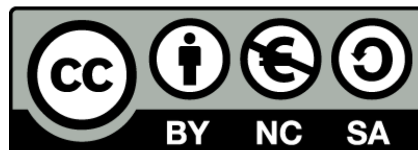
TEM techniques for the development of ULSI technology

Alberto Romano Rodríguez



Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement- NoComercial – CompartirIgual 4.0. Espanya de Creative Commons***.

Esta tesis doctoral está sujeta a la licencia ***Reconocimiento - NoComercial – CompartirIgual 4.0. España de Creative*** **Commons.**

This doctoral thesis is licensed under the ***Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Spain License.***

UNIVERSIDAD DE BARCELONA

FACULTAD DE FISICA

DEPARTAMENTO DE FISICA APLICADA Y ELECTRONICA

TEM TECHNIQUES FOR THE DEVELOPMENT OF ULSI TECHNOLOGY

ROMANO RODRÍGUEZ, A.

Memoria presentada por Alberto Romano-Rodríguez para optar al grado de Doctor en Ciencias Físicas

UNIVERSIDAD DE BARCELONA

FACULTAD DE FISICA

DEPARTAMENTO DE FISICA APLICADA Y ELECTRONICA

TEM TECHNIQUES FOR THE DEVELOPMENT OF ULSI TECHNOLOGY

Memoria presentada por
Alberto Romano Rodríguez
para optar al grado de
Doctor en Ciencias Físicas

Don Joan Ramon Morante i Lleonart y don Albert Cornet i Calveras, Catedráticos de Electrónica de la Facultad de Física de la Universidad de Barcelona

CERTIFICAN que la memoria "TEM TECHNIQUES FOR THE DEVELOPMENT OF ULSI TECHNOLOGY", que presenta Alberto Romano Rodríguez para optar al grado de Doctor en Ciencias Físicas, ha sido realizada bajo su dirección.

Barcelona, Febrero de 1991

Joan Ramon Morante i Lleonart

Albert Cornet i Calveras

A la Silvia

El presente trabajo se enmarca dentro del programa de investigación sobre caracterización de semiconductores y dispositivos electrónicos que se desarrolla en la Cátedra de Electrónica de la Facultad de Física de la Universidad de Barcelona. La colaboración entre la Cátedra, el Interuniversitair Microelectronika Centrum (IMEC), Leuven (Bélgica), y el Rijksuniversitair Centrum Antwerpen (RUCA), Antwerpen (Bélgica), y la beca de formación de postgrado en el extranjero, otorgada por el Ministerio de Educación y Ciencia, han permitido la realización de la mayor parte de este trabajo en los dos laboratorios del extranjero.

Primeramente quiero agradecer a mis directores, doctores Joan Ramon Morante i Lleonart y Albert Cornet i Calveras, Catedráticos de Electrónica de la Universidad de Barcelona, por haberme aceptado en su grupo de investigación, haberme ofrecido la oportunidad de realizar este trabajo y, por tanto, de profundizar en la aplicación de la microscopía electrónica de transmisión a materiales semiconductores.

I would like to thank especially Dr. Jan Vanhellemont (IMEC) for his day to day following of the work, for fruitful discussions about defects in silicon and for his continuous encouragement.

Prof. Dr. ir. Herman Maes is acknowledged for offering me the possibility to work in the analytical group (ARS) of IMEC. Prof. Dr. Jef Van Landuyt (RUCA) is also acknowledged for giving me access to the electron microscopes of his laboratory.

Lic. Ann de Keersgieter and Dr. Hans Norström (now at the Swedish Institute of Microelectronics, Kista, Sweden) kindly supplied most of the experimental material used in this work and are also thanked for many stimulating discussions. The expertise of Dr. Ingrid de Wolf in Raman microprobe spectroscopy has been of great use. Prof. Dr. ir. Wilfried Vandervorst provided the SIMS and SRP measurements for correlation with TEM. The critical reading of the manuscript by Dr. Hugo Bender is especially appreciated. Patricia Van Marcke helped in the development and in the application of the specimen preparation techniques.

The main part of the TEM observations have been carried out in RUCA. The careful preparation of the photographic material is due to Freddy Schallenberg and André de Munck. Without them this work would look quite different. Roger Van Ginderen and Jan Eysermans had the responsibility of keeping the microscopes in good working conditions, with big success.

I also would like to acknowledge all the colleagues from the three laboratories that helped me by creating a positive atmosphere and all the people that, direct or indirectly, collaborated in the realization of this work.

Quiero expresar mi más sincero agradecimiento a mis padres, por el continuo apoyo y los valiosos comentarios recibidos a lo largo de este trabajo.

Finalment haig d'agraïr especialment a la Sílvia el seu suport incondicional i l'entusiasme manifestat sempre per la realització d'aquest treball.

Leuven (Belgium), February 1991.

INDEX

0. **INTRODUCTION** i

1. **EXPERIMENTAL TECHNIQUES** I.1
1.0 INTRODUCTION I.1
1.1 TRANSMISSION ELECTRON MICROSCOPY I.2
1.1.0 General remarks I.2
1.1.1 Two beam imaging I.5
1.1.1.1 Stacking faults I.11
1.1.1.2 Dislocations I.13
1.1.1.3 Dislocation loops I.13
1.1.1.4 Examples I.14
1.1.2 High resolution electron microscopy I.19
1.2 SECONDARY ION MASS SPECTROMETRY I.24
1.3 SPREADING RESISTANCE PROBE I.27
REFERENCES CHAPTER 1 I.31

2. **SPECIMEN PREPARATION TECHNIQUES** II.1
2.0 INTRODUCTION II.1
2.1 CROSS-SECTION SPECIMENS II.3
2.1.1 Sawing of the wafer II.4
2.1.2 Glueing of the stripes II.6
2.1.3 Flat grinding II.6
2.1.4 Dimpling II.9
2.1.5 Ion milling II.11
2.2 PLAN VIEW SPECIMENS II.12
2.2.1 Cutting a disc II.16
2.2.2 Mechanical thinning II.16
2.2.3 Final thinning II.17
2.2.3.1 Chemical thinning II.17
2.2.3.2 Ion milling II.18
2.3 LOCAL THINNING II.20
2.3.1 Cross-section preparation II.20
2.3.1.1 Sawing of the wafer II.21
2.3.1.2 Glueing of the stripes II.21
2.3.1.3 Mechanical prethinning and dimpling II.23
2.3.1.4 Ion milling II.26
2.3.1.5 Finished specimen II.29

2.3.2 Plan view preparation II.29
2.4 OTHER SEMICONDUCTOR MATERIALS II.32
2.5 CONCLUSIONS II.34
REFERENCES CHAPTER 2 II.36

3. ISOLATION TECHNIQUES III.1
3.0 INTRODUCTION III.1
3.1 EXPERIMENTAL DETAILS III.2
3.2 LOCAL OXIDATION OF SILICON III.3
3.2.1 Principles of the process III.3
3.2.2 Geometry effects III.6
3.2.3 Gate oxide thinning III.7
3.2.4 Drawbacks and alternatives to the LOCOS process III.7
3.3 SEALED INTERFACE LOCAL OXIDATION III.10
3.3.0 General remarks III.10
3.3.1 Nitridation of the pad oxide III.11
3.3.1.1 Experimental procedure III.11
3.3.1.2 Results III.12
3.3.1.3 Fully processed transistors III.12
3.3.1.4 Asymmetry in closely packed structures III.14
3.3.2 Nitride layer in direct contact with the silicon substrate III.17
3.3.2.1 Processing conditions III.17
3.3.2.2 TEM observations III.19
3.4 LOCAL OXIDATION OF POLYSILICON OVER SILICON III.19
3.4.0 Introduction III.19
3.4.1 LOPOS with polycrystalline silicon III.20
3.4.2 LOPOS with microcrystalline silicon III.20
3.4.2.1 Experimental conditions III.21
3.4.2.2 Geometrical aspects III.23
3.4.2.3 Stresses in LOPOS structures III.26
3.4.2.4 Field oxide thinning III.33
3.5 LOCOS ON SIMOX SUBSTRATES III.35
3.5.0 Introduction III.35
3.5.1 Experimental conditions III.37
3.5.2 TEM observations III.37
3.6 CONCLUSION III.40
REFERENCES CHAPTER 3 III.42

4. JUNCTION DELINEATION IV.1
4.0 INTRODUCTION IV.1
4.1 STARTING MATERIAL IV.4
4.2 TEM COMBINED WITH PREFERENTIAL ETCHING IV.5
4.2.1 HF-HNO_3 system IV.7
4.2.1.1 Etching behaviour IV.7
4.2.1.2 Delineation with this system IV.10
4.2.1.3 Limitations of this solution IV.14
4.2.2 CrO_3-HF system IV.16
4.2.2.1 Etching behaviour IV.16
4.2.2.2 Delineation experiments IV.17
4.2.3 Limitations of the chemical etching method IV.22
4.3 IN-SITU ELECTRON IRRADIATION IN THE HVEM IV.22
4.3.0 Introduction IV.22
4.3.1 Observations IV.23
4.3.1.1 boron doped silicon IV.24
4.3.1.2 Phosphorus IV.32
4.3.1.3 Arsenic IV.34
4.3.2 Discussion IV.37
4.4 CONCLUSIONS IV.40
REFERENCES CHAPTER 4 IV.44

5. GENERAL CONCLUSIONS AND FURTHER RESEARCH V.1

0. INTRODUCTION

In the last decade transmission electron microscopy has become one of the most powerful tools for the structural characterization of crystalline materials. Especially in the field of microelectronics development and production, the information obtained by this analytical tool has been of vital importance for the improvement of the different processing steps. With the reduction of the dimensions of the devices TEM is the only technique that can give information with enough spatial resolution.

In this work applications of transmission electron microscopy (TEM) for the study of processing steps of ULSI technology will be presented. In the first chapter special emphasize is given to the theoretical background of transmission electron microscopy, needed for the interpretation of the obtained results. Some details about Secondary Ion Mass Spectrometry (SIMS) and Spreading Resistance Probe (SRP) are also given, as they will be used for cross-correlation in the following chapters.

In semiconducting research and production the processes for the fabrication of devices are restricted to a small volume of the starting material. Only the areas close to the surface, about 5 μm deep, are used for the operation of the devices. Both for plan view and cross-section TEM investigation of integrated circuits special specimen preparation techniques had to be developed. These techniques are described in full detail in chapter two.

Chapters 3 and 4 deal with the application of TEM for the characterization of two processing steps that are required for the fabrication of ULSI devices. Chapter 3 presents a structural study of the geometry and the defect generation in the substrate for one of the most important aspects of the process: the device isolation. In it modifications of the LOCOS process will be given, which try to solve the specific problems of LOCOS when applied to submicron devices.

In chapter 4 the delineation of shallow junctions is treated. The dopant profiles have to be well characterized and analytical techniques have to be used which have a high spatial resolution and a high sensitivity for dopant concentration. Especially the determination of two-dimensional dopant profiles poses severe problems. Special TEM techniques will be presented, which allow to determine the lateral dopant spread below implantation masks. Especially promising is the application of in-situ electron irradiation in a high voltage transmission electron microscope (HVEM) to this topic. It will be shown that the formation of extended defects in preferential areas of the specimens, induced by the high energy electrons, is dependent on the dopant concentration. Correlation with SIMS or SRP will allow to determine this doping level in a quite accurate way.

Finally in chapter 5 the most relevant conclusions of this work will be presented. Further research on the treated topics will be suggested.

1. EXPERIMENTAL TECHNIQUES

1.0 INTRODUCTION

In this chapter the different experimental techniques that are applied throughout this work will be presented. The main technique used is the transmission electron microscopy (TEM). The other techniques that will be used are employed only for comparison or confirmation of the observed phenomena in the microscope. The theoretical background and formulae necessary for the interpretation of the micrographs will be discussed. This is presented in section 1.1, where the two main modes of image formation in the TEM are revised, i.e, diffraction contrast and phase contrast imaging. Examples will be given that show how to characterise defects in the electron microscope. For this purpose material which will be discussed more in detail in chapters 3 and 4 will be used.

In sections 1.2 and 1.3 Secondary Ion Mass Spectrometry (SIMS) and Spreading Resistance Probe (SRP) will be briefly described, respectively. These techniques will be used for comparison or confirmation of the observed behaviours in the electron microscope. One dimensional SIMS or SRP are used for calibration of the different delineation techniques. The delineated depth is compared with one of these two techniques and the corresponding level is assigned. For this reason it is interesting to have a basic knowledge of them.

In addition to these techniques some results from Raman microprobe spectroscopy and Scanning Electron Microscopy (SEM) will be employed in punctual cases and the information supplied by them will be discussed when necessary.

In chapter 4 some simulations of dopant profiles will be given, which have been obtained using commercial simulation programs, like SUPREM, SUPRA, ICECREM and PREDICT. They are compared with SIMS or SRP when these data are available.

1.1 TRANSMISSION ELECTRON MICROSCOPY

1.1.0 General remarks

Transmission electron microscopy is an ideal tool for the characterization and identification of lattice defects in crystalline materials. Its construction is analogous to a transmission optical microscope and both are schematically represented in figure 1.1. The optical microscope makes use of visible light and glass lenses, while the electron microscope works with an electron beam and electromagnetic lenses. Due to the high absorption of electrons in air it is necessary to accelerate them in vacuum and, for this reason, the microscope column has to be a vacuum tube. The source of electrons is a filament, commonly made out of tungsten or LaB_6, that emits electrons by thermal emission. These electrons are emitted in all directions and by using a wehnelt cathode they are oriented towards the column and are next accelerated by the voltage difference between the cathode and the anode, typically in the order of 100 to 400 kV, although some instruments work with acceleration voltages up to 3 MV. The illumination system is located below the anode and is composed by two (or, sometimes, three) condensor lenses and two apertures, that limit the angular spread of the electrons to some mrad. After the electrons cross the specimen the imaging system forms an image on a fluorescent screen, photographic plate or a TV camera. The imaging system is formed by the objective lens, the diffraction lens and the intermediate and projector lenses.

Depending on the settings of the lenses of the imaging system, two imaging modes will be obtained. The first is a direct image of the specimen, while the second gives an image of the back focal plane of the objective lens, which, at its time, is the reciprocal lattice plane. This imaging mode is called diffraction mode.

