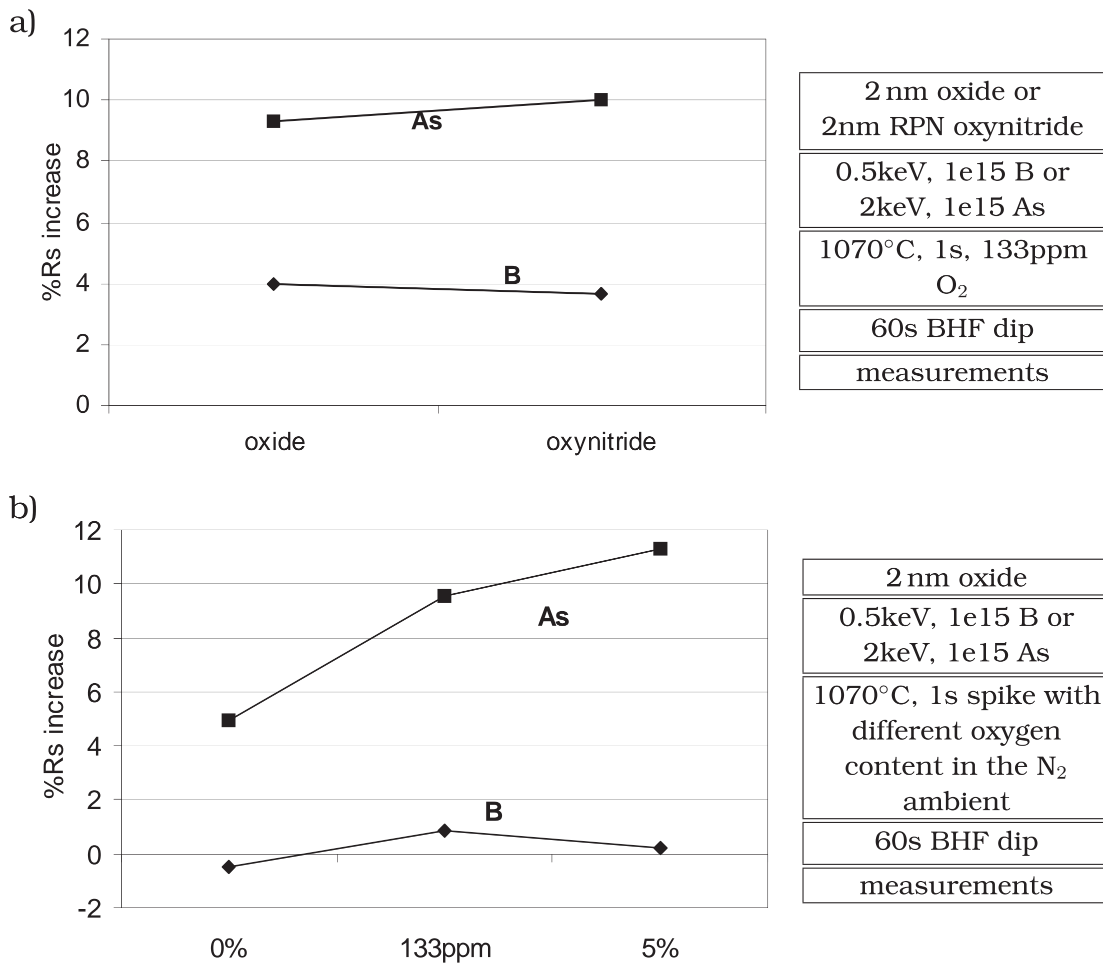

Figure 5.11: Sheet resistance increase measured immediately after the HF dip for B and As samples. The influence of different screening oxides (a) or different oxygen content in the annealing ambient (b) are compared. (Samples from lot P020334 (a) and P020465 (b))

For B, no significant influence of the oxidation during anneal on the pile-up is observed. This is consistent with the ERD profiles shown before (section 4.5.2), where oxidation was observed to even slightly reduce the pile-up height at the interface. This should lead to a reduced amount of pile-up atoms on the silicon side of the interface.

However, for the As implanted samples, the electrical results show a dependence of the initial R_s increase on the amount of oxygen in the annealing ambient, while the ERD profiles revealed no significant differences between the 133 ppm and 5% cases.

It is therefore suggested for the arsenic case, that the activation of the pile-up dopant atoms is better if the surface is oxidized during anneal. The rearrangement process that takes place during oxidation seems to improve the probability that a given dopant atom is activated and incorporated on a substitutional lattice site instead of in an interfacial dopant cluster. The results therefore suggest that the activation of the pile-up dopants is governed by the oxidation and not so much by the thermal activation processes known from bulk silicon.

5.6.3 Influence of the thermal budget

The influence of the spike annealing temperature on the electrical properties of the pile-up is investigated. The samples were implanted with 2 keV BF_2, 0.5 keV B or 2 keV As (all 1e15 cm^{-2}) and annealed under standard conditions (1s spike anneal, 133 ppm oxygen), but with the peak temperature varied between 950 °C and 1070 °C.

Figure 5.12b shows that the pile-up becomes much more important for a low thermal budget. For 950 °C, the relative pile-up conductivity is by 9 (As) or 5 (B) percentage points higher than at the high thermal budget of a 1070 °C anneal.

However, it has to be kept in mind that for the low temperatures, a much higher total sheet resistance is seen, because of a lower dopant activation in the bulk part of the junction. From a comparison of the R_s increase after native oxide regrowth with the absolute R_s data, the conductivity of the pile-up can be estimated, analogous the the technique presented in the previous section (5.5.1). The conductivity in the pile-up and the silicon top layer is displayed in 5.12b. Only for the 950 °C arsenic case, a significantly lower value is seen, whereas for all other cases, the pile-up conductivity seems to be almost independent on the temperature.

[6] The values given here are calculated as $1/R_{s,pile\text{-}up}$, thus they indicate a sheet conductivity of the pile-up and the entire silicon layer deactivated during etch and native oxide regrowth.