The diffraction pattern is very useful for obtaining information on the type of the material subjected to observation. If the material is amorphous, diffuse rings will be obtained, due to the random scattering of the electrons by the atoms of the

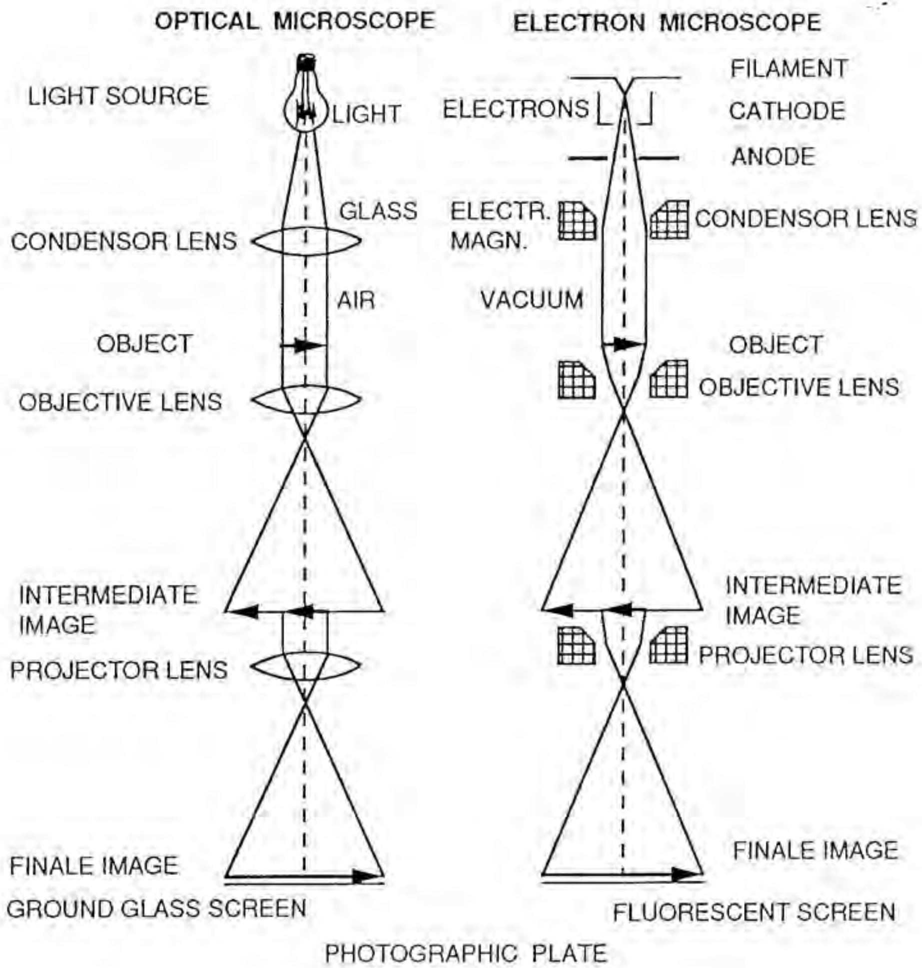


Figure 1.1: Schematic representation of the transmission optical microscope and the transmission electron microscope, showing their correspondences.

specimen. Two types of crystalline materials can be distinguished. If it is monocrystalline, spot patterns will be obtained, of which the aspect will depend on the orientation of the crystal related to the incident beam. For a polycrystalline specimen a ring pattern will be observed, the rings being very sharp. This is due to the random orientation of the different crystallites, each giving rise to a point diffraction pattern. A superposition of the diffraction patterns arising from each crystal will form concentric rings, whose distances to the centre of the pattern being equal to the distances of the corresponding spots in the single crystal diagram.

For a monocrystalline material, by appropriate orientation of the real and diffracted images, the crystallographic directions in the image can be deduced. Furthermore the diffraction pattern of such a material can reveal the presence of second phases in the matrix by the appearance of extra spots, which give information on the nature of the inclusions. Some defects, like twins, also give rise to additional spots if their size is large enough.

This shows that TEM is specially suited for the structural investigation of crystalline materials in very small areas. Some information can also be obtained from amorphous materials. It is possible to distinguish between two different amorphous layers, because their contrast is dependent on the atomic number of their elements, as the scattering and absorption of the electrons is proportional to Z. This explains why two different layers, for example silicon oxide and silicon nitride, can be distinguished (see chapter 3). The same dependence is valid for monocrystalline materials, although the diffraction conditions play, generally, a more important role.

In the next sections the two main modes of contrast formation employed in this work will be discussed. They are diffraction contrast and phase contrast. For the first the contrast is strongly influenced by the orientation of the specimen related to the electron beam, while for the latter the contrast is formed by the interference of several diffraction beams.

1.1.1 Two beam imaging

For the interpretation of the diffraction contrast, the most simple situation is the so-called two-beam condition. In that case the specimen is oriented such that only the transmitted and one diffracted beam are strongly excited, which can be seen in the diffraction diagram. Inserting an aperture in the back focal plane, the objective or contrast aperture, one of the beams can be selected. If the central beam is selected, the so-called bright field image (BF) is formed. In this mode a hole in the specimen (where no diffraction takes place and the beam passes unscattered) will appear clear. If instead of selecting the transmitted beam only the strongly diffracted one is allowed to pass through the aperture, the image will be formed in dark field mode (DF). As the beam that passes through the hole is not scattered, no contribution of these electrons will take place and it will appear dark on the screen. Finally it is possible to select another beam that is none of these two in order to form the image. Typically this beam is of the same family as the strongly diffracted one and the image so formed is called weak beam image (WB). This mode is very useful for the observation and identification of small defects present in the specimen.

For a perfect crystal without lattice defects, the only contrast will result from thickness variations of the specimen or from bending, because the specimen is very thin. But if a defect is present in the crystal, its strain field will give rise to a change in the orientation of the lattice planes in its neighbourhood. As a consequence these areas will diffract the incident beam differently and their image will be darker or lighter than the one of the undistorted crystal. In the dark field mode the defects are light against the surrounding dark background, giving a better contrast than for the bright field mode. The attainable resolution is of the order of 10 to 20 nm. By selecting a weak beam the resolution can be improved to about 2.5 to 5 nm.

The microscope employed in this work for these type of imaging is, mainly, a high voltage transmission electron micro-

scope JEOL 1250 (maximum 1250 kV) operated at 1000 kV. The specimens are mounted in a rotation-tilt holder. The main advantages of this microscope, compared to a conventional 100 or 200 kV, is that the penetration depth of the electrons is much higher. At 1000 kV easily through 5 µm of silicon can be observed with a good contrast, while for 100 kV this thickness is only about 0.5 µm. As a consequence, the transparent area per specimen is increased by a factor of 100. This is particularly interesting to study materials with low defect densities. As the defects can be studied in thicker areas, also a better insight will be gained on the bulk behaviour and the defect interaction and bulk defects can be observed simultaneously.

However this microscope has also some drawbacks. At that high voltage the efficiency of the fluorescent screen is less than for 100 kV electrons. A second disadvantage is that for the higher energies the contrast of the image decreases substantially. In addition to these points it is necessary to protect the user against X-ray radiation, which is achieved by a thick lead glass in the observation window, which also substantially lowers the intensity that reaches the eye for observation. As a consequence it is difficult to work at high magnifications, at which the settings of the exact focus and of the objective astigmatism are critical. The magnification range commonly employed lies between 3000x and 50000x, but in certain cases 100000x is possible to obtain.

Another effect for which care has to be taken is the production of radiation damage by the electron beam. This damage can be due to two different mechanisms: heavy ions accelerated by the high voltage and that create a cascade in the silicon lattice [1.1] or the electrons themselves, because their energy is above the displacement threshold to create Frenkel pairs (see, for example, [1.2]). In chapter 4 this effect will be used to study the influence of the dopant concentration on the creation of radiation induced defects. Anyway, when studies on extended defects have to be carried out this radiation damage does not disturb the image and can be easily distinguished. However these specimens can be subsequently studied by high resolution electron microscopy, in which case one has to be very careful to

decide if a certain defect is due to a processing step or has been generated by the radiation. The extended nature of the defects in high resolution specimens is lost, as only very thin sections can be observed in this last mode.

As the point resolution of the high voltage electron microscope in the side entry mode is of the order of 0.7 nm, no lattice image of the silicon can be obtained. Therefore the microscope is only used for two-beam imaging and for the observation of small precipitates in the symmetrical beam orientation, without the aim of lattice resolution. The symmetrical orientation is also used for geometry studies in which, generally, a [110] section is observed, which is perpendicular to the wafer surface and to the dominant direction of structures used in ULSI processing.

The theoretical background for the interpretation of the images of defects has been developed about 20 years ago. A large number of reference books are available, e.g., the ones by Amelinckx [1.3], Heidenreich [1.4] and Hirsch et al. [1.5]. In the following only a brief synthesis is given. The terminology which is used is the one by Hirsch et al. [1.5].

The steady state function $\Psi(\bar{r})$ of an electron with energy eE can be obtained from the Schrödinger equation:

$$\nabla^2 \Psi(\bar{r}) + \frac{8\pi^2 me}{h^2} \left[E + V(\bar{r}) \right] \Psi(\bar{r}) = 0 \tag{1.1}$$

where $V(\bar{r})$ is the lattice potential. The electron mass, m, and the energy, E, are related to the rest mass, m_o, and accelerating potential, E_o, by, respectively:

$$m = \frac{m_o}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{1.2}$$

$$E = E_o \frac{1 + \frac{eE_o}{2m_o c^2}}{1 + \frac{eE_o}{m_o c^2}} \tag{1.3}$$

The relativistic corrections become important only if the accelerating voltage is above 100 kV, but they do not affect the general expressions that will describe the defect contrast and the conditions to obtain it.

For an undistorted crystal the lattice potential can be expressed as a Fourier expansion, i.e.:

$$V(\bar{r}) = \frac{h^2}{2me} \sum_g U_g \exp(2\pi i \, \bar{g}\bar{r}) \tag{1.4}$$

where $\bar{g}$ are the different reciprocal lattice vectors and, consequently, the different spots in the diffraction diagram. Under these conditions the solution to the wave function $\Psi(\bar{r})$ can be written, according to the Bloch theorem, in the form:

$$\Psi(\bar{r}) = \sum_g C_g \exp\left[2\pi i \, (\bar{k} + \bar{g})\bar{r}\right] \tag{1.5}$$

Replacing $\Psi(\bar{r})$ and $V(\bar{r})$ in equation (1.1) by their values from equations (1.4) and (1.5), gives rise to the basic equations describing the dynamical theory in its general case, i.e., considering all diffraction vectors:

$$\left[\bar{K}^2 - (\bar{k} + \bar{g})^2\right] C_g + \sum_{g'} U_{g'} C_{g-g'} = 0 \tag{1.6}$$

In this equation the summation is carried out for all reciprocal lattice vectors except for $\bar{g}' = 0$. Furthermore:

$$\bar{K}^2 = U_o + \bar{\chi}^2 \tag{1.7}$$

where $\bar{\chi}$ is the incident wave vector, i.e., $|\bar{\chi}| = 1/\lambda$, λ being the electron wavelength.

Equation (1.6) is quite complicated if one takes into account all the reciprocal lattice vectors. But if only two of them are strongly excited, the contribution of all the other terms is very small and can be neglected. In this way the equation system of (1.6) reduces to only two equations:

$$(\bar{K}^2 - \bar{k}^2)\ C_o + U_{-g}\ C_g = 0$$
$$U_g\ C_o + \left[\bar{K}^2 - (\bar{k} + \bar{g})^2\right]\ C_g = 0 \qquad (1.7)$$

The solution of these equations for a perfect crystal yields the so called rocking curve, which explains the bending and extinction contours.

We will assume now that the crystal we are dealing with is no longer a perfect crystal, but that lattice defects are present. In this case the potential will no longer be periodical. But it will be still possible to express the potential as a Fourier expansion, where the coefficients for the different $\bar{g}$ vectors vary with the position. This can be written in the following form:

$$V(\bar{r}) = \frac{h^2}{2me} \sum_g u_g(\bar{r})\ \exp\ (2\pi i\ \bar{g}\bar{r}) \qquad (1.8)$$

Similarly the wave functions can be expressed as:

$$\Psi(\bar{r}) = \sum_g \Phi_g(\bar{r})\ \exp\ \left[2\pi i\ (\bar{\chi} + \bar{g} + \bar{s}_g)\bar{r}\right] \qquad (1.9)$$

In this formula $\bar{s}_g$ represents the deviation from the Bragg position taken only in the z direction. Substituting formulae (1.8) and (1.9) in the Schrödinger equation and assuming that the coefficient of each exponential can be set to zero, a new set of equations is obtained:

$$(\bar{\chi} + \bar{g} + \bar{s}_g)_z \frac{\partial \Phi_g}{\partial z} = \sum_{g'} \pi i\ u_{g-g'}(\bar{r})\ \Phi_{g'}(\bar{r})\ \exp\ \left[2\pi i\ (s_{g'} - s_g) z\right]$$
$$- (\bar{\chi} + \bar{g})_x \frac{\partial \Phi_g}{\partial x} - (\bar{\chi} + \bar{g})_y \frac{\partial \Phi_g}{\partial y} + \frac{i}{4\pi} \nabla^2 \Phi_g \qquad (1.10)$$

In our case the beam is incident along the z axis or very close to it, so that the x and y components of $\bar{\chi}$ are much smaller than the z component. In addition Φ_g must be varying sufficiently slow. Under these assumptions the last three terms

of equation (1.10) can be neglected. This approach is called the column approximation, for which the equations can be solved in each individual column (x,y) independently of the neighbouring columns. This approximation is true for relatively thin crystals, where the scattering is quite small.

When a defect is present in the specimen it can be described by a vector $\bar{R}(\bar{r})$, which gives the displacement of the surrounding lattice due to the defect. It is assumed that this value varies slowly, and then the deformable ion approximation can be introduced. The potential in the imperfect crystal at a certain position $\bar{r}$ can be replaced by $\bar{r} - \bar{R}(\bar{r})$. Introducing this value in formula (1.8) and comparing with (1.4) it can be deduced that the potential coefficients in the Fourier expansion are now:

$$u_g(\bar{r}) = U_g \exp\left[-2\pi i\ \bar{g}\ \bar{R}(\bar{r})\right] \qquad (1.11)$$

If the small difference between the z projection of $(\bar{\chi} + \bar{g} + \bar{s}_g)$ and the projection in the z direction of $\bar{K}$, called K_z, is neglected, a parameter called the extinction distance, ξ_g, can be defined as follows:

$$\xi_g = \frac{K_z}{|U_g|} \qquad (1.12)$$

This extinction distance corresponds, in the above mentioned approximation, to the thickness difference between two points that give rise to the same contrast.

Formula (1.10) can therefore be expressed as:

$$\frac{d\Phi_g}{dz} = \sum_{g'} \frac{\pi i}{\xi_{g-g'}} \Phi_{g'} \exp\left[2\pi i\ (s_{g'} - s_g) z + 2\pi i\ (\bar{g}' - \bar{g})\bar{R}\right]$$

$$(1.13)$$

This is the general expression of the contrast formation for an imperfect crystal, where the defect is represented by the displacement $\bar{R}(\bar{r})$. In this approximation it has been assumed that there is no absorption of the electrons by the specimen, which is only the case for very thin specimens. If one wants to

account for the absorption in the general formula, it can be done by introducing a complex potential, which is equivalent to using a complex extinction distance.

As mentioned above, in order to simplify the analysis, the specimen is oriented in order to have only two strongly excited beams, the transmitted and one diffracted. Then the contribution of the other beams can be neglected, because of their much lower intensity. Only two waves will thus be present, namely Φ_o and Φ_g. Under these conditions the displacement introduced by a defect can be described simply by a phase shift, $2\pi\bar{g}\bar{R}$. Using these equations the rules that determine the defect contrast can be deduced when the displacement field is known and then numerical calculations are possible.

In the next subsections examples of the images obtained for typical defects that appear in silicon after treatments used in integrated circuit processing will be presented. All the pictures are taken from samples that are studied in the following chapters.

1.1.1.1 Stacking faults

The displacement field $\bar{R}(z)$ of a stacking fault at a depth t under the surface of the specimen has the form:

$$\begin{aligned} \bar{R}(z) &= 0 \qquad \text{for } z < t \\ \bar{R}(z) &= \bar{R} \qquad \text{for } z > t \end{aligned} \tag{1.14}$$

where $\bar{R}$ is the displacement vector. The phase change, as indicated in the previous section, is given by $2\pi\bar{g}\bar{R}$. The conditions for which extinction is obtained, i.e., the defect will be invisible, are:

$$\bar{g}\,\bar{R} = n, \text{ where } n = 0, 1, 2, \ldots \tag{1.15}$$

In the cases where $\bar{g}\,\bar{R} \neq n$, a fringe contrast will be obtained for an inclined stacking fault. If the deviation parameter, $\bar{s}_g$, is zero, i.e., we are in the exact Bragg conditions,

the fringes will have a depth periodicity of 0.5 ξ_g. Contrary, a stacking fault lying parallel to the specimen surface will show one single dark or bright contrast fringe over the whole defect.

Stacking faults can be extrinsic, i.e, formed by the introduction of additional atomic planes, also called of interstitial type, or intrinsic (vacancy type), when formed by removal of lattice planes. Its character can be deduced if one follows the rules first proposed by Gevers et al. [1.6]. If the diffraction vectors employed are of the type $\bar{g}_{200}$, $\bar{g}_{222}$ and $\bar{g}_{440}$, $\bar{g}$ points towards a dark outer fringe in the dark field image for extrinsic faults and towards a bright outer fringe for intrinsic faults. This rule is reversed if the diffraction vectors are from the family $\bar{g}_{111}$, $\bar{g}_{220}$ and $\bar{g}_{400}$. An additional detail is that for these three diffraction vectors the top fringe, corresponding to the incidence surface, has the same contrast in both bright and dark field images, while the bottom fringe has the opposite contrast.

In practice the diffraction vectors which are employed for the determination of displacements fields in diamond type structures are $\bar{g}_{111}$ and $\bar{g}_{220}$, because they give the highest contrast. $\bar{g}_{400}$ is normally employed for confirmation of the predicted values from the other two reflections. For this type of investigations, plan view specimens are better suited than cross-sections because it is easier to orient the specimen in the microscope and it is possible to have a better overview of the defects present over a large area. As commonly the upper surface of the silicon wafers is a [001] plane, the reflection $\bar{g}_{200}$ is forbidden and will not be present, except for very thick specimens, where double diffraction takes place. If one works on cross-sections, this reflection will be very weak compared with the $\bar{g}_{400}$, so that in practice it is very difficult to obtain such a two-beam condition. In the examples of determination of displacement fields that will be given for illustration only the diffraction vectors $\bar{g}_{111}$, $\bar{g}_{220}$ and $\bar{g}_{400}$ will be employed.

1.1.1.2 Dislocations

The displacement vector for a mixed dislocation is given by the isotropic elasticity theory as:

$$\bar{R} = \frac{1}{2\pi}\left\{\bar{b}\phi + \bar{b}_e \frac{\sin 2\phi}{4(1-\nu)} + (\bar{b} \times \bar{u})\left(\frac{1-2\nu}{2(1-\nu)} \ln|\bar{r}| + \frac{\cos 2\phi}{4(1-\nu)}\right)\right\} \quad (1.16)$$

where ν is the Poisson's ratio, ϕ is the angle between the position of the vector, $\bar{r}$, and the dislocation line, $\bar{b}$, the Burgers vector of the dislocation, $\bar{b}_e$ is the edge component of the Burgers vector and $\bar{u}$, the line direction. The conditions that have to be fulfilled in order to get extinction are:

$$\bar{g}\cdot\bar{b} = 0 \quad (1.17a)$$

$$\bar{g}\cdot\bar{b}_e = 0 \quad (1.17b)$$

$$\bar{g}\cdot(\bar{b} \times \bar{u}) = 0 \quad (1.17c)$$

If the first of the three conditions is fulfilled but not the other two, residual contrast will be observed. Depending on the character of the Burgers vector and the value of $\bar{g}$, residual contrast can be observed and in certain conditions it can be as strong as the contrast for other conditions. This makes critical the choice of the correct diffraction conditions ($\bar{g}$ and $\bar{s}_g$). This requires that not only the image but also the diffraction pattern are always recorded.

For a dislocation the curvature of the lattice planes will be the largest close to the dislocation core. The image will thus not be on the position of the core, but near to it. Depending, again, on the diffraction conditions a displacement of the image of the dislocation line can be observed.

1.1.1.3 Dislocation loops

Similarly to the dislocations and stacking faults, the interstitial or vacancy nature of closed dislocation loops can be determined from the position of the intensity minimum as a func-

tion of the sign of the diffraction vector and the deviation from the Bragg position. The sign of the Burgers vector follows the rule:

$$(\bar{g}\cdot\bar{b})s_g > 0 \quad \text{outside contrast}$$
$$(\bar{g}\cdot\bar{b})s_g < 0 \quad \text{inside contrast} \qquad (1.18)$$

The magnitude s_g is defined positive when the reciprocal lattice point corresponding to $\bar{g}$ lies inside the Ewald sphere. The positive direction around the loop is taken clockwise looking down to the crystal from the electron gun and the sense of the Burgers vector is obtained with the FS/RH rule (final-start-right hand) [1.7].

When $\bar{n}$ is the upside normal to the loop plane in the electron microscope indexing system (that is reversed in the photographs because they are printed upside down), the character of the loops can be determined by:

$$\bar{n}\cdot\bar{b} > 0 \quad \text{for interstitial type loops}$$
$$\bar{n}\cdot\bar{b} < 0 \quad \text{for vacancy type loops} \qquad (1.19)$$

1.1.1.4 Examples

Working in the two beam conditions and obtaining images in which the defects are in extinction, it is possible to determine the displacement field of the defects as explained in the previous paragraphs. The overall procedure can be found in all the books referenced and is shortly summarised below.

1) A two-beam situation has to be obtained using low order reflections, e.g., $\bar{g}_{111}$ or $\bar{g}_{220}$, in an area with a thickness of about 3-8 ξ_g and without bend contours, so that the $\bar{s}_g$ vector does not change along the defect.

2) At least two or three different two beam conditions have to be obtained for which the dislocation is in extinction. It is necessary to record both the image and the diffraction dia-

gram, so that $\bar{g}$ can be correctly and unequivocally determined. For this purpose it is necessary to know which is the rotation between the image and the diffraction pattern.

3) The normal $\bar{n}$ has also to be determined. In the case of semiconductor material the normal is, for [001] oriented wafers, [001] direction when the specimen is prepared in plan view, and, mostly, [110] or [010] for a cross-section.

Condition 2) can be relaxed if only one extinction condition has been obtained, but several images in which the defect is in contrast have been obtained.

Figure 1.2 shows the dislocation loops generated after the annealing of a boron ion implanted wafer. Images with different $\bar{g}$ vectors have been recorded. At the top three loops are labelled from A to C. The loops A and C are both of the type $\frac{1}{2}$ [110], while loop B is of the type $\frac{1}{2}$ [$\bar{1}$10]. This can be proved by using the following table:

g	**g·b (A,C)**		**g·b (B)**	
[040]	2	contrast	2	contrast
[$\bar{2}$20]	0	extinction	2	contrast
[1$\bar{1}\bar{1}$]	0	extinction	1	contrast
[111]	1	contrast	0	extinction
[220]	2	contrast	0	extinction

The most common extended defects in silicon are of interstitial type, with very few exceptions. This makes usually the character determination not necessary. However, if this type of analysis has to be performed, the two lower micrographs should be used. It can be observed that defects A and C present an external lobe with the reflection $\bar{g}_{220}$, while with the opposite reflection the lobe is internal.

Similarly, figure 1.3 shows a stacking fault observed under three different diffraction conditions. In the upper one, fringes perpendicular to the diffraction vector are observed. In the central figure the SF is in extinction, while in the lower

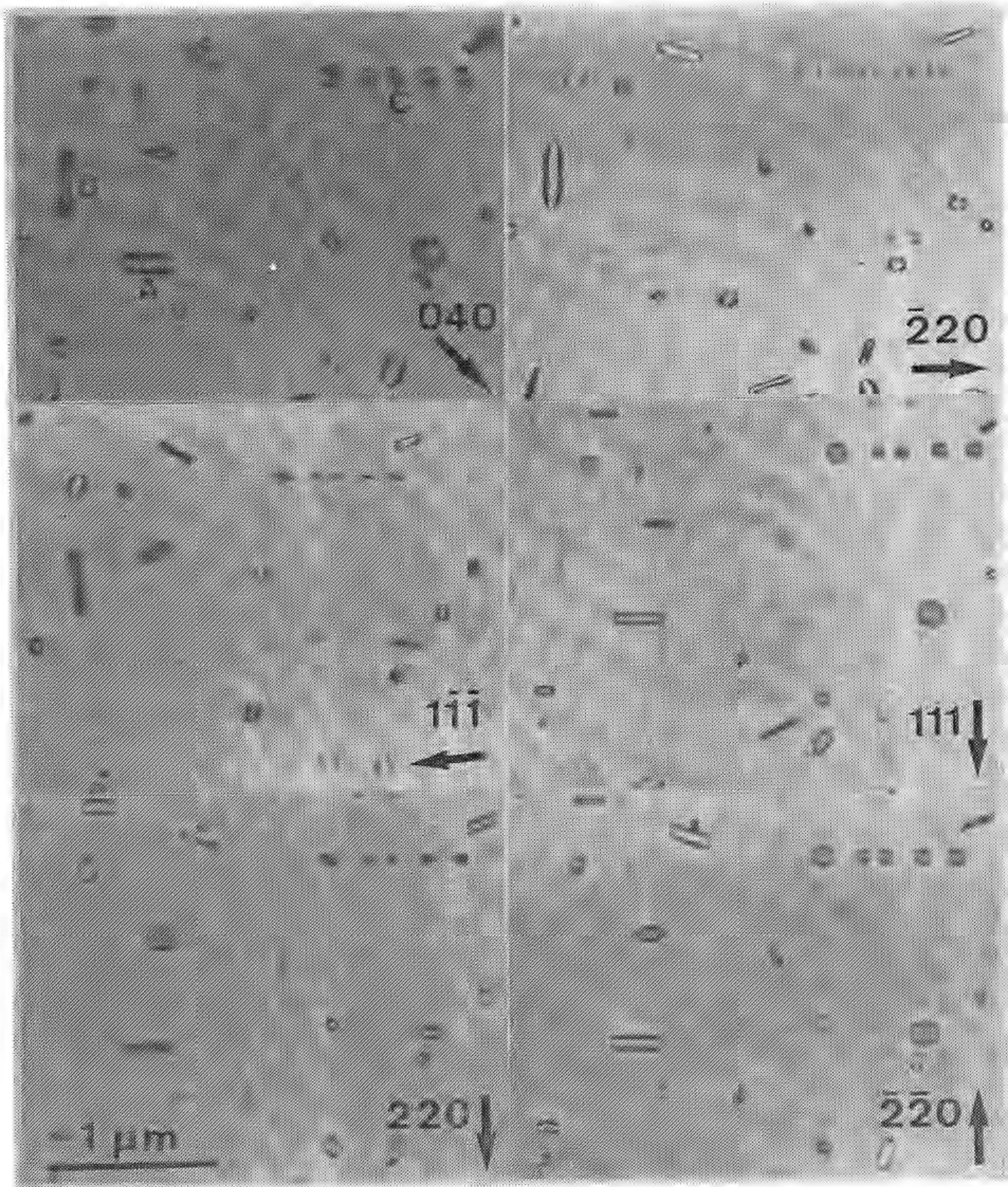


Figure 1.2: Determination of the Burgers vector of dislocation loops formed by ion implantation of boron and annealing of the wafer in a conventional furnace.

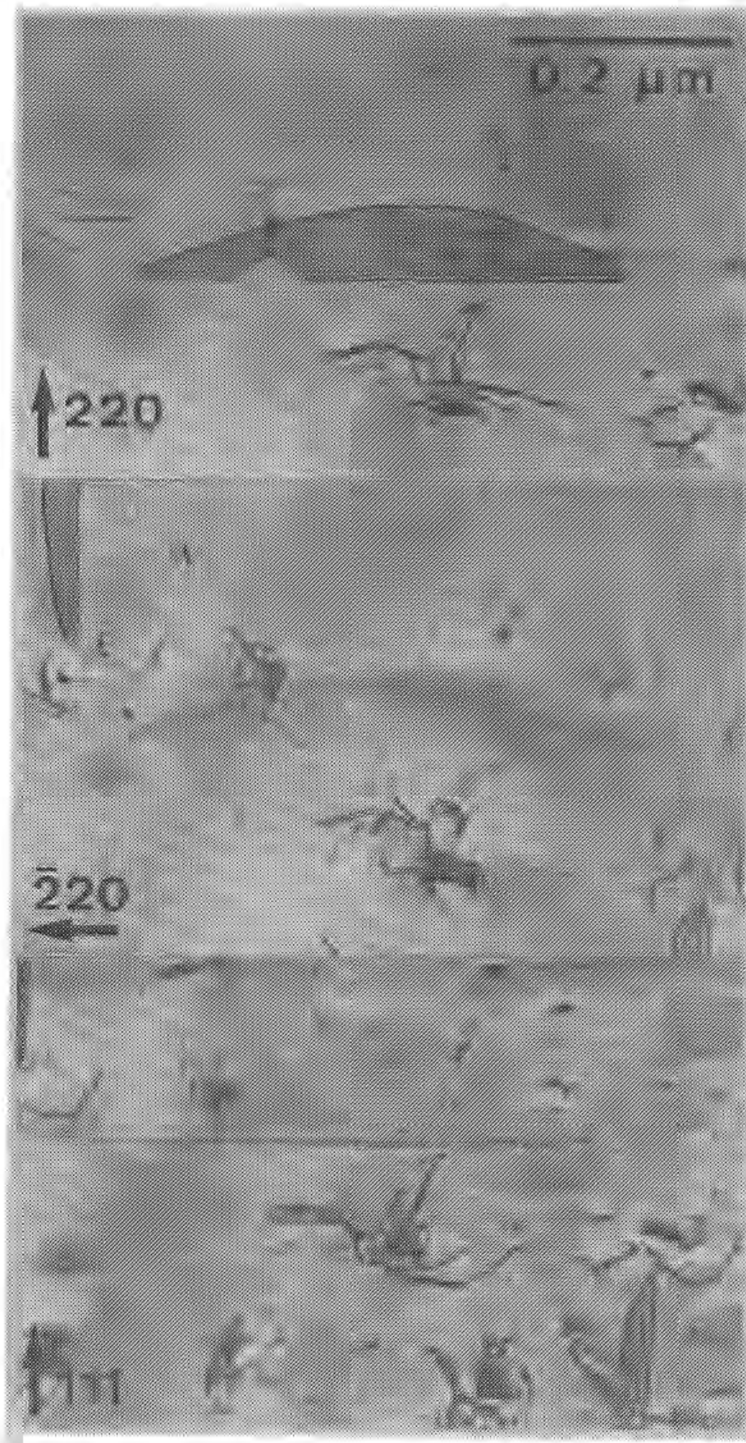


Figure 1.3: Diffraction contrast from a surface stacking fault.