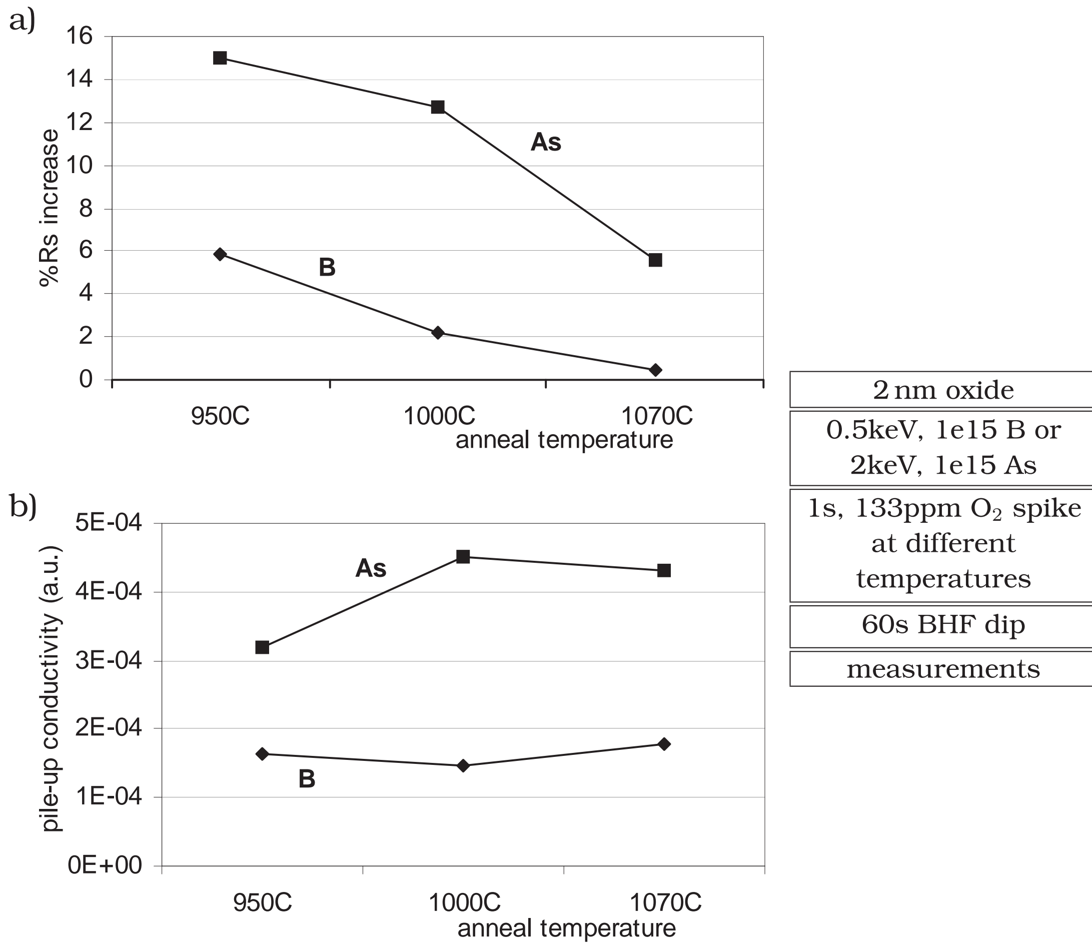

Figure 5.12: Sheet resistance increase after a BHF dip for B, BF_2 and As samples, spike annealed in a 133 ppm oxygen ambient for 1s at different temperatures (a). Graph b shows the estimated conductivity in pile-up in arbitrary[6] units for the same samples. (Samples from lot P020465)

This result agrees with the conclusion from the last section, indicating that the activation of the pile-up dopants is mainly governed by the oxidation process, and not so much by the thermal diffusion and activation mechanisms that control activation in the bulk silicon. The oxidation induced pile-up activation seems to depend much less on the processing temperature than the bulk activation.

5.7 Summary

In order to investigate the electrical activity of the pile-up dopants and their contribution to the overall junction conductivity, a special surface treatment was used. The surface oxide together with the pile-up was removed in a BHF or DHF dip. The consecutive native oxide regrowth process was observed, measuring simultaneously sheet resistance and oxide thickness.

It is observed that the removal of the pile-up together with the surface oxide causes an initial jump in the sheet resistance of up to 10%. This is an indication that at least a part of the pile-up dopants is electrically active. However, initial electrical surface effects strongly influence the measured value. The subsequent native oxide regrowth consumes silicon, deactivates the dopants in the consumed layer and thereby causes a continuous increase in sheet resistance.

From the results, an estimation of the active dose in the pile-up was calculated. While in all cases, only 10-20% of the dopants trapped in the pile-up are active, very high active concentrations of $10^{21}\,\mathrm{cm}^{-3}$ and more were observed in the pile-up, much higher than the typical active bulk concentration of the junctions. At least for B, the estimated active concentration in the pile-up also exceeds the bulk solid solubility limit.

Such high active concentrations are possible because the formation process is governed by trapping of dopant atoms at interfacial sites during the oxidation process. The thermodynamical solubility limits of the bulk material don't apply for this case. Also, oxide induced stress in the silicon close to the interface is believed to facilitate the activation of such high concentrations [124].

A comparison of results from different samples shows the influence of the various anneal parameters. While the presence of nitrogen in the oxide increases the total dopant dose in the pile-up, it has no significant effect on the pile-up conductivity. Oxidation during anneal is observed to increase the activation of As dopants, but has no significant influence on B implanted junctions. Reducing the temperature of the spike anneal drastically deteriorates the activation of the dopants in the junction's bulk part, but has only little influence on the pile-up ac-

tivation. This demonstrates once more that the pile-up formation is not so much governed by temperature-dependent thermodynamical processes, but more by the lattice rearrangement at the interface during oxidation.

6 Conclusions

In order to follow the path determined by the ITRS, the IC manufacturing industry introduces new CMOS technology nodes every 2-3 years. Every technology node reduces the length and width of all transistor or interconnect structures roughly by a factor of 0.7, thereby doubling the number of transistors per unit area. If the lateral extension (i.e. the gate length) of a transistor is reduced, also the vertical size needs to be scaled accordingly, in order to avoid short channel effects and other detrimental electrical effects. Consequently, the dopants in the source and drain contacts ("junctions") come closer to the surface with each technology node, and the influence of any surface effects on the dopant behavior becomes increasingly important. The aim of this thesis was to investigate one of these effects, the dopant pile-up found at silicon-oxide interfaces.

The samples analyzed for this work were implanted with BF_2, B and As at ultra-low energies and spike annealed under various conditions. Since SIMS proved not to be accurate enough for near-surface profiling, high resolution ERD and RBS were used to gain accurate dopant profiles with sub-nm resolution. SIMS profiles helped to optimize the dose accuracy of the measurements. Complementary results were obtained from TEM and EDX measurements. SRP and sheet resistance measurements were used to investigate the electrical activity of the dopants.