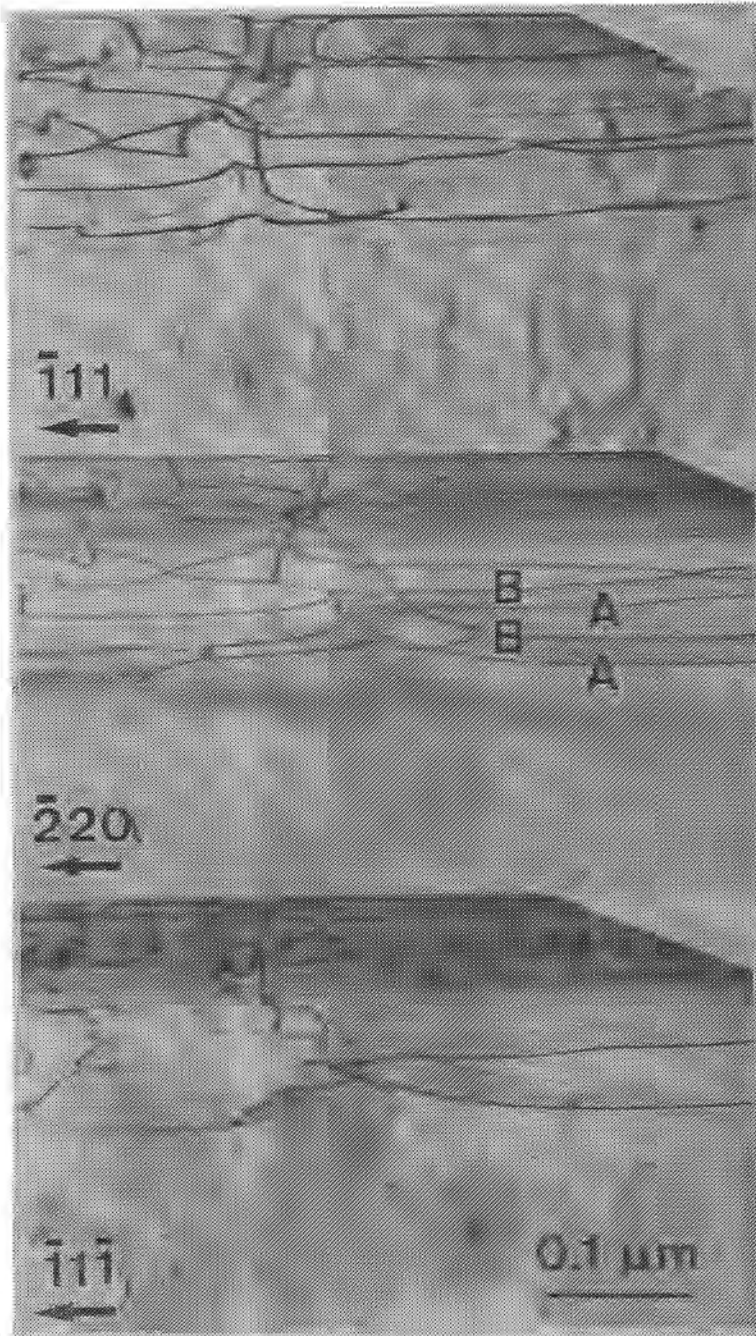


Figure 1.4: Dislocations generated by the high stresses during the LOCOS process. Pairs of dislocations lying in the same glide plane and having opposite Burgers vector (A and B) are visible.

one it is observed edge on. This implies that it lies in a {111} plane. This stacking faults have a displacement vector which is perpendicular to their habit plane, thus it is a $\frac{1}{3}$ [111].

The dislocations in figure 1.4 have been generated at the edge of a LOCOS mask. Two types of dislocations are visible: A and B. For the upper and lower images when one of the two types is in extinction, the other is in contrast, and the contrary. As will be commented in chapter 3, they lye in the same plane and have alternating Burgers vector.

1.1.2 High resolution electron microscopy

The second major mode of image formation is the phase contrast, which is used to obtain the highest resolution. A very good review on the conditions necessary to obtain high resolution images can be found in the book by Spence [1.8].

This imaging technique requires that the specimen is symmetrically oriented, i.e., that in the diffraction diagram beams opposite one to another have the same intensity. An objective aperture big enough to include the central beam and a number of diffracted beams is inserted and the image is formed by their interference. As a consequence, the image will appear to be strongly dependent on the lens aberration and the defocus settings. Depending on the characteristics of the microscope, a point to point resolution of the order of the atomic distances can be achieved. A drawback of this mode is that only very thin areas, less than 30 nm thick, can be observed with a good resolution. For thicker areas of the specimen the absorption and scattering is too high. An additional problem of this imaging mode is that in order to be able to image the structure of a defect, it has to be oriented along the axis of incidence of the electron beam and has to extend through nearly the whole specimen thickness. In that case the interaction between the beam and the defect is high enough to be imaged. If the defect is only one or two atomic layers thick, the rest of the crystal will contribute much more and no clear image of the defect can be obtained.

The ultimate resolution of a microscope can only be reached in quite restricted conditions. They include the use or a very intense light source with a very high degree of coherence, i.e. LaB_6 filaments or field emission guns. In addition the objective lens should have the smallest possible aberrations. This is obtained by having an objective pole piece with cylindrical symmetry, which will strongly limit the movement and orientation of the specimen. Commonly the specimen can only be tilted ±10°, thus allowing only to reach a very limited number of crystallographic sections. Recently instruments have become available allowing to tilt the specimen over much larger angles (±20°) while still keeping the high resolution.

The high resolution images in this work have been obtained with a Jeol 200 CX microscope with a double tilt goniometer stage. The point to point resolution of the instrument is about 0.25 nm. For silicon material this value allows only to resolve the [110] section. The thickness that can be used is very small, making the useful area per specimen very limited. A high density of defects is thus necessary to make the probability of having one of them in that area high enough.

The area that can be observed in high resolution can be increased by iterative thinning and observation of the specimen, as will be explained in the next chapter. This will allow to scan a complete device or to observe areas with a relatively low density of defects with high resolution.

As already pointed out before, the images obtained by working in this mode are formed by the interaction of the central beam with several diffracted beams. A more detailed study reveals that it is necessary to include also the interaction of the electron beam with the specimen, the formation of a diffraction pattern, the influence of the lens aberrations and defocus on the phase of the diffracted beams and the interference of these beams to form the image.

The theoretical treatment described in section 1.1.1 can also be applied for the present working mode, as for the deduction of the general formulae no special assumption has been

made. Thus, it is necessary to solve equation (1.6) to obtain the amplitudes of the diffracted beams. The diffraction pattern is the Fourier transform of the wave function at the exit plane of the object. The lens aberrations which are present will result in a phase shift of the diffracted beams compared to the transmitted. The aberrations that will introduce the major phase shift will be the spherical aberration and the defocus value. It can be shown that they give rise to a phase shift of the form:

$$\chi(q) = \frac{\pi}{2} C_s \lambda^3 q^4 + \pi \Delta f \lambda q^2 \qquad (1.20)$$

where C_s is the spherical aberration constant, typical for each instrument, λ is the electron wavelength, Δf is the defocus value and q is the distance in the reciprocal space.

The image can be expressed as the Fourier transform of the multiplication of the diffracted beams with the phase shift exp $(-i\chi)$ and an aperture function which limits the number of beams used to form the image.

Usually it is difficult to interpret the obtained image in a straightforward way, because only a planar projection of the atomic columns is obtained. For this reason image calculation and matching has to be carried out. Several models for the observed defect have to be used and simulated, choosing as solution the one that gives the best matching between the experimental and the calculated images. In this work no image calculation has been used as the goal of the high resolution images is not a structure determination but only a confirmation of the results obtained with conventional two beam imaging.

Figure 1.5 shows a high resolution image of a silicon-silicon oxide interface. The difference between the image of an amorphous and a crystalline material can be easily observed. In the figure each white dot represents the projection of two atom columns, separated by 0.143 nm, which is too small to be resolved by the Jeol 200 CX microscope.

Figure 1.6 corresponds to a high resolution image of a dislocation loop in boron ion implanted and annealed silicon. In

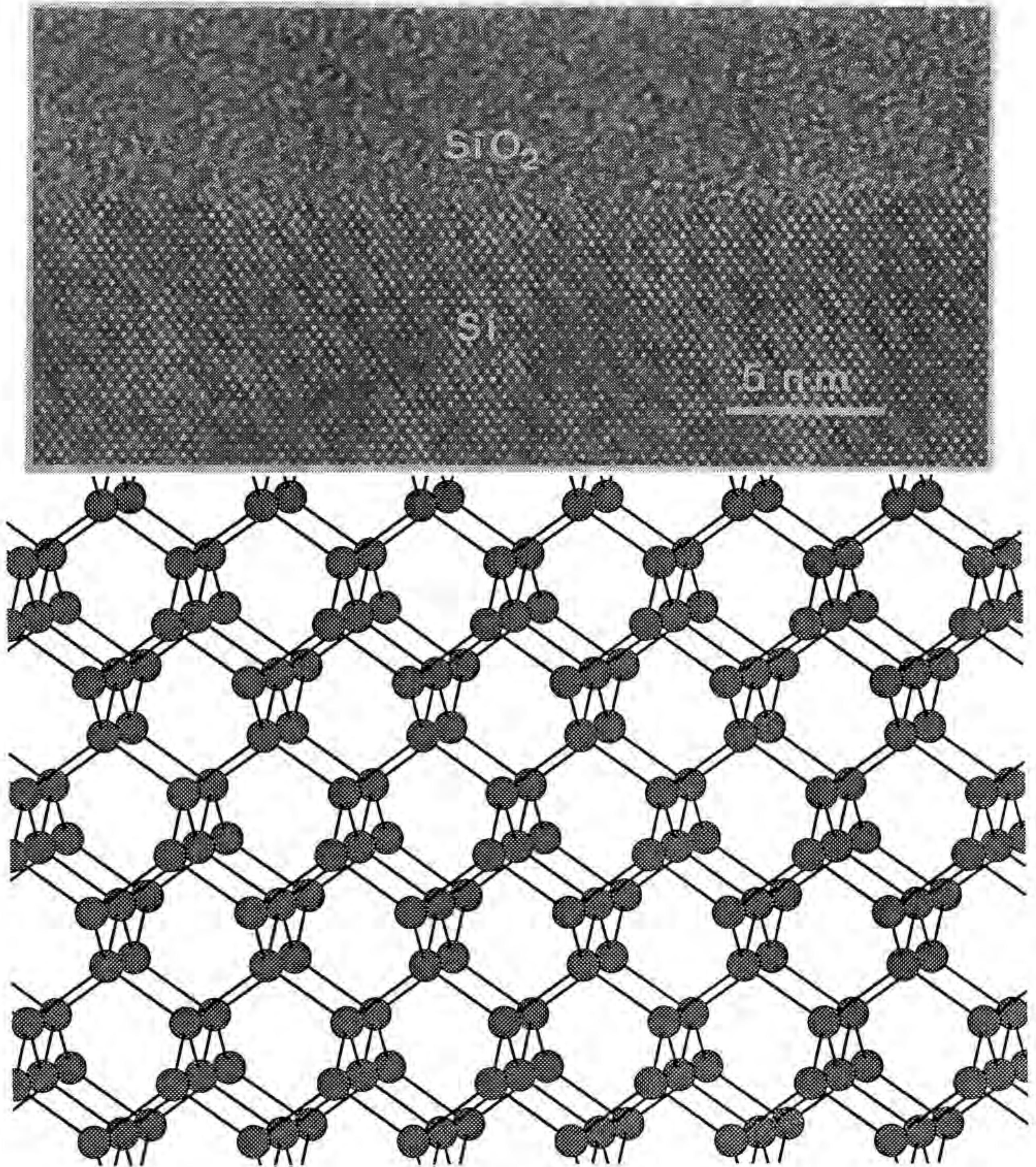


Figure 1.5: a) High resolution image of the silicon-silicon oxide interface. The direction of observation is [110]. b) Crystallographic structure of the silicon in this projection. The distance between two atoms is 0.143 nm, which cannot be resolved by the microscope. Each white dot corresponds to two atomic columns.

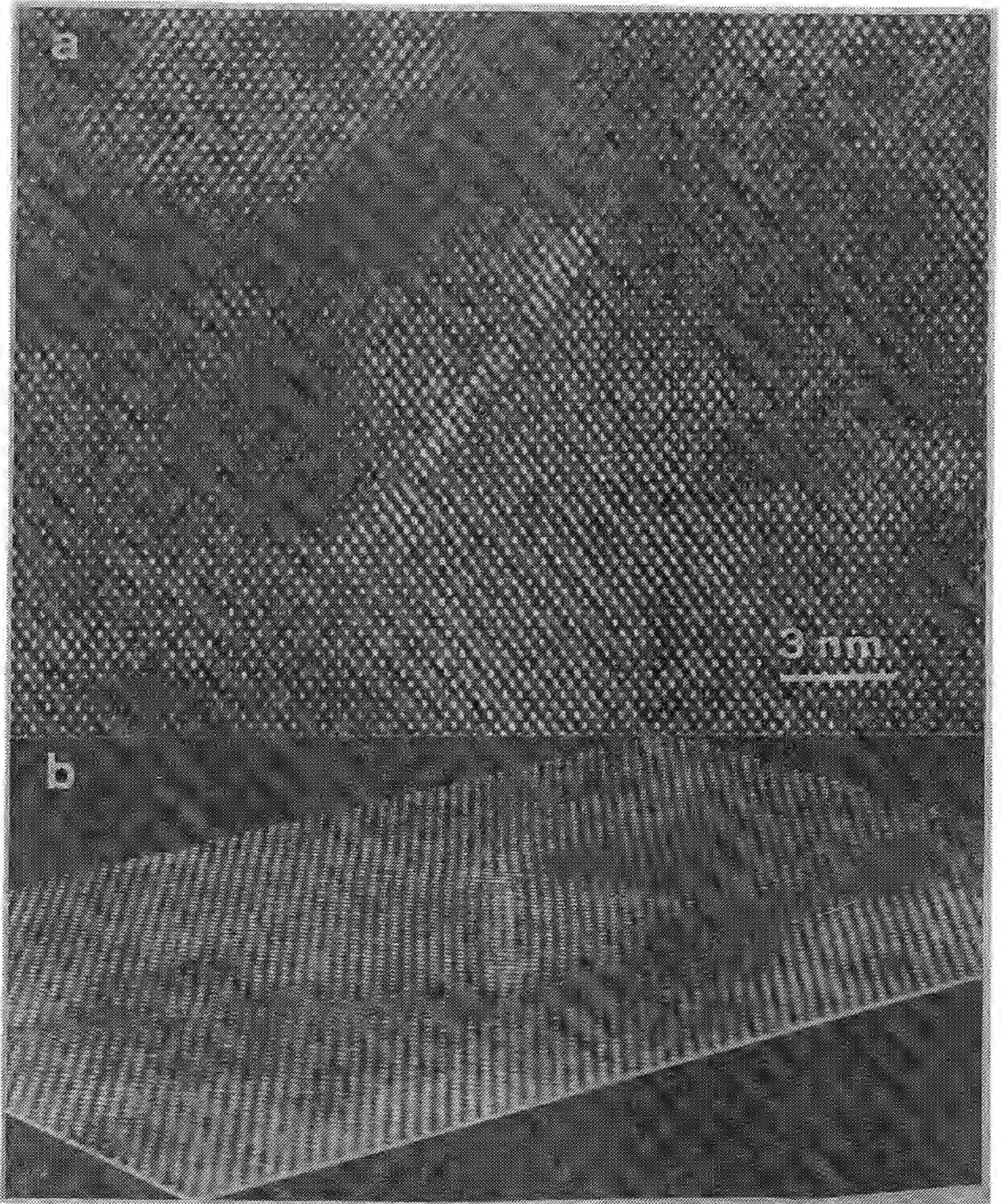


Figure 1.6: High resolution image of the edge segments of a dislocation loop in boron ion implanted and annealed silicon a) as seen in the microscope and b) using a special recording technique to enhance the additional lattice plane.

this specimen the segments that have a pure edge component are in the thin area, so that the additional atomic plane can be easily distinguished.