The samples were observed to lose up to 50% of the nominally implanted dose during implant (self-sputtering) and anneal (outdiffusion). From the dopants retained in the sample, up to 70% are concentrated in a pile-up at the silicon-oxide interface, for the range of doses investigated. While for boron, this percentage increases for higher implanted doses, the opposite is the case for arsenic. The results suggest that for boron, doses of not more than $1\text{-}2 \cdot 10^{15}\,\mathrm{cm}^{-2}$ can be retained and activated in the bulk part of the junction, for the range of implant energies and the spike anneal used. For arsenic, no indication for such a limit was found.

Oxidation during anneal was found to be necessary for pile-up creation. A model is proposed that includes the atomic rearrangement at the silicon/oxide interface during oxidation as the main cause for pile-up formation. During the

rearrangement process, trapping sites like dangling bonds are created that block dopant atoms in a very narrow region. The width of the pile-up is therefore restricted to the immediate vicinity of the Si/oxide interface, thus it is expected to be not more than around two to three monolayers wide.

However, if a significant amount of oxide is grown during anneal, an additional effect needs to be considered. Dopants are trapped into the interface during the entire annealing process. In the case of boron, every atom stays at its trapping site, even if the interface continues to proceed. Therefore, the final pile-up extends measurably into the oxide, and the final pile-up width is determined by the amount of oxygen grown during anneal. Arsenic, on the contrary, is not soluble in oxide. The moving oxide/silicon interface therefore drags the already accumulated pile-up dopants along, and consequently the pile-up never becomes wider than the mentioned few monolayers.

The second important factor for the formation of a pile-up is an uphill diffusion of the implanted dopants towards the pile-up. This diffusion is mediated by the large amounts of silicon self-interstitials that are present in deeper regions of the sample due to the implant process. During anneal, they diffuse towards the surface, which acts as a sink, and drag dopant atoms along. Arriving at the Si/oxide interface, the dopants get trapped due to the mechanism described above.

These observations were made for conventional junctions, implanted with ultra low energies and annealed in a RTP spike anneal. For SPER, a similar, but much less pronounced pile-up is observed at the Si/oxide interface, at least for boron. Additional to the uphill diffusion mechanism, it is proposed that the dopants are carried along with the moving interface between the amorphous and the crystalline silicon.

Electrical investigations have shown that high active dopant concentrations are observed in the pile-up. While only a small fraction of only 10-20% of the pile-up dopants are active, high active concentrations of $10^{21}\,\mathrm{cm}^{-3}$ and above are observed, far above the typical active concentration in the bulk part of the junction. At least for boron, these values also exceed the bulk solid solubility limit by up to one order of magnitude.

It is proposed that the activation mechanism is not the one of a thermal equilibrium, as for conventional diffusion and activation. Instead, the interfacial rearrangement process during oxidation creates sites that can trap dopants in very high doses. A small fraction of these dopants is active, but their number is not limited by the thermodynamical equilibrium that leads to a solid solubility limit in bulk silicon.

It is suggested that the activity of the pile-up is of special interest in two main areas in a junction. First, the contact resistance between channel and extension might be reduced by the fact that channel as well as pile-up are located next to the interface to the gate oxide. Secondly, a pile-up at the interface between HDD and silicide could influence their contact resistance by narrowing the Schottky barrier. Further research needs to be done to investigate how strong these effects are and whether they can be used to optimize the total resistance of a transistor.

With this work, profiles of the pile-up are presented in unprecedented detail. For the first time, high-resolution ERD and RBS profiles are used to investigate the pile-up with nanometer resolution, complementing and extending various theoretical and experimental results published before. It is shown that the interfacial pile-up is not only trapping significant fractions of the implanted dopants, but also contributes to the junction conductivity. The goal for junction engineering is therefore not necessarily to avoid the pile-up, but rather to optimize its properties.

For the future, this work will serve as a basis for more extensive investigations on new junction formation technologies, such as SPER or LTA, that use very different physical mechanisms for junction formation. It will help to understand and to simulate the behavior of dopants in ultra-shallow junctions, a critical point in the development of the upcoming IC manufacturing technologies beyond the 65 nm node.

New analysis techniques such as SSRM or the Nanoprofiler will be able to add valuable results to the discussion of the electrical activity of the pile-up dopants, and allow a much more direct and precise observation of the electrical effects than it is possible today.

Acknowledgements

There are a lot of people whom I would like to thank for a variety of reasons.

First of all, I would like to thank my daily supervisor Richard Lindsay, without whom this thesis would never have been possible. He kept my work on track, helped to design and evaluate all my experiments, and he never got tired of interpreting the most strange results, even at the permanent risk of having his brain fried. Thank you very much to Prof. Dr. Fred Koch, who supervised my thesis and provided me with all the necessary scientific background. Many thanks also to the companies Infineon and IMEC, especially to Georg Tempel, Klaus Schrüfer, Reinhard Mahnkopf, Karen Maex and Rudy Cartuyvels for providing me with financial and technical support.

I am very grateful to Wilfried Vandervorst for the long discussions on the interpretation of his own and other group's analysis results. Many thanks to Andreas Bergmaier, Luc Geenen, Ilse Hoflijk, Olivier Richard and Andreas Rucki who spent a lot of their time analyzing my samples and evaluating the results with me, to the IMEC P-Line crew for processing my wafers, and to Bartek Pawlak, Rob Lander, Ray Duffy, and all the people of the USJ team at IMEC for lots of discussions and help.

Last but not least, my special thanks go to Amal, Aude, Bartek, Gosia, Guilherme, Ric, Valia and many others for being great friends and colleagues in Leuven and Brussels, and, most of all, to my parents for making my studies and this PhD thesis possible!

Abbreviations

BED	Boron Enhanced Diffusion
BEOL	Back End Of Line
CMOS	Complementary Metal-Oxide-Semiconductor (technology)
CMP	Chemical-mechanical polishing
CVD	Chemical Vapor Deposition
DPN	Decoupled Plasma Nitridation
DSA	Dynamic Surface Anneal
EDX	Energy Dispersive X-ray spectroscopy, also EDS
EELS	Electron Energy Loss Spectrometry
ERD	Elastic Recoil Detection
FEOL	Front End Of Line
FET	Field Effect Transistor
FPP	Four Point Probe (also 4PP)
HDP-CVD	High Density Plasma CVD
IC	Integrated Circuit
ICM	IC Manufacturer
I/I	Ion Implantation
ITRS	International Technology Roadmap for Semiconductors
LTA	Laser Thermal Annealing
MEIS	Medium Energy Ion Scattering
MOS	Metal-Oxide-Semiconductor (transistor technology)
OED	Oxygen Enhanced Diffusion
PLAD	Plasma Doping
P^2LAD	Pulsed Plasma Doping
PS	Probe Spacing
PVD	Physical Vapor Deposition
RBS	Rutherford Backscattering Spectrometry
RPN	Remote Plasma Nitridation
R_s	Sheet resistance
RTA	Rapid Thermal Annealing
RTP	Rapid Thermal Processing
RTO	Rapid Thermal Oxidation