1.2 SECONDARY ION MASS SPECTROMETRY

Secondary ion mass spectrometry, SIMS, is a versatile and powerful technique for the determination of impurity profiles that are present in a sample. A recent review of its principles can be found in reference [1.9]. The technique is element specific and capable of detecting all elements, as well as isotopes and molecular species.

The principle of the technique is represented schematically in figure 1.7 and is based on the destructive removal of material from a sample by sputtering. This material is subsequently analyzed using a mass spectrometer. The primary ion beam, with energies of the order of 1 to 20 keV, impinges on the sample and atoms are sputtered or ejected from the sample. Most of the atoms are neutral and cannot be detected with conventional SIMS, but a small fraction, of about 1%, is positively or negatively charged and can be analyzed. The mass/charge ratio is determined and detected as a mass spectrum, a count number or displayed on a fluorescent screen or a CRT.

Sputtering is a process in which the incident ions lose their energy mainly by momentum transfer as they come to rest within the solid at a depth of about 1 to 10 nm. In this process they displace atoms within the sample. Sputtering takes place when the atoms near the surface receive sufficient energy from the incident ions to be ejected from the sample. The escape depth of the sputtered atoms is generally a few monolayers. The ion bombardment gives also rise to ion implantation and lattice damage and care has to be taken in the interpretation of the results.

The choice of the primary beam has a big importance in the sputtering yield of the different elements or clusters. For electropositive elements, like boron and aluminium in silicon,

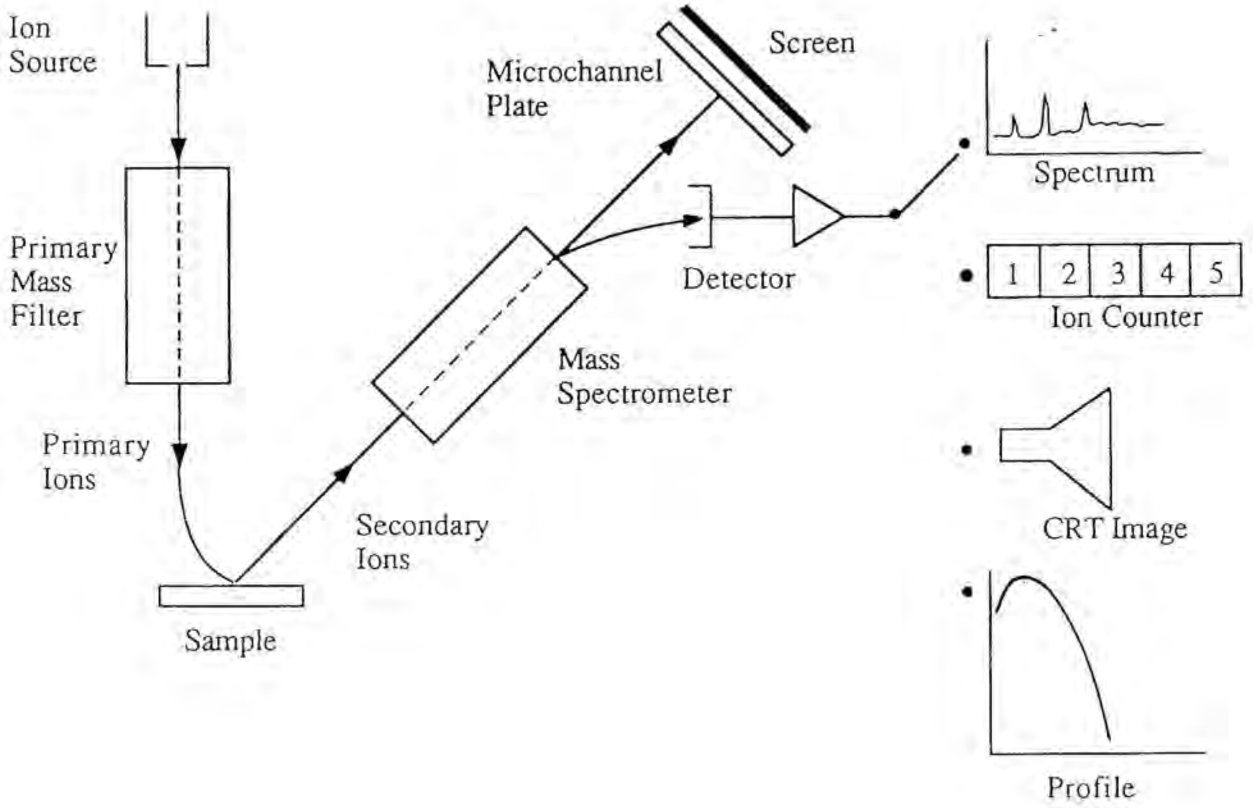


Figure 1.7: Schematic representation of the SIMS instrumentation.

an electronegative beam, commonly O_2^+, is used because it produces mainly positive charged secondary ions. The situation is reversed for electronegative elements, like phosphorus, arsenic and antimony in silicon, that have a greater yield when sputtered with electropositive ions like Cs^+.

Since the escape depth of the secondary ions is small, a few nanometer, this technique is ideal for the study of thin films or dopant profiles. Continued bombardment will lead to the erosion of the surface, so that this fact can be used for the investigation of deeper regions. A depth profiling for a certain element is obtained by tuning the mass spectrometer to this particular element. If a constant primary beam is used for the bombardment, the secondary ions are counted as a function of the sputter time. If the sample is not homogeneous, it is rather complicated to convert this count number into a concentration versus depth and it is difficult to obtain quantitative information. Anyhow, in the case of implanted samples the silicon wafer is quite homogeneous and the only introduced impurities are the dopant atoms. It is, thus, possible to obtain concentration as a function of depth for our case.

The main positive characteristics of SIMS are summarised below:

- A quite low detection limit, that in certain cases reaches 10^{14} to 10^{15} atoms/cm^3. This value will strongly dependent on the components of the matrix and on the element that has to be studied. In the application to silicon and its common dopants the detection limit is, mostly, of the order of 10^{15} atoms/cm^3. However, for phosphorus it is higher, of 10^{16} atoms/cm^3, because phosphorus and silicon-hydrogen clusters have a quite similar mass, so that there can be problems with the separation in the mass spectrometer.

- The lateral resolution of the machine is limited by the diameter of the primary beam, which has typical values of about 1 to 2 μm.

- The theoretical depth resolution of SIMS is of the order of 1 to 2 nm, which corresponds to the depth from which the secondary ions are sputtered. Due to practical problems and because there is not only sputtering but also impingement of part of the atoms present in the layer to be studied (collisional mixing), the depth resolution increases to a value of about 5 to 10 nm.

Some drawbacks of this technique are:

- The so-called matrix effect. The degree of ionisation of the different elements can be strongly dependent on the matrix, making the quantification difficult if the matrix is not well known. This is generally not the case for dopants in silicon.

- During the sputtering process, a crater is formed in the sample. Most of the ejected ions come from the bottom of the crater, but also sputtering takes place at its walls, giving rise to the edge effect. This effect can be partially solved by using adequate magnetic lenses that eliminate the ions that do not come from the centre of the crater.

- Collisional mixing of the ions of the different depths can take place, which reduces the depth resolution, as already pointed out.

All the SIMS spectra that are presented in this work are already converted into quantitative concentrations, giving directly dopant profiles.

1.3 SPREADING RESISTANCE PROBE

The spreading resistance probe [1.10], SRP, is an electrical dopant profiling technique that is widely spread in the semiconductor research and development laboratories. The information that can be obtained with this technique is the resistivity of the sample and, by an adequate algorithm, the electrically active dopant concentration can be deduced. Thus, the re-

sults of SIMS and SRP do not have to coincide, unless total activation of the dopant takes place.

The working principle of this technique is based on the resistance which is associated with the divergence of the current lines that emanate when a small-tipped electrical probe is placed onto the silicon surface. The equipotential lines form concentric circles, whose centre is at the probe. The current flowing in the direction of the electric field is concentrated at the probe tip and spreads out radially from the tip. Hence the name spreading resistance. In this technique the spreading resistance of a reproducibly formed point-contact is measured. This can be achieved by using one, two or three probes, although the configuration of two probes is the commercially available.

To make a spreading resistance measurement, a known current is applied between two probes and the voltage drop is measured across these probes, so that the spreading resistance is obtained, $R_{SR} = V/I$. This resistance is measured in a very small volume of silicon immediately under the probe. In the two probe method, the resistivity ρ is related to R_{SR} by the relationship:

$$R_{SR} = \frac{\rho}{2\ a} \tag{1.21}$$

where a is an empirical quantity, which is related to the effective electrical contact radius. For a hemispherical probe with radius r, $a=\pi r$, while for a cylindrical probe of radius r, $a=2r$.

To perform a spreading resistance measurement to determine the carrier profile, the wafer has to be prepared first. This is accomplished by angle lapping the wafer with an angle θ, typically in the order of a few minutes of degree, up to 5 degrees. In this way an enlarged cross-section is obtained. The magnification can then be calculated from:

$$\Delta z = \Delta x \sin \theta \tag{1.22}$$

where Δx represents the distance between two successive measurements and Δz, the corresponding increment in depth. Two tung-

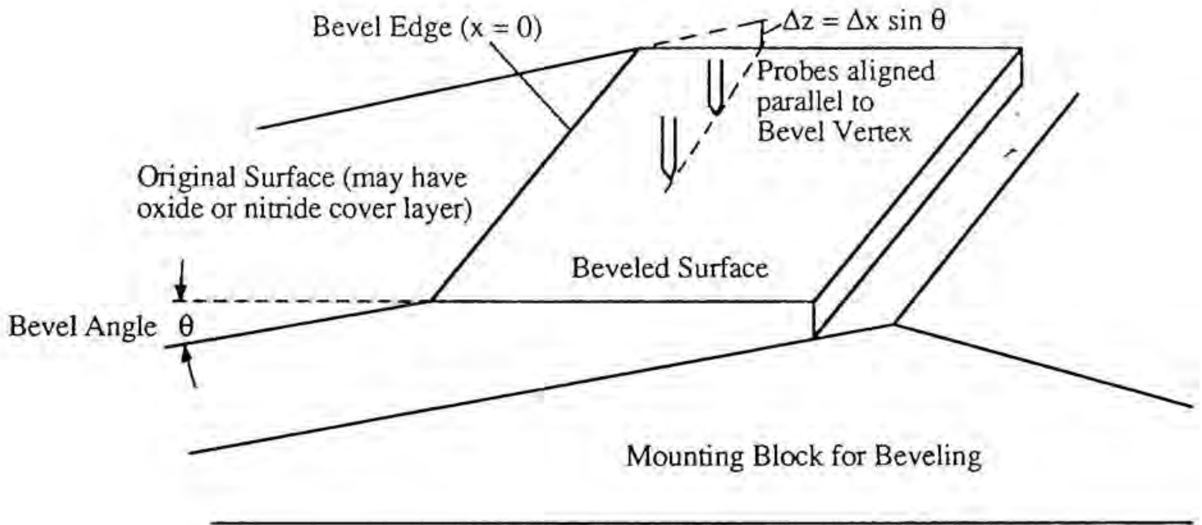


Figure 1.8: Schematic representation of the SRP technique.

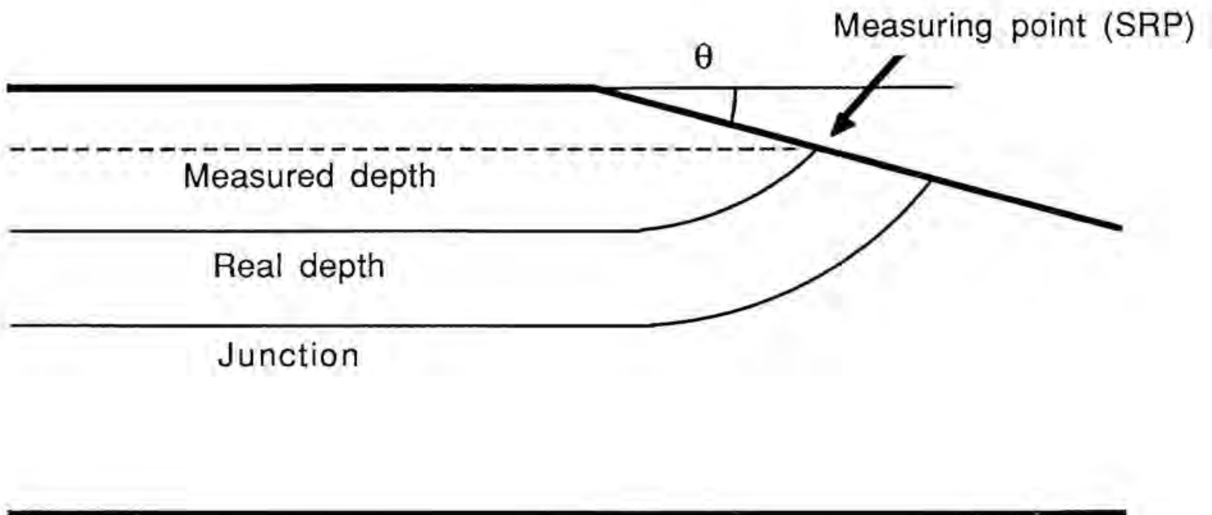


Figure 1.9: The carrier spilling effect in SRP samples.

sten-osmium point probes, separated about 30 μm one to each other and having a size of about 1.5 to 3 μm, are used. They are placed in a way that the line that connects them is parallel to the edge of the wedge, as shown in figure 1.8. A pressure of about 5 g is applied to each probe, so that good ohmic contact with the specimen is assured. Due to this pressure the tips penetrate in the sample 5 to 10 nm. Next a small tension, 5 mV, is applied between the two tips and the resulting current is measured. Scanning over an area large enough to contain the complete junction is necessary. In this way a set of data is obtained, each corresponding to a certain distance from the bevel edge and, thus, to a certain depth. The final and most complicated step is to deconvolute the values of resistivity and convert them into dopant concentration profiles.