SACVD	Sub-Atmospheric CVD
SCE	Short Channel Effects
SIMS	Secondary Ion Mass Spectrometry
STI	Shallow Trench Isolation
SPER	Solid Phase Expitaxial Regrowth
SRP	Spreading Resistance Profiling
SSRM	Scanning Spreading Resistance Microscopy
TED	Transient Enhanced Diffusion
TEM	Transmission Electron Microscopy
TEOS	tetra-ethyl-ortho-silicate: Si-(O-$C_2H_5)_4$
ULEI	Ultra-low Energy (Ion) Implanter
USJ	Ultra-Shallow Junction
X_j	Junction Depth
XPS	X-ray Photoelectron Spectrometry

Bibliography

Even though all internet links listed in this bibliography were accurate during editing of this thesis, the author cannot guarantee that the respective pages will not change during or after publication. The pages referred to may change address, be removed from the internet (dead links), or their content may change at any time. The author has no influence on this.

[1] J. Frühauf, R. Lindsay, A. Bergmaier, W. Vandervorst, G. Tempel, K. Maex, G. Dollinger, and F. Koch. Electrical activity of B and As segregated at the Si-SiO_2 interface. In *Mat. Res. Soc. Symp. Proc.*, volume 717, page C3.4, San Francisco, CA, 2002.

[2] J. Frühauf, R. Lindsay, W. Vandervorst, K. Maex, A. Bergmaier, G. Dollinger, and F. Koch. Characterization of the B and As pile-up at the Si-SiO_2 interface. In *Proc. USJ-2003*, pages 399–404, Santa Cruz, CA, 2003.

[3] R. Lindsay, A. Lauwers, J. Frühauf, M. de Potter, and K. Maex. Effect of implant oxide on ultra shallow junction formation. In *Proc. USJ-2001*, Napa, CA, 2001.

[4] R. Lindsay, A. Lauwers, J. Frühauf, M. de Potter, and K. Maex. Effect of implant oxide on ultra shallow junction formation. *J. Vac. Sci. Tech. B, Microelectron. Nanometer Struct.*, 20(6):2225–8, 2002.

[5] W. Vandervorst, T. Janssens, R. Loo, M. Caymax, I. Peytier, R. Lindsay, J. Frühauf, A. Bergmaier, and G. Dollinger. An (un)solvable problem in SIMS: B-interfacial profiling. In *Proc. SIMS-XIII*, Nara, 2001.

[6] W. Vandervorst, T. Janssens, R. Loo, M. Caymax, I. Peytier, R. Lindsay, J. Frühauf, A. Bergmaier, and G. Dollinger. An (un)solvable problem in SIMS: B-interfacial profiling. *J. Appl. Surf. Sci.*, 203-204:371, January 2002.

[7] W. Vandervorst, T. Janssens, J. Frühauf, I. M. Ross, A. Cullis, J. A. Van den Berg, A. Bergmaier, and G. Dollinger. Assessment of the near-surface profiling capabilities of SIMS. In *Proc. USJ-2003*, Santa Cruz, CA, 2003.

[8] W. Vandervorst, T. Janssens, J. Frühauf, I. M. Ross, A. Cullis, J. A. Van den Berg, A. Bergmaier, and G. Dollinger. Assessment of the near-surface profiling capabilities of SIMS. *J. Vac. Sci. Tech. B*, (to be published).

[9] G. E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8), April 1965 (ftp://download.intel.com/research/silicon/moorespaper.pdf).

[10] Intel presentation, 2002 (http://www.fiatech.org/pdfs/stout.pdf).

[11] G. E. Moore. No exponential is forever... but we can delay 'forever'. In *International Solid State Circuits Conference (ISSCC)*, 2003 (ftp://download.intel.com/research/silicon/ Gordon_Moore_ISSCC_021003.pdf).

[12] ITRS, The International Technology Roadmap for Semiconductors, 2002 updated edition (http://public.itrs.net/).

[13] http://www.intel.com/research/silicon/mooreslaw.htm, 2003.

[14] M. H. Juang. A practical device scheme for designing the dopant profile of source/drain extension region in sub-quarter-micron p-MOSFET's. *Solid State Electronics*, 43(12):2209–13, December 1999.

[15] P. Packan. Scaling transistors into the deep-submicron regime. *MRS Bulletin*, 25(6):18–21, June 2000.

[16] J. D. Plummer and P. B. Griffin. Material and process limits in silicon VLSI technology. In *Proc. of the IEEE*, volume 89(3), pages 240–58, March 2001.

[17] R. Kasnavi, P. B. Griffin, and J. D. Plummer. Ultra low energy arsenic implant limits on sheet resistance and junction depth. In *2000 Symp. on VLSI Technology. Digest of Tech. Papers*, pages 112–13, Honolulu, HI, 2000.

[18] A. J. Murrell, E. J. H. Collart, M. A. Foad, and D. Jennings. Process interactions between low-energy ion implantation and rapid-thermal annealing for optimized ultrashallow junction formation. *J. Vac. Sci. Tech. B*, 18(1):462–7, January 2000.

[19] A. Agarwal, H.-J. L. Gossmann, A. T. Fiory, V. C. Venezia, and D. C. Jacobson. Ultra-shallow junction formation using ion implantation and rapid thermal annealing: Physical and practical limits. In *Electrochem. Soc. Symp. Proc.*, volume 2000-9, pages 49–60, Toronto, 2000.

[20] A. Agarwal, H.-J. L. Gossmann, A. T. Fiory, V. C. Venezia, and D. C. Jacobson. Limits of ultra-shallow junctions by ion implantation and rapid thermal annealing, 2000 (http://www.iwai.ae.titech.ac.jp/pdf/2000_0901/Agarwal.pdf).

[21] J. F. Ziegler. SRIM, The Stopping and Range of Ions in Matter. Implant simulation software package. Version 2003.20 (http://www.srim.org/).

[22] U. Littmark and J. F. Ziegler. Ranges of energetic ions in matter. *Phys. Rev. A*, 23(1):64–72, January 1981.

[23] X. Lu, L. Shao, X. Wang, J. Liu, W.-K. Chu, J. Bennett, L. Larson, and P. Ling. Cluster-ion implantation: An approach to fabricate ultrashallow junctions in silicon. *J. Vac. Sci. Tech. B*, 20(3):992–4, May 2002.

[24] A. T. Fiory and K. K. Bourdelle. Electrical activation kinetics for shallow boron implants in silicon. *Appl. Phys. Lett.*, 74(18):2658–60, May 1999.

[25] V. Privitera, E. Schroer, F. Priolo, E. Napolitani, and A. Carnera. Electrical behavior of ultra-low energy implanted boron in silicon. *J. Appl. Phys.*, 88(3):1299–306, August 2000.

[26] H. Föll. Hyperscripts: Semiconductors, http://www.tf.uni-kiel.de/matwis/amat/semi_en/kap_3/backbone/r3_1_1.html.

[27] A. Bratschun. The application of rapid thermal processing technolgy to the manufacture of integrated circuits—an overview. *J. Electron. Mater.*, 28(12):1328–32, December 1999.

[28] A. Agarwal, H.-J. Gossmann, and A. T. Fiory. Effect of ramp rates during rapid thermal annealing of ion implanted boron for formation of ultra-shallow junctions. *J. Electron. Mater.*, 28(12):1333–39, December 1999.

[29] D. F. Downey, S. W. Falk, A. F. Bertuch, and S. D. Marcus. Effects of “fast” rapid thermal anneals on sub-keV boron and BF_2 ion implants. *J. Electron. Mater.*, 28(12):1340–44, December 1999.

[30] V. I. Kuznetsov, A. B. Storm, G. J. Snijders, C. de Ridder, T. A. M. Ruijl, J. C. G. v.d. Sanden, and E. H. A. Granneman. Levitor 4000: An advanced RTP system based on conductive heat transfer. In *Proc. 197th Meeting of The Electrochem. Soc.*, page #564, Toronto, Canada, 2000.

[31] B. J. Pawlak, R. Lindsay, R. Surdeanu, X. Pages, W. Vandervorst, and K. v. d. Jeugd. The role of F with Ge pre-amorphisation in forming pMOS junctions for the 65nm CMOS technology node. In *Electrochem. Soc. Symp. Proc.*, volume 2003-14, pages 99–104, Paris, 2003.

[32] A. T. Fiory. Recent developments in rapid thermal processing. *J. Electron. Mater.*, 31(10):981–987, October 2002.

[33] R. Lindsay. Private communication.

[34] J. Ross, S. McCoy, K. Elliott, J. Gelpey, D. F. Downey, and E. A. Arevalo. Characterizing implant behavior during flash rtp by means of backside diagnostics. In *Proc. RTP 2002*, volume 02EX574, pages 99–105. IEEE, 2002.

[35] J. A. Kittl, P. G. Sanders, M. J. Aziz, D. P. Brunco, and M. O. Thompson. Complete experimental test of kinetic models for rapid alloy solidification. *Acta mater.*, 48:4797–811, 2000.

[36] X. Luo, S. B. Zhang, and S.-H. Wei. Understanding ultrahigh doping: The case of boron in silicon. *Phys. Rev. Lett.*, 90(2):026103/1–4, January 2003.

[37] B. Pawlak, R. Lindsay, R. Surdeanu, B. Dieu, L. Geenen, I. Hoflijk, O. Richard, R. Duffy, T. Clarysse, B. Brijs, W. Vandervorst, and C. J. J. Dachs. Chemical and electrical dopant profile evolution during solid phase epitaxial regrowth. In *Proc. USJ-2003*, page 227, Santa Cruz, CA, 2003.

[38] G. L. Olson and J. A. Roth. Kinetics of solid phase crystallization in amorphous silicon. *Mat. Sci. Rep.*, 3(1):1–78, 1988.

[39] R. Lindsay, B. Pawlak, J. A. Kittl, K. Henson, S. Giangrandi, R. Duffy, R. Surdeanu, W. Vandervorst, X. Pages, K. van der Jeugd, P. Stolk, and K. Maex. Leakage optimisation of ultra-shallow junctions formed by solid phase epitaxial regrowth (SPER). In *Proc. USJ-2003*, pages 65–72, Santa Cruz, CA, 2003.

[40] G. Chase. Ohms per square what!, ESD Journal, (http://www.esdjournal.com/techpapr/ohms.htm).

[41] T. Clarysse, D. Vanhaeren, and W. Vandervorst. Impact of probe penetration on the electrical characterization of sub-50 nm profiles. *J. Vac. Sci. Tech. B*, 20(1):459–466, January 2002.

[42] T. Clarysse, P. Eyben, N. Duhayon, M. W. Xu, and W. Vandervorst. Carrier spilling revisited: On-bevel junction behavior of different electrical depth profiling techniques. *J. Vac. Sci. Tech. B*, 21(2):729–36, March 2003.

[43] F. Giannazzo, V. Raineri, V. Privitera, and F. Priolo. High-resolution scanning capacitance microscopy by angle bevelling. *Mat. Science in Semiconductor Processing*, 4(1-3):77–80, 2001.

[44] S. T. Dunham, N. Collins, and N. Jeng. Improved analysis of spreading resistance measurements. *J. Vac. Sci. Technol. B*, 12(1):283–289, Jan-Feb 1994.

[45] V. Korenivski. The van der Pauw technique, http://tau.nanophys.kth.se/cmp/hall/node5.html.

[46] P. Eyben, S. Denis, T. Clarysse, and W. Vandervorst. Progress towards a physical contact model for scanning spreading resistance microscopy. In *Europ. Mat. Res. Soc. Proc.*, pages E–V.2, 2002.

[47] W. Vandervorst, T. Clarysse, and P. Eyben. Spreading resistance roadmap towards and beyond the 70 nm technology node. *J. Vac. Sci. Tech. B*, 20(1):451–8, January 2002.

[48] T. Clarysse, W. Vandervorst, E. J. H. Collart, and A. Murrell. Electrical characterization of ultrashallow dopant profiles. *J. Electrochem. Soc.*, 147(9):3569–74, September 2000.

[49] P. De Wolf, R. Stephenson, T. Trenkler, T. Hantschel, and W. Vandervorst. Status and review of two-dimensional carrier and dopant profiling using scanning probe microscopy. *J. Vac. Sci. Technol. B*, 18(1):361–8, Jan.-Feb. 2000.

[50] W. Vandervorst, T. Clarysse, N. Duhayon, P. Eyben, T. Hantschel, M. Xu, T. Janssens, H. de Witte, T. Conard, J. Deleu, and G. Badenes. Ultra shallow junction profiling. In *Int. El. Dev. Meeting (IEDM) 2000. Techn. Digest*, pages 429–32, San Francisco, CA, 2000.