Theoretically it is possible to measure any dopant concentration in a silicon wafer. A serious problem for the quantification is introduced by the so-called carrier spilling [1.11], which becomes important for dopant concentrations below 10^{17} cm^{-3}. This effect is schematically represented in figure 1.9 for a varying dopant profile, where the higher doped regions are close to the surface of the specimen. When a bevel is created, the lower doped regions are exposed, so that the probes can scan and measure them. By the removal of the higher doped layers the boundary conditions for the charge distribution in these areas change, giving rise to a "bending" of the free carrier concentration. When deconvoluting the dopant profile from the resistivity measurement, the dopant profile will drop faster than the real profile in the sample. As a consequence a correction has to be made to account for the carrier spilling. An important drawback in this moment is that no standard correction is available and that the correction varies with the sample preparation, making it difficult to apply to any sample. The correction of the carrier spilling has not been applied to any of the profiles used in chapter 4. For this reason only the results corresponding to doping levels above 10^{17} cm^{-3} are used.

REFERENCES CHAPTER 1

[1.1] P. Werner and M. Pasemann, Ultramicroscopy 7, 267 (1982).

[1.2] M.D. Matthews and S.J. Ashby, Phil. Mag. A 27, 1313 (1973).

[1.3] S. Amelinckx, Sol. St. Phys., Suppl. 6 (1964).

[1.4] R.D. Heidenreich, Fundamentals of Transmission Electron Microscopy, Interscience, New York, 1964.

[1.5] P.B. Hirsch, A. Howie, R.B. Nicholson, D.W. Pashley and M.J. Whelan, Electron Microscopy of Thin Crystals, Butterworths, London, 1967.

[1.6] R. Gevers, A. Art and S. Amelinckx, phys. stat. sol. 3, 1563 (1963).

[1.7] J.P. Hirth and J. Lothe, Theory of Dislocations, Mc Graw Hill, New York, 1968.

[1.8] J.C.H. Spence, Experimental High-resolution Electron Microscopy, Clarendon Press, Oxford, 1981.

[1.9] A. Benninghoven, F.G. Rudenauer and H.W. Werner, Chemical Analysis vol. 86, eds. P.J. Elving, J.D. Winefordner and I.M. Kolthoff, John Wiley and Sons, New York, 1987.

[1.10] R.G. Mazur and P.H. Dickey, J. Electrochem. Soc. 113, 255 (1966).

[1.11] S.M. Hu, J. Appl. Phys. 53, 1499 (1982).

2. SPECIMEN PREPARATION TECHNIQUES

2.0 INTRODUCTION

As shown in the previous chapter, the use of the TEM for investigations in solid state physics has been generalized in the last years, thanks to the high spatial resolution of the microscopes nowadays available. One of the major drawbacks of the TEM is that the specimen have to be extremely thin in order to be transparent to the electrons. This can give two different problems: from one side, preparation of specimens is not straightforward and special techniques have to be developed in order to minimize the artifacts introduced; from the other, due to the small area that can be used for observation, small statistics is obtained and sometimes it is difficult to generalize the obtained information. This happens, for example, when looking at dislocations in integrated circuits. The dislocation line can be several microns long and only a small portion of it can be observed in the conventional TEM.

In the last years a large number of papers has been devoted to the preparation of specimens of semiconducting materials transparent for the electron beam. Different thinning methods have been proposed, the principals of which are the following:

- Cleavage of the specimen [2.1,2.2], which has the principal advantages that the procedure is performed without the use of any special equipment and that the wedge geometry is well known. The major disadvantage is however that the crystallographic sections in which the specimen cleaves are basically limited to the [110] orientation, so that the area of view is always very narrow.

- Ion milling with [2.3-2.6] and without [2.7,2.8] mechanical prethinning. This is a fast and reproducible method of preparation. Principal advantages are the reproducibility and the wide transparent area which is obtained. Disadvantages are the difficulty of thinning in a localized area and also the possi-

bility of introducing radiation damage in the specimen, although an improvement has been recently presented [2.9].

- Chemical thinning of the specimen with [2.6,2.10] or without [2.11,2.12] mechanical prethinning. This technique is very useful for plan view preparation, but is difficult when applied to cross-section specimens due to the different etching rate of the different layers of which the sample is composed.

Also very special techniques have been developed in order to thin the sample at a predetermined region, but then dedicated equipment, which is not commonly available, has to be used. For example:

- Mechanical thinning until the complete sample is thin enough for its inspection with the electron microscope [2.13]. Special and expensive equipment has to be used, requiring a skilled operator to obtain the samples. A drawback for high resolution investigation is that further thinning until perforation is required and that the bending of the sample close to the perforation makes it unstable and difficult to orient.

- Lithographic techniques and reactive ion etching [2.14], which require also special and expensive equipment, but offer the possibility of preparing samples of selected regions.

- Use of a Scanning Ion Microscope [2.15] allowing simultaneous observation and cutting of the areas around the structure that has to be investigated.

- Use of a Precision Ion Milling System, which incorporates an SEM to the Ion Miller, so that the etching process can be monitored continuously.

The technique used throughout this work has been published in various papers [2.6,2.9] and is based on mechanical prethinning followed by ion milling or chemical etching of the sample. A Laser Autoterminator in the milling stage allows to stop the milling just before or at the moment when a very small hole is formed. Prethinning is performed in two steps: flat grinding followed by dimple grinding [2.16]. The use of a dimpling device

allows to obtain a very thin central area, reducing the required ion milling time to a few hours. At the same time a reasonably resistant border of the sample is still kept which allows to handle the thinned specimen easily with tweezers. An additional advantage is that the described preparation is reasonably fast and that the obtained specimens are of high quality. Furthermore no special skills or sophisticated equipment are required.

Procedures both for plan view and for cross-section TEM specimen preparation of semiconductors and integrated circuits are presented in detail. Also the extension of the technique for the preparation of prespecified regions of integrated circuits and for difficult semiconductors materials, as InP, are presented.

2.1 CROSS-SECTION SPECIMENS

The cross-section preparation procedure is schematically represented in figure 2.1. Its principal steps are the following:

a) Sawing of the wafer in order to have stripes of rectangular size 2 mm x 350 μm x wafer thickness.

b) Glueing of two of the stripes, surface to surface. In this way the surfaces are protected against damaging during the subsequent mechanical prethinning procedure.

c) Grinding and polishing of the sample from both sides to remove the sawing damage and to obtain a sample with smooth and flat parallel sides.

d) Dimpling of the flat parallel specimen in order to thin the central part while still having relatively thick borders allowing to pick up the specimen using tweezers.

e) Ion milling until perforation using low settings, i.e. moderate high tension and current, in order to keep the damage introduced in the specimen at the lowest possible level and to

keep the operational time of the milling equipment as high as possible.

2.1.1 Sawing of the wafer

The sawing of the wafer is performed with a thin-blade diamond saw used for cutting integrated circuits (Model 1100 Wafer Dicing Saw, Micro Automation Inc.). This saw is computer controlled and uses a very thin blade (30 μm) rotating at about 40.000 rps. The cutting speed is very high and of the order of 5 mm/s. To take away the material removed during the cut and to cool the wafer a water flux is used during sawing. The equipment can be programmed in order to perform cuts in two different directions with any angle between them. For cross-section TEM preparation a set of parallel cuts is performed in two perpendicular directions, one with a spacing of 2 mm and the other with a spacing of 380 μm. The result are prismatic stripes of 1.97 mm x 350 μm x wafer thickness. From the area of interest two stripes are selected to be glued.

It is necessary that the wafer lies very flat related to the surface of the cutting table and that it is well fixed to avoid deviations of the cutting direction. For this purpose a piece of self-adhesive plastic is used. The wafer is glued to this plastic with the back face against it. Then the whole assembly is put on a hot plate at 90°C to have the maximum adhesion of the wafer and to reinforce the glueing power. After a few minutes the whole is removed and fixed on the table of the saw using adhesive band. When special structures are present the wafer has to be oriented first rotating the sawing table in order to perform the cuts in well determined directions determined by these structures. For this purpose the video camera of the equipment is very useful. Once correctly oriented the cutting procedure can start and is fully automatic. The cutting takes then about 15 minutes.

The principal advantages of using a saw instead of cleaving the specimen are that the dimensions of the stripes are well defined and that the cut can be performed in all crystallographic directions and not only in the [110]. This is very useful when

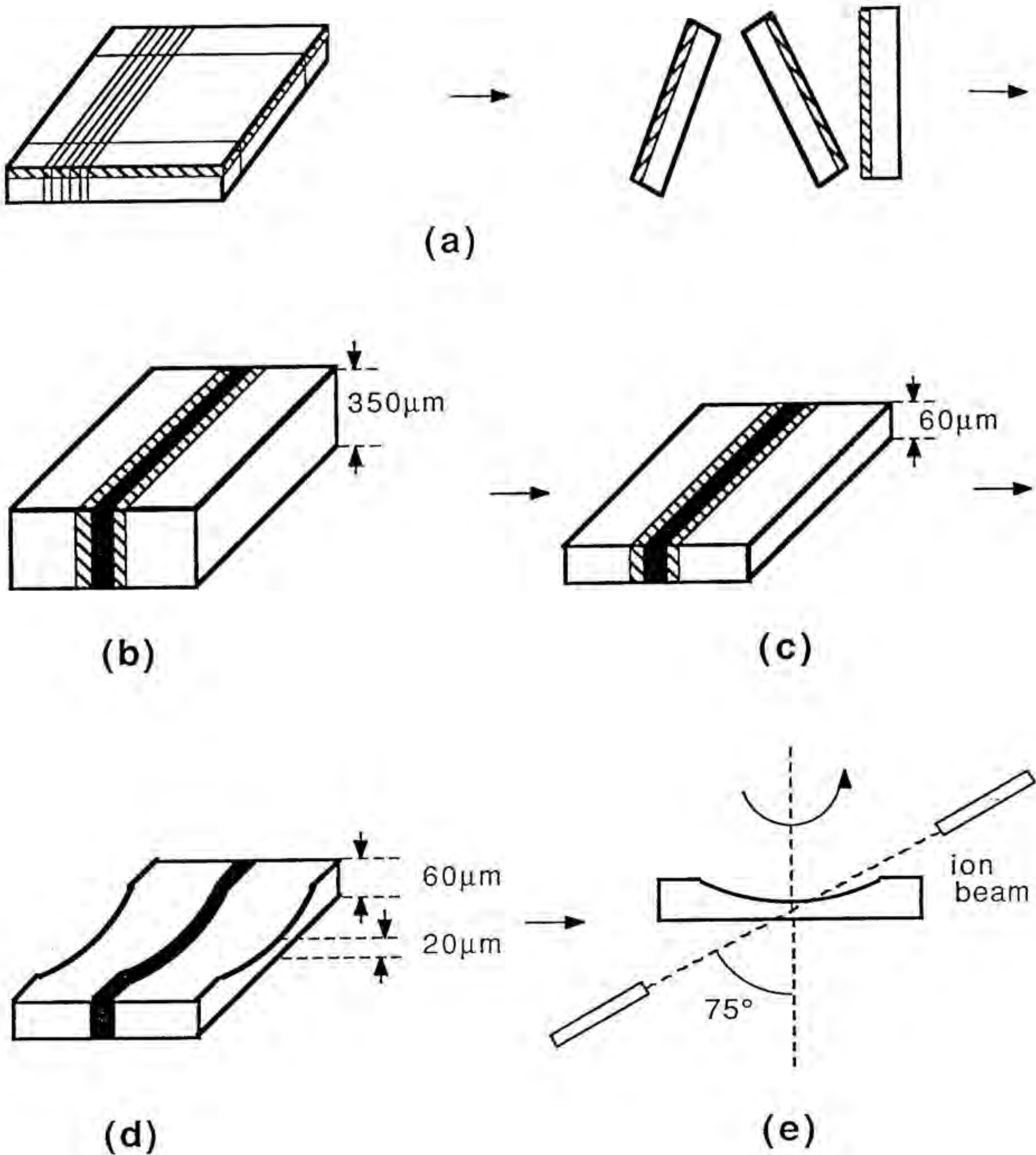


Figure 2.1: Schematic representation of the cross-section preparation procedure. The different steps are: a) Cutting of the wafer; b) glueing of the stripes; c) flat grinding and polishing of the sample; d) dimpling; e) ion milling.

features have to be studied which are only visible along well defined directions parallel to the surface of the substrate. A typical example is the study of crystalline precipitates with a well known orientation relationship with the semiconductor substrate. Cutting along the appropriate directions will allow to image the lattice of the precipitate using HREM, however without being able to resolve the surrounding substrate lattice, as observed in figure 2.2, which corresponds to an antimony precipitate in a silicon matrix [2.17].

Another example of a more technologically important application of the sawing technique is the study of the geometry of integrated circuit structures oriented along directions differing from the more conventional [110] on the silicon substrate [2.7]. Especially in those cases where the strong anisotropy of silicon plays an important role in the physical processes involved in a processing step of integrated circuits, the use of this sawing technique proves very useful to study the influence of the orientation of the structures on the silicon wafer on geometrical parameters or on defect generation.

2.1.2 Glueing of the stripes

The stripes selected from the area of interest are glued together active surface to active surface using M-610 Bond (Measurements Group, Inc.), making a small pressure with tweezers in order to minimize the thickness of the glue between the stripes. The whole operation is performed on a flat teflon base as the glue does not stick on it. After the glueing the base with the stack on it is introduced in a furnace set at 150°C for about 15 minutes. When taken out of the oven the sample is easily removed from the base by exerting a small lateral pressure on it.

2.1.3 Flat grinding

This step is carried out with the Gatan Disc Grinder. The cross-section stack is glued to the support cylinder with the face that was against the teflon base against it, because this is initially the flattest surface.