[51] American Socienty for Testing and Materials. *Standard Practice for Conversion Between Resistivity and Dopant Density for Boron-Doped and Phosphorous-Doped Silicon*. Annual Book of ASTM Standards: F723-88, 1988.

[52] B. Thurber (NIST). Hall effect measurements, http://www.eeel.nist.gov/812/hall.html.

[53] M. V. Ramana Murty. Sputtering: the material erosion tool. *Surface Science*, (500):523–44, 2002.

[54] G. A. Cooke, T. J. Ormsby, M. G. Dowsett, C. Parry, A. Murrell, and E. J. H. Collart. Use of two beam energies in secondary ion mass spectrometry analysis of shallow implants: Resolution-matched profiling. *J. Vac. Sci. Tech. B*, 18(1):493–5, January 2000.

[55] H. D. Witte, W. Vandervorst, and R. Gijbels. Modeling of bombardment induced oxidation of silicon. *J. Appl. Phys.*, 89(5):3001–11, March 2001.

[56] K. Wittmaack, K. H. Eisenrith, W. Hösler, R. von Criegern, and A. Sears. Boron surface contamination on silicon as the ultimate marker for determining dose losses in ultra-shallow depth profiling. In *Proc. USJ-2001*, Napa, CA, 2001.

[57] J. J. Serrano, H. D. Witte, W. Vandervorst, B. Guzmán, and J. M. Blanco. Simulation of the initial transient of the Si^+ and O^+ signals from oxygen sputtered silicon by means of independent models on sputtering and secondary ionization. *J. Appl. Phys.*, 89(9):5191–8, May 2001.

[58] C. M. Frey. *Atomar auflösende Tiefenmikroskopie mit hochenergetischen, schweren Ionen*. PhD thesis, Physik-Department, TU München, D-85747 Garching (Germany), 1997.

[59] A. Bergmaier. *Quantitative Elementanalytik an Diamantschichten mittels elastischer Vorwärtsstreuung (ERD) hochenergetischer schwerer Ionen*. PhD thesis, Physik-Department, TU München, D-85747 Garching (Germany), 1999.

[60] G. Dollinger, C. Frey, A. Bergmaier, and T. Faestermann. Elastic recoil detection with single atomic layer depth resolution. *Nucl. Instr. and Meth. B*, 138:603, 1998.

[61] P. Neumaier, G. Dollinger, A. Bergmaier, I. Genchev, L. Gorgens, R. Fischer, C. Ronning, and H. Hofsass. High-resolution elastic recoil detection utilizing bayesian probability theory. *Nucl. Instr. and Meth. B*, 183(1-2):48–61, July 2001.

[62] J. A. Van den Berg, S. Zhang, D. G. Armour, S. Whelan, D. Goldberg, E. J. H. Collart, P. Bailey, and T. C. Q. Noakes. Medium energy ion scattering for the characterisation of damage profiles of ultra shallow B implants in Si. *Nucl. Instr. and Meth. in Phys. Res. B*, (183):154–165, 2001.

[63] J. A. Van den Berg, D. G. Armour, S. Zhang, S. Whelan, H. Ohno, T.-S. Wang, A. G. Cullis, E. H. J. Collart, R. D. Goldberg, P. Bailey, and T. C. Q. Noakes. Characterization by medium energy ion scattering of damage and dopant profiles produced by ultrashallow B and As implants into Si at different temperatures. *J. Vac. Sci. Tech. B*, 20(3):974–83, May 2002.

[64] D. Briggs and M. P. Seah, editors. *Practical Surface Analysis*, volume 2: Ion and Neutral Spectroscopy, chapter 10, pages 577–613. John Wiley & Sons, New York, 1992.

[65] P. Bailey and S. Richards. MEIS facility at Daresbury Laboratory, http://wserv1.dl.ac.uk/ENG/MEIS/facility/index.htm.

[66] Background on scanning transmission electron microscopy (STEM), http://www.nih.gov/od/ors/dbeps/ssfr/tech.htm.

[67] N. Dietz, D. J. Stephens, G. Lucovsky, and K. J. Bachmann. In-situ multilayer film growth characterization by brewster angle reflectance differential spectroscopy. In *Mat. Res. Soc. Symp. Proc.*, volume 324, pages 27–32, Boston, MA, 1994.

[68] How to make ellipsometry measurements, http://www.jawoollam.com/Tutorial/Tutorial_5.html.

[69] J. W. Colby and L. E. Katz. Boron segregation at Si-SiO_2 interface as a function of temperature and orientation. *J. Electrochem. Soc.*, 123(3):409–12, March 1976.

[70] R. B. Fair and J. C. C. Tsai. Theory and direct measurement of boron segregation in SiO_2 during dry, near dry and wet O_2 oxidation. *J. Electrochem. Soc.*, 125(12):2050–8, December 1978.

[71] G. Charitat and A. Martinez. Boron segregation at Si-SiO_2 interface during neutral anneals. *J. Appl. Phys.*, 55(8):2869–73, April 1984.

[72] P. Küpper. Private communication.

[73] L. Geenen. Private communication.

[74] T. Janssens. Private communication.

[75] W. Vandervorst. Private communication.

[76] N. Cowern and C. Rafferty. Enhanced diffusion in silicon processing. *MRS Bulletin*, 25(6):39–43, June 2000.

[77] H. Bracht. Diffusion mechanisms and intrinsic point-defect properties in silicon. *MRS Bulletin*, 25(6):22–27, June 2000.

[78] H.-J. Gossmann, T. E. Haynes, P. A. Stolk, D. C. Jacobson, G. H. Gilmer, J. M. Poate, H. S. Luftman, T. K. Mogi, and M. O. Thompson. The interstitial fraction of diffusivity of common dopants in Si. *Appl. Phys. Lett.*, 71(26):3862–4, December 1997.

[79] S. C. Jain, W. Schoenmaker, R. Lindsay, P. A. Stolk, S. Decoutere, M. Willander, and H. E. Maes. Transient enhanced diffusion of boron in Si. *J. Appl. Phys.*, 91(11):8919–41, June 2002.

[80] R.-D. Chang, H.-P. Chiang, H.-W. Liu, L.-W. Ho, P.-C. Chiang, J.-R. Tsai, and J. Lin. Observation of transient enhanced diffusion in B^+-implanted Si by buried boron isotopes. *Jpn. J. Appl. Phys.*, 39(11):6136–8, November 2000.