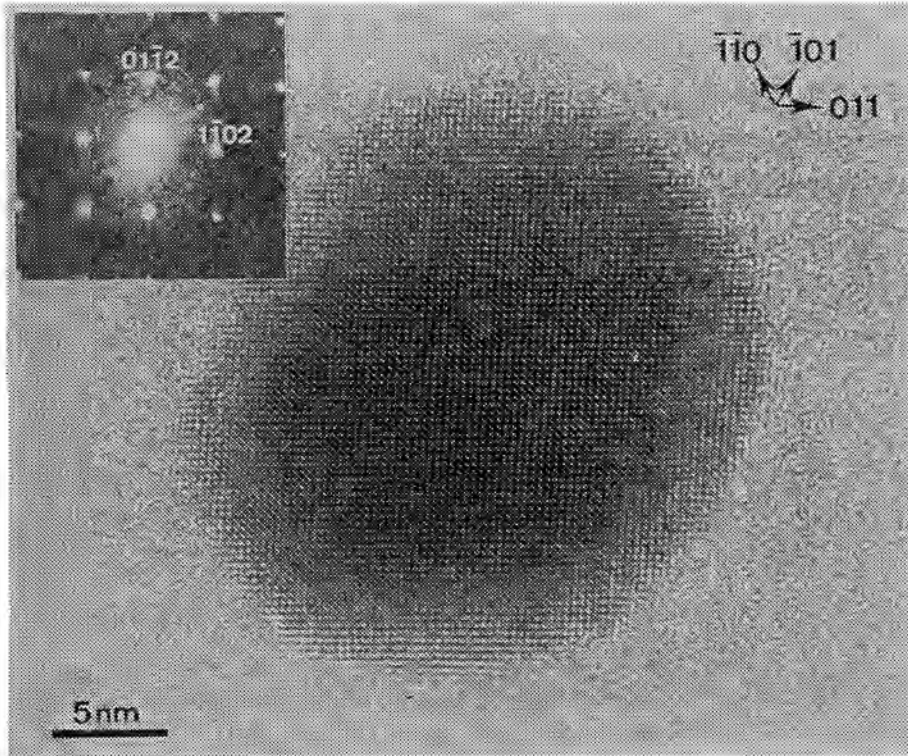


Figure 2.2: HREM image of an antimony precipitate in the [$\bar{2}$021]$_{Sb}$ orientation. The silicon lattice is not resolved as its orientation is approximately 3° off the [111]$_{Si}$ orientation around the [231]$_{Si}$ axis. Such a result can only be obtained using a sawing technique to cut the substrates. (Courtesy of H. Bender [2.17]).

First the support cylinder has to be heated using a hot plate kept at a temperature of about 130°C. When the cylinder is hot enough a drop of low melting point wax (e.g. the wax supplied by Gatan) is put on its top surface. Immediately the specimen is put on the cylinder and the whole is taken away from the hot plate and put on a metal base to diffuse rapidly the heat to solidify the wax. At the same time a small pressure has to be exerted on the surface of the cross-section to reduce as much as possible the thickness of the wax film between the specimen and the support cylinder. If too much wax is still present an appreciable error in measuring the thickness of the sample will be made. At the same time the specimen has to be well embedded in wax to avoid to be crashed during grinding. When the whole is cold enough, the cylinder is mounted in the grinding tool. This is a heavy steal cylinder of 7.5 cm diameter and 1.5 cm height with a concentric cylindrical hole of 0.9 cm in the centre, where the support cylinder with the sample can be introduced. At the end of this hole a steal knob (control knob) acts as stop for the support cylinder. This knob can be turned clock or counter-clockwise in order to set the relative height between the support cylinder and the grinding tool and is connected to a scale which indicates directly the specimen thickness if the zero has been set when both are at the same height.

Now the grinding itself has to be performed. The zero has to be adjusted in order to control accurately the remaining thickness of the sample. For this purpose the control knob has to be turned until the grinding tool and the support cylinder are at the same height, using an optical microscope glass slide to control it. Then the zero mark is made to coincide with the zero of the scale. Turning the control knob the cylinder is lowered relative to the holder until the specimen does not overpass the surface of the holder, typically 370 µm (due to the thickness of the wax and to the glue that overpasses the two stripes) and this value is displayed in the control knob scale. Fine abrasive paper is glued to a flat rectified glass base in order to grind the sample as flat as possible. Then water is added as lubricant and the grinding tool is put on it. The grinding is performed moving the tool over the paper in "8" shape without

exerting any pressure. The sample is ground down to about 300 μm with the 40 μm paper in 25 μm steps, further to 290 μm with the 5 μm paper (both papers supplied by 3M) in 2 to 3 μm steps and finally with the 0.3 μm paper (supplied by Buehler Ltd) for about 4 μm more in two steps, controlling optically that the surface is smooth and that no scratches are present.

Next the cylinder is removed from the Grinder. To remove the specimen first the wax located around it will be removed using a wood stick with a cotton rub and chloroform. Then the remaining wax is located only under the specimen, with a thickness of around 2 μm. The whole cylinder-specimen is put on the hot plate set at the same temperature as for glueing the specimen and after 10 to 15 seconds the wax is liquid and a small pressure is made with tweezers perpendicular to the longest side of the specimen, until it begins to displace. The rest of the wax assures that the specimen remains glued to the tweezers. The specimen is introduced in a beaker with chloroform to dissolve the wax and, next, dried with methanol.

Now the specimen is turned around and the polished face glued against the cylinder. Using again 40 μm and 5 μm grinding paper the specimen is ground down first to 70 μm and secondly to 60 μm.

Normally the flat thinning of the specimen takes about 20 minutes.

2.1.4 Dimpling

The Gatan Dimple Grinder (from now on called dimpler) consists of a base where the sample is installed and that turns around a vertical axis and a wheel that turns around a horizontal one. Putting some abrasive solution on the sample, lowering the wheel and connecting the rotation of both, a small crater is formed in the specimen. In our machine the base turns at a constant speed of about 6 rpm, the rotation of the wheel can be varied continuously from 1 to 8 rps, the weight varied from 0 to 40 g and the diameter of the wheel is 15 mm.

Without removing the sample from the cylinder, it is mounted in the dimpler. As before the zero has to be adjusted very accurately and the sample has to be centred in order to dimple the specimen in its centre. For this purpose a centring ring is supplied. The specimen is dimpled using 3 µm diamond compound to the thickness of 20 µm in about 20 minutes (diamond compound and lubrication liquid specially for hard materials, type DE/SY, supplied by PRESI S.A.). The weight applied to the wheel is 30 g and the rotation speed of the wheel is 2 rps. After this a very fine polishing compound (alumina 0.05 µm) and a wheel, approximately of the same diameter, with a felt stripe glued on its perimeter, is used to remove the surface damage. Sometimes this final step is performed with Syton X30 (Colloidal Silica, with particle size of approximately 25 nm, supplied by Monsanto), which is used in microelectronics for very fine polishing. In both cases the weight is lowered to 25 g and the speed increased to 3 or 4 rps. This final step takes normally around 10 minutes, but if scratches are not totally removed longer times are required.

Further dimpling down to 10 µm thickness is possible, but results are difficult to reproduce, due to variations of the thickness of the wax from one sample to another.

Again the specimen is removed from the cylinder, as explained previously, and cleaned thoroughly.

One of the major problems encountered with the dimpler is damaging of the dimpling wheel. During the dimpling not only material of the sample is removed but also of the wheel. The wheel, the surface of which was originally of spherical shape, has finally an uncontrolled shape and this produces an irregular dimpling of the sample. In the worst case deep scratches are created in the sample which cannot be removed by the final polishing step. In this case the wheel has to be replaced or rectified by removing material and leaving its diameter slightly smaller.

2.1.5 Ion milling

The dimpled sample is mounted on a 2.3 or 3 mm copper grid with a rectangular slot of 1.5 x 1 mm^2, putting M-610 Bond all around the aperture and gently placing the specimen on the grid. A small pressure has to be applied to each end of the sample to reduce the glue thickness between the grid and the dimpled specimen. This is important because when the amount of inserted glue is too large the thinnest area of the specimen is not located in the focusing point of the ion beams and, consequently, the milling time is increased considerably and the specimen quality will be much poorer. To cure the glue the sample is again heated in the furnace at 150°C for about 10 minutes, because now no mechanical steps are required.

Next the sample is mounted in the specimen holder and introduced into the Gatan Duo Ion Mill (in the rest of the chapter called ion miller). The settings which are used are a gun current and a high tension of 0.6 mA and 5 kV respectively. The angle of incidence with the normal to the specimen is 75°, the specimen holder turns continuously at about 1 to 2 rpm in order to avoid preferential milling in particular directions and argon ions are used. Typically the ion milling takes 1 to 2 hours, depending on the state of the cathode plates and the real thickness of the dimpled area of the specimen.

The Laser Autoterminator proves very helpful to stop the milling just at the moment when a small hole is formed. It consists of an helium-neon laser beam perpendicular to the surface of the sample and a detector at the other side of it. When the specimen is very thin or a small hole is formed, part of the light can pass through it. When the light intensity exceeds a prefixed value the milling is automatically stopped. The sensitivity level can be set at the appropriate value depending on the type of sample and/or of the size of the hole that is wanted. For silicon the sensitivity of the Autoterminator has to be calibrated carefully because silicon is transparent to light before it is to electrons.

The specimen holder with the cross-section sample is removed from the ion miller and the specimen inspected using an optical microscope. If the hole is located in the glue line this assures that the surface of the specimen will be thin enough, because it is located at the edges of the hole. When optical inspection allows to recognize device structures in cross-section further thinning can be performed until the required area is reached.

Figure 2.3 shows an optical image of a completely finished specimen using partly transmitted light. The large transparent region is visible in the centre of the specimen.

The described cross-section preparation technique is routinely used for silicon and for gallium arsenide on silicon materials with a high yield of good samples, useful both for conventional and high resolution TEM. A typical result is given in figure 2.4, showing a cross-section HREM image of the interface area between two bonded wafers. The lattice of the upper wafer is not resolved as it has a slightly different orientation. Steps can be observed on the substrate surface.

2.2 PLAN VIEW SPECIMENS

The principal steps followed in this procedure are represented schematically in figure 2.5. They are:

a) Cutting a disc from the wafer using an ultrasonic cutter.

b) Flat grinding of the disc only from the backside.

c) Dimpling to 20 μm.

d_1) Chemical thinning or d_2) ion milling, depending on the nature of the sample.

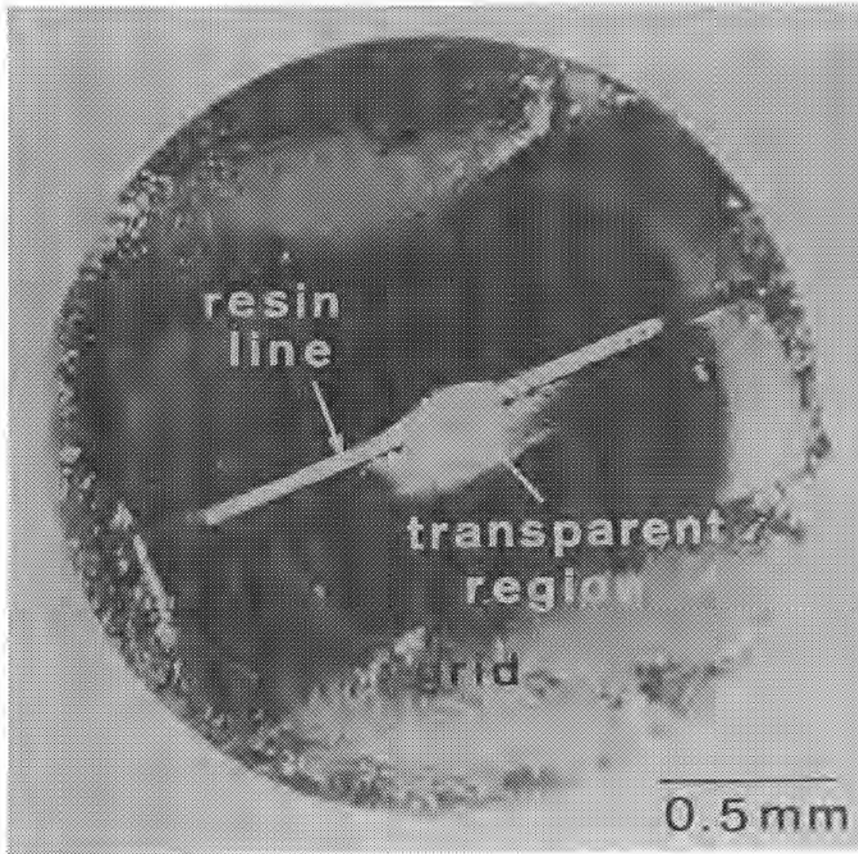


Figure 2.3: Cross-section specimen observed using transmitted and reflected light, showing the large transparent area.

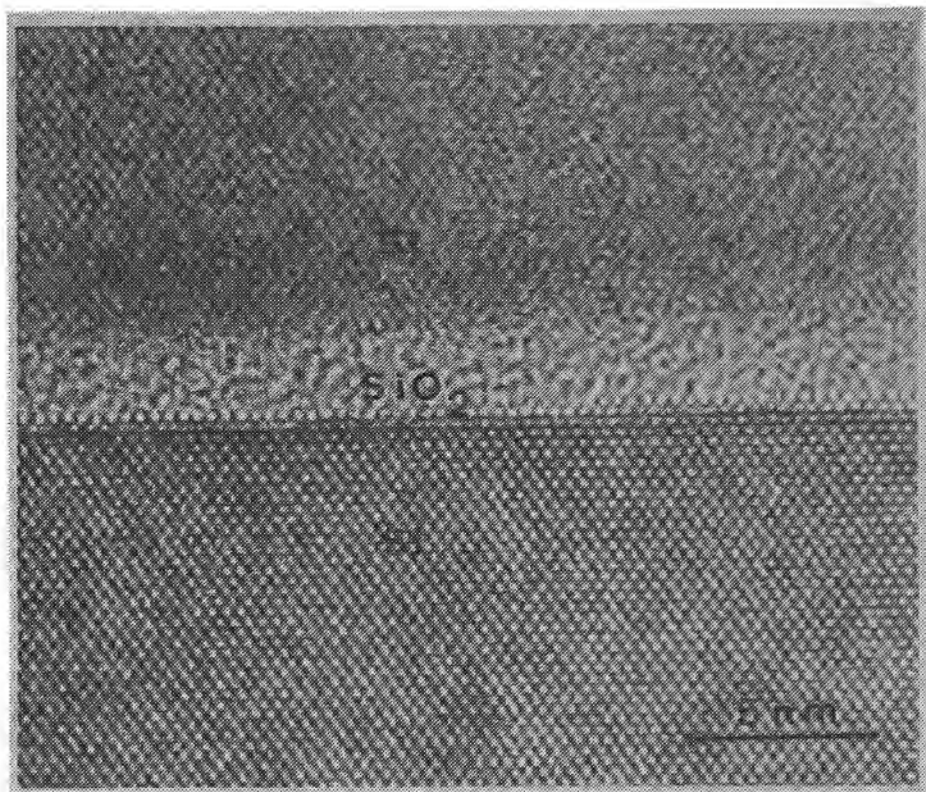


Figure 2.4: Cross-section HREM image of the bonding area between two wafers. The 2.1 nm native oxide can be measured very accurately. Steps on the silicon substrate are also observed.