[81] P. A. Stolk, H.-J. Gossmann, D. J. Eaglesham, D. C. Jacobson, C. S. Rafferty, G. H. Gilmer, M. Jaraiz, J. M. Poate, H. S. Luftman, and T. E. Haynes. Physical mechanisms of transient-enhanced dopant diffusion in ion-implanted silicon. *J. Appl. Phys.*, 81(9):6031–50, May 1997.

[82] A. T. Fiory and K. K. Bourdelle. Thermal activation of shallow boron-ion implants. *J. Electron. Mater.*, 28(12):1345–52, December 1999.

[83] G. Mannino, P. A. Stolk, N. E. B. Cowern, W. B. de Boer, A. G. Dirks, F. Roozeboom, J. G. M. van Berkum, P. H. Woerlee, and N. N. Toan. Effect of heating ramp rates on transient enhanced diffusion in ion-implanted silicon. *Appl. Phys. Lett.*, 78(7):889–91, February 2001.

[84] S. T. Dunham, A. H. Gencer, and S. Chakravarthi. Modeling of dopant diffusion in silicon. *IEICE Trans. Electron.*, E82-C(6):800–12, June 1999.

[85] A. Agarwal, H.-J. Gossmann, D. J. Eaglesham, L. Pelaz, D. C. Jacobson, T. E. Haynes, and Y. E. Erokhin. Reduction of transient diffusion from 1–5 keV Si^+ ion implantation due to surface annihilation of interstitials. *Appl. Phys. Lett.*, 71(21):3141, November 1997.

[86] A. Agarwal, H.-J. Gossmann, D. J. Eaglesham, S. B. Herner, A. T. Fiory, and T. E. Haynes. Boron-enhanced diffusion of boron from ultralow-energy ion implantation. *Appl. Phys. Lett.*, 74(17):2435–37, April 1999.

[87] H.-H. Vuong, H.-J. Gossmann, L. Pelaz, G. K. Celler, D. C. Jacobson, D. Barr, J. Hergenrother, D. Monroe, V. C. Venezia, C. S. Rafferty, S. J. Hillenius, J. McKinley, F. A. A. Stevie, and C. Granger. Boron pileup and clustering in silicon-on-insulator films. *Appl. Phys. Lett.*, 75(8):1083–5, August 1999.

[88] H. C.-H. Wang, C.-C. Wang, C.-S. Chang, T. Wang, P. B. Griffin, and C. H. Diaz. Interface induced uphill diffusion of boron: An effective approach for ultrashallow junction. *IEEE El. Dev. Lett.*, 22(2):65–7, February 2001.

[89] G. A. Sai-Halasz, K. T. Short, and J. S. Williams. Antimony and arsenic segregation at Si-SiO_2 interfaces. *IEEE El. Dev. Lett.*, 6:285, June 1985.

[90] F. Lau, L. Mader, C. Mazure, C. Werner, and M. Orlowski. A model for phosphorus segregation at the silicon-silicon dioxide interface. *Appl. Phys. A*, 49:671–5, 1989.

[91] R. Baierle, M. J. Caldas, J. Dabrowski, H.-J. Müssig, and V. Zavodinsky. A unified microscopic mechanism for donor deactivation in Si. *Physica B*, 273-274:260–3, December 1999.

[92] J. Dabrowski, R. A. Casali, H.-J. Mussig, R. Baierle, M. J. Caldas, and V. Zavodinsky. Mechanism of dopant segregation to SiO_2/Si(001) interfaces. *J. Vac. Sci. Tech. B*, 18(4):2160–4, July 2000.

[93] J. Dabrowski, H.-J. Mussig, V. Zavodinsky, R. Baierle, and M. J. Caldas. Mechanism of dopant segregation to SiO_2/Si(001) interfaces. *Phys. Rev. B*, 65(24):245305, June 2002.

[94] H.-H. Vuong, C. S. Rafferty, S. A. Eshraghi, J. L. Lentz, P. M. Zeitzoff, M. R. Pinto, and S. J. Hillenius. Effects of oxide interface traps and transient enhanced diffusion on the process modeling of pMOS devices. *IEEE Tran. Elec. Devices*, 43(7):1144–52, July 1996.

[95] H.-H. Vuong, C. S. Rafferty, S. A. Eshraghi, J. Ning, J. R. McMacken, S. Chaudhury, J. McKinley, and F. A. Stevie. Dopant dose loss at the Si-SiO_2 interface. *J. Vac. Sci. Tech. B*, 18(1):428–34, January 2000.

[96] A. Shima, T. Jinbo, J. Ushio, J.-H. Oh, K. Ono, M. Oshima, and N. Natsuaki. Investigation of a model for the segregation and pile-up of boron at the SiO_2/Si interface during the formation of ultra-shallow p^+ junctions. In *Int. El. Dev. Meeting (IEDM) 2000. Techn. Digest*, pages 519–22, San Francisco, CA,, 2000.

[97] A. Shima, T. Jinbo, N. Natsuaki, J. Ushio, J.-H. Oh, K. Ono, and M. Oshima. A model for the segregation and pileup of boron at the SiO_2/Si interface during the formation of ultrashallow p^+ junctions. *J. Appl. Phys.*, 89(6):3458–63, March 2001.

[98] N. J. Chou, Y. J. van der Meulen, R. Hammer, and J. Cahill. Auger and ellipsometric study of phosphorus segregation in oxidized degenerate silicon. *Appl. Phys. Lett.*, 24(4):200–2, February 1974.

[99] S. A. Schwarz, C. R. Helms, W. E. Spicer, and N. J. Taylor. High resolution auger sputter profiling study of the effect of phosphorus pileup on the Si–SiO_2 interface morphology. *J. Vac. Sci. Tech.*, 15(2):227–30, March 1978.

[100] S. A. Schwarz, R. W. Barton, C. P. Ho, and C. R. Helms. Studies of phosphorus pile-up at the Si-SiO_2 interface using auger sputter profiling. *J. Electrochem. Soc.*, 128(5):1101, May 1981.

[101] Y. Sato, M. Watanabe, and K. Imai. Characterization of phosphorus pile-up at the SiO_2/Si interface. *J. electrochem. Soc.*, 140(9):2679–82, September 1993.

[102] Y. Sato, K. Imai, H. Yomezawa, and T. Shigematsu. Detection of phosphorus pileup at SiO_2/Si interface. *J. electrochem. Soc.*, 140(12):L176–7, December 1993.

[103] Y. Sato, J. Nakata, K. Imai, and E. Arai. Arsenic pileup at the SiO_2/Si interface. *J. electrochem. Soc.*, 142(2):655–60, February 1995.

[104] P. B. Griffin, S. W. Crowder, and J. M. Knight. Dose loss in phosphorus implants due to transient diffusion and interface segregation. *Appl. Phys. Lett.*, 67(4):482–4, July 1995.

[105] R. Kasnavi, P. Pianettam, Y. Sun, R. Mo, P. B. Griffin, and J. D. Plummer. Characterization of arsenic dose loss at the Si/SiO_2 interface using high resolution X-ray photoelectron spectrometry. In *Int. El. Dev. Meeting (IEDM) 1998. Techn. Digest*, pages 721–4, San Francisco, CA, 1998.

[106] R. Kasnavi, Y. Sun, R. Mo, P. Pianetta, P. B. Griffin, and J. D. Plummer. Characterization of arsenic dose loss at the Si/SiO_2 interface. *J. Appl. Phys.*, 87(5):2255–60, March 2000.

[107] M. Koh, K. Egusa, H. Furumoto, T. Shirakata, E. Seo, K. Shibahara, S. Yokoyama, and M. Hirose. Quantitative evaluation of dopant loss in 5-10 keV As ion implantation for low-resistive, ultrashallow source/drain formation. *Jpn. J. Appl. Phys.*, 38(1/4B):2324–8, April 1999.

[108] S. W. Crowder, P. M. Rousseau, J. P. Snyder, J. A. Scott, P. B. Griffin, and J. D. Plummer. The effect of source/drain processing on the reverse short channel effect of deep sub-micron bulk and SOI NMOSFETs. In *Int. El. Dev. Meeting (IEDM) 1995. Techn. Digest*, pages 427–30, Washington, DC, USA, 1993.

[109] T. Topuria, E. M. James, N. D. Browning, and Z. Ma. Direct atomic scale characterization of interfaces and doping layers in field-effect transistors. *Appl. Phys. Lett.*, 79(1):132–134, July 2001.

[110] T. Topuria, N. D. Browning, and Z. Ma. Dopant segregation in field effect transistors characterized by scanning transmission electron microscopy. In *AVS 2002 Conf.*, 2002.

[111] W. Vandervorst, T. Janssens, B. Brijs, R. Lindsay, E. Collart, D. Kirkwood, G. Mathot, and G. Terwagne. Profile changes and self-sputtering during low energy ion implantation. In *Mat. Res. Soc. Symp. Proc.*, volume 717, page C7.2, San Francisco, CA, 2002.

[112] R. B. Fair. Anomalous B penetration through ultrathin gate oxides during rapid thermal annealing. *IEEE El. Dev. Lett.*, 20(9):466–9, September 1999.

[113] S. Mukherjee and P. Evans. Studies on diffusion of boron through silicon oxide films. *Thin Solid Films*, (14):299, 1972.

[114] T. Aoyama, K. Suzuki, H. Tashiro, Y. Toda, T. Yamazaki, Y. Arimoto, and T. Ito. Boron diffusion through pure silicon oxide and oxynitride for metal-oxide-semiconductor devices. *J. Electrochem. Soc.*, 140(12):3624–7, December 1993.

[115] M. Navi and S. T. Dunham. Investigation of boron penetration through thin gate dielectrics including role of nitrogen and fluorine. *J. Electrochem. Soc.*, 145(7):2545–8, July 1998.

[116] A. T. Fiory. Thin SiO_2 films grown for brief oxidation times. *J. Electron. Mater.*, 28(12):1358–64, December 1999.

[117] F. Koch. Private communication.

[118] ISU College of Engineering. Semiconductor processing and fabrication, http://mosfet.isu.edu/classes/Semiconductor/ProcClass7Student.doc.

[119] MICROTECH Knowledge Base. Wet etching of silicon dioxide, http://www.microtechweb.com/kb/sio_etch.htm.

[120] J. Bühler, F.-P. Steiner, and H. Baltes. Silicon dioxide sacrificial layer etching in surface micromachining. *J. Micromech. Microeng.*, 7(1):R1–13, 1997.

[121] F. Iacona, V. Raineri, F. La Via, A. Terrasi, and E. Rimini. Arsenic redistribution at the SiO_2/Si interface during oxidation of implanted silicon. *Phys. Rev. B*, (58):10990–9, 1998.

[122] T. Dittrich, H. Angermann, W. Füssel, and H. Flietner. Electronic properties of the HF-passivated Si(111) surface during the initial oxidation in air. *Phys. Stat. Sol. (a)*, 140:463–70, 1993.

[123] T. Dittrich. Private communication.

[124] B. Sadigh, T. J. Lenosky, M.-J. Caturla, A. A. Quong, L. X. Benedict, T. Diaz de la Rubia, M. M. Giles, M. Foad, C. D. Spataru, and S. G. Louie. Large enhancement of boron solubility in silicon due to biaxial stress. *Appl. Phys. Lett.*, 80(25):4738–40, June 2002.

Index

A

Anomalous diffusion 54

B

Back-End-Of-Line (BEOL) 7
BED 55
Buffered HF (BHF) 109

C

Carrier spilling 25
Channelling 13

D

Decoupled Plasma Nitridation (DPN) ... 83
Diffusion 54, 102
Dopant loss 63
DPN 83

E

EDX 39, 51
EELS 39
EFTEM 39
Ellipsometry 40
ERD 34, 48
Extension 5, 9

F

Flash annealing 18
Four Point Probe (FPP) 24, 107
Front-End-Of-Line (FEOL) 3

H

Hall effect 29
HDD 6
Heavily nitrided oxide 83
HF dip 98, 109, 111

I

In-situ deposition 21
Ion implantation (I/I) 12
ITRS 1, 9

L

Laser Annealing (LTA) 18

M

MEIS 37
Metallization 7
Moore, Gordon 1

N

native oxide growth 122
NO 83

O

OED 55
Outdiffusion 64
Oxidation 88
Oxide (pure) 82
Oxynitride (NO) 83

P

Plasma Doping (PLAD) 14
Probe junction 25
Probe Spacing (PS) 29

R

RBS 36, 48
RTP 14

S

Sample preparation 61
Segregation coefficient 45
Self-Sputtering 63

Shallow Trench Isolation (STI) 5
Sheet resistance 23
Silicide 7
SIMS 31, 49
Spacer 6
SPER 19, 103
SRP 27, 132

T

Technology node 1
TED 55
TEM 37, 51, 111

U

Uphill diffusion 56
USJ 9

V

Van-der-Pauw structure 26