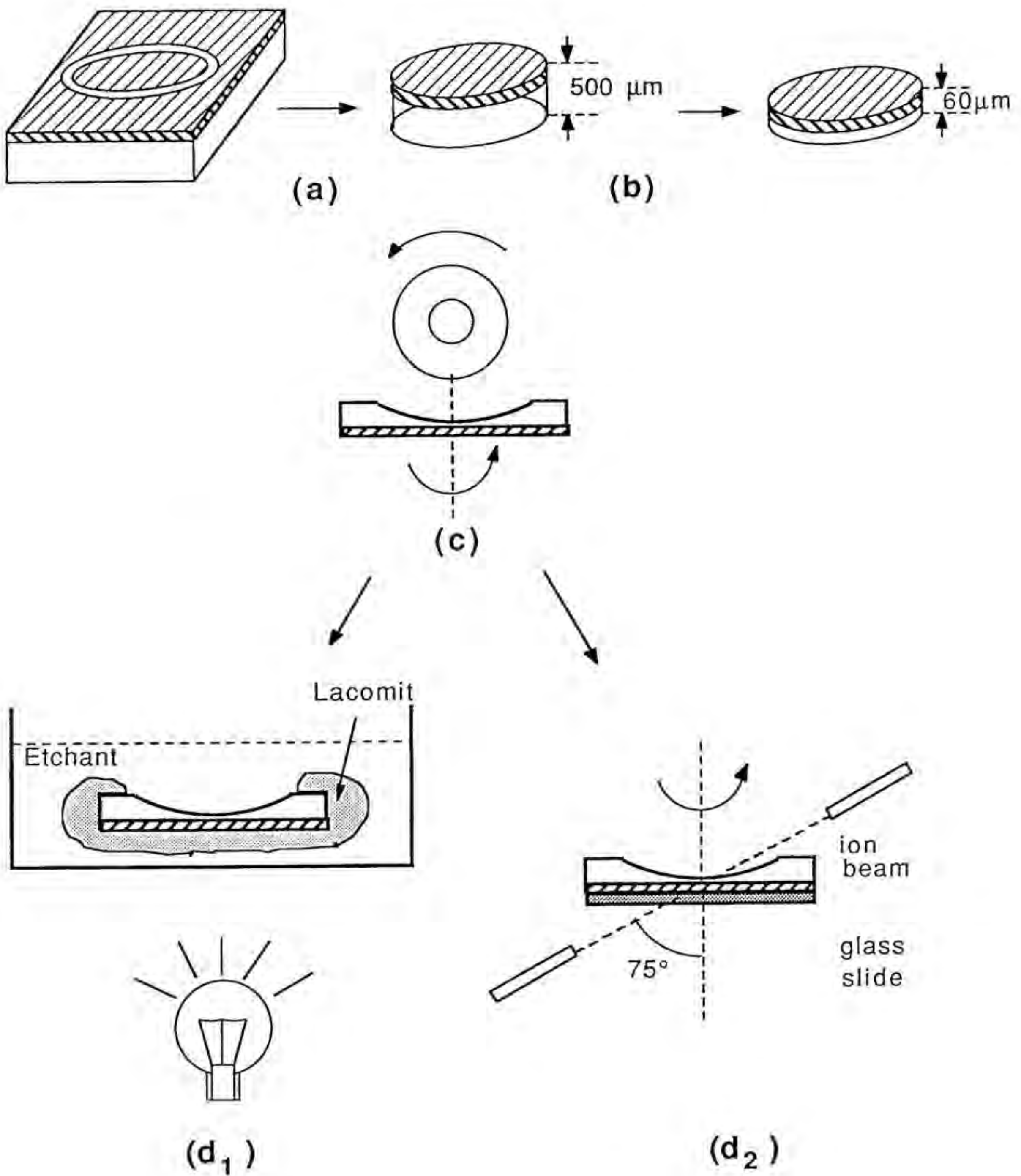


Figure 2.5: Schematic representation of the plan view preparation procedure: a) Ultrasonic cutting; b) flat grinding and polishing; c) dimpling; d_1) chemical thinning; d_2) one sided ion milling.

2.2.1 Cutting a disc

For this purpose the Gatan Ultrasonic Disc Cutter is used and the piece of wafer from which a specimen will be made is glued on a heavy flat metal base.

First the zero setting has to be adjusted with the metal base under the cutting tool. Then the piece of wafer is glued with wax to the metal base in the same way as indicated for the cross-section and with the active face against it for protection. In special cases where a specific region of the wafer has to be investigated the piece of wafer will be glued with the active face on top, increasing however the risk of damaging the surface during drilling and subsequent removal. The base with the wafer is located under the cutting tool and some abrasive powder solved in water is put on the region of the wafer that has to be cut. The cutting tool is lowered until the appropriate pressure is obtained and the ultrasounds connected. In approximately 10 seconds a disc of 2.3 or 3 mm diameter is cut from a 500 μm thick wafer. This procedure can be repeated for more discs on the same wafer without dismounting it, only displacing the base and locating another part of the wafer under the cutting tool and it takes in total about 10 minutes.

To unglue the disc the same procedure as explained for the cross-section specimens is used. Cleaning with chloroform to remove the remaining wax and final drying with methanol and nitrogen gas blow is used.

2.2.2 Mechanical thinning

The specimen is glued with wax first to a glass slide used for optical microscopy with the surface on which the features of interest are, against the glass slide. Then using abrasive silicon carbide paper and water the disc is hand ground to a final value of about 200 μm. This is done instead of using only the fine abrasive paper because it is much faster and the papers can be used for longer times. Also at a thickness of 200 μm the disc is still thick enough to handle it with tweezers.

Next the specimen is unglued from the glass slide, cleaned with chloroform and methanol and remounted on the support cylinder of the Gatan system, the cylinder is introduced in the grinding tool and thinned first to 70 µm and secondly to 60 µm using 40 µm and 5 µm grinding paper, respectively. Then the cylinder is put in the dimpler and the specimen dimpled to about 20 µm. All these steps are performed in the same way and with the same settings as for the cross-sections. Optionally the specimen can be glued to a copper grid to increase the rigidity. This step is performed typically in 25 minutes.

2.2.3 Final thinning

This stage can be realized in two different ways, depending on the specimen:

- Chemical thinning is used for most of the silicon samples.

- Ion milling is used only in special cases when the silicon wafer has layers of different materials deposited on it for which the etching solution used for silicon is not appropriate.

2.2.3.1 Chemical thinning

The sample is covered either with Lacomit or with selfadhesive plastic leaving the centre of the dimpled side without cover. In the second case two pieces of selfadhesive plastic of 2×2 cm^2, one of a transparent plastic and the other of an opaque one, are used. The specimen is glued to the centre of the transparent one with the active surface against it and in the opaque one a circular hole of about 1.5 mm in diameter is cut away. This plastic is then glued on the first one, putting the hole on the centre of the specimen and assuring that no liquid can penetrate between both plastics. This has to be done very carefully because the solution can attack the front side making the specimen useless. This preparation procedure is schematically represented in figure 2.5d_1, and is the same as published by De Veirman et al. [2.10].

A transparent plastic recipient is used to contain the etching solution and is put on a transparent plastic base which is illuminated from the bottom side in order to control optically the state of the specimen using transmitted light. The solution employed is $HF:HNO_3=1:8$ at room temperature when using 48%HF and 65%HNO_3. The specimen is introduced in the freshly prepared solution. When the sample begins to become red transparent, still about 10 μm of material is remaining. It is thinned further until the transmitted light turns light yellow and is then taken away from the solution and introduced in distilled water to stop the etching reaction. After this, acetone is used to dissolve the Lacomit or the glue of the selfadhesive foil. The specimen is cleaned further with trichloroethylene and finally with methanol and dried with nitrogen gas.

2.2.3.2 Ion milling

In some cases the thinning of the specimen by wet etching is not possible due to the different etching behaviour of the materials of the specimen. This is the case e.g. for gallium arsenide on silicon. Gallium arsenide is difficult to thin using the above mentioned solution because it is highly anisotropic, thus yielding poor quality specimens. Using backside ion milling to thin the dimpled plan view specimen allows however to circumvent this problem, as illustrated in figure 2.6, showing a plan view HVEM micrograph of the defects in a 3 μm thick MBE GaAs layer on Si.

After dimpling the specimen is mounted in the holder of the ion miller, protecting the active surface with a very thin glass disc, as seen in figure 2.5d_2. This is cut from a cover-glass of optical microscope using the Ultrasonic Disc Cutter. Then the specimen holder is introduced into the ion miller and milled from the backside with the same conditions as for the cross-sections. In this case the Autoterminator is extremely useful, because gallium arsenide is not transparent to light when it is to electrons. So the highest sensitivity of the Autoterminator has to be used in order to stop the milling at the moment when only a very small hole is formed. In this case the milling time takes

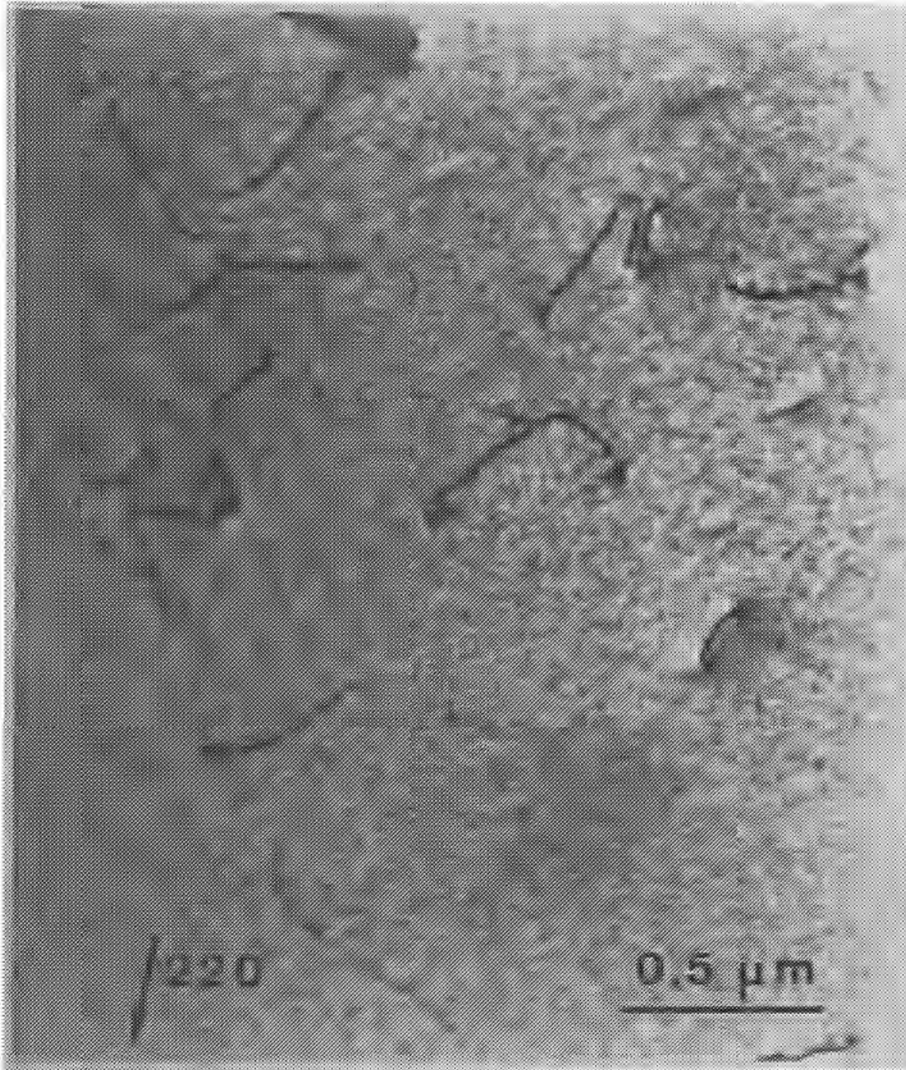


Figure 2.6: Plan view HVEM observation of defects in a 3 µm thick MBE GaAs on Si thinned with back-side ion milling.

again about 1 to 2 hours, depending on the initial thickness of the sample and on the state of the guns.

One side milling without protection of the surface is not performed because of problems with redeposition of the sputtered material at the side where no ion beam is incident. This is due to the fact that the vacuum of the Gatan ion miller with diffusion pump is not good enough to avoid this problem.

2.3 LOCAL THINNING

In the microelectronics industry and research, TEM has become one of the most powerful tools for the structural characterization and investigation of lattice defects and interface properties.

For the conventional preparation technique described in the previous section specially designed structures for TEM investigation are fabricated based on the repetition of the pattern of interest [2.5]. This approach allows to have a high probability of reaching that structure when preparing the TEM specimen using the conventional technique. A serious drawback are the special masks which have to be designed for these test structures. This approach is only possible for the development and optimization of a process. Mostly, when a problem appears in the production line, no such structures are available and the problem has to be solved on-line. In this case the final product has to be investigated itself and a prespecified region has to be reached, for example, when certain transistors have substantially degraded electrical characteristics.

2.3.1 Cross-section preparation

The cross-section preparation technique is basically the same as represented in figure 2.1, but some minor changes have been introduced. For this procedure the cross-section stack will be formed by a stripe containing the structure to be studied and a glass stripe of approximately the same size. This will allow to look through the glass stripe during the thinning procedure,

so that the position of the feature of interest can be always controlled. Also the prethinning is performed in more steps and it is necessary to be extremely careful during preparation. Finally, one side ion milling with lower settings is used to make the milling more controllable.

2.3.1.1 Sawing of the wafer

The sawing of the wafer and of the glass plate is performed in the same way as described in section 2.1.1. Rectangular stripes of the same size are cut from the integrated circuit, locating the device of interest at 100 μm from one of the long edges of the rectangle and also displaced 100 μm from the centre of the same long side. This has been found to be a good compromise between the time required for the preparation and the reproducibility. The structure can also be located further away from the centre, but then longer ion milling times are required and also the slope of the specimen increases, leading to a smaller transparent area [2.16]. An example of a cut structure can be seen in figures 2.7a and c. The arrow indicates the transistor which will be used throughout this section to illustrate the thinning procedure. The transistor dimensions are 10 μm long by 0.6 μm wide. At its left, other transistors of the same length but with widths ranging from 0.7 to 1.5 μm can be seen.

2.3.1.2 Glueing of the stripes

A stripe of the wafer containing the desired structure and a glass stripe of the same size are glued together using M-610 Bond. Figure 2.7b shows the top view of the glued stack looking through the glass stripe, illustrating that the region with the structure to be studied is still visible.

The main advantages of this type of glue are: i) that it is transparent to light, ii) that it forms a thin layer of about 1 μm between the two stripes when they are gently pressed together using tweezers, iii) that the same glue can be used for up to three months without loosing glueing power, iv) that the glue is not solved when taking off the sample of the prethinning holder, and, v) that the etching rate in the ion miller is quite close to the one of silicon (when using the settings described in this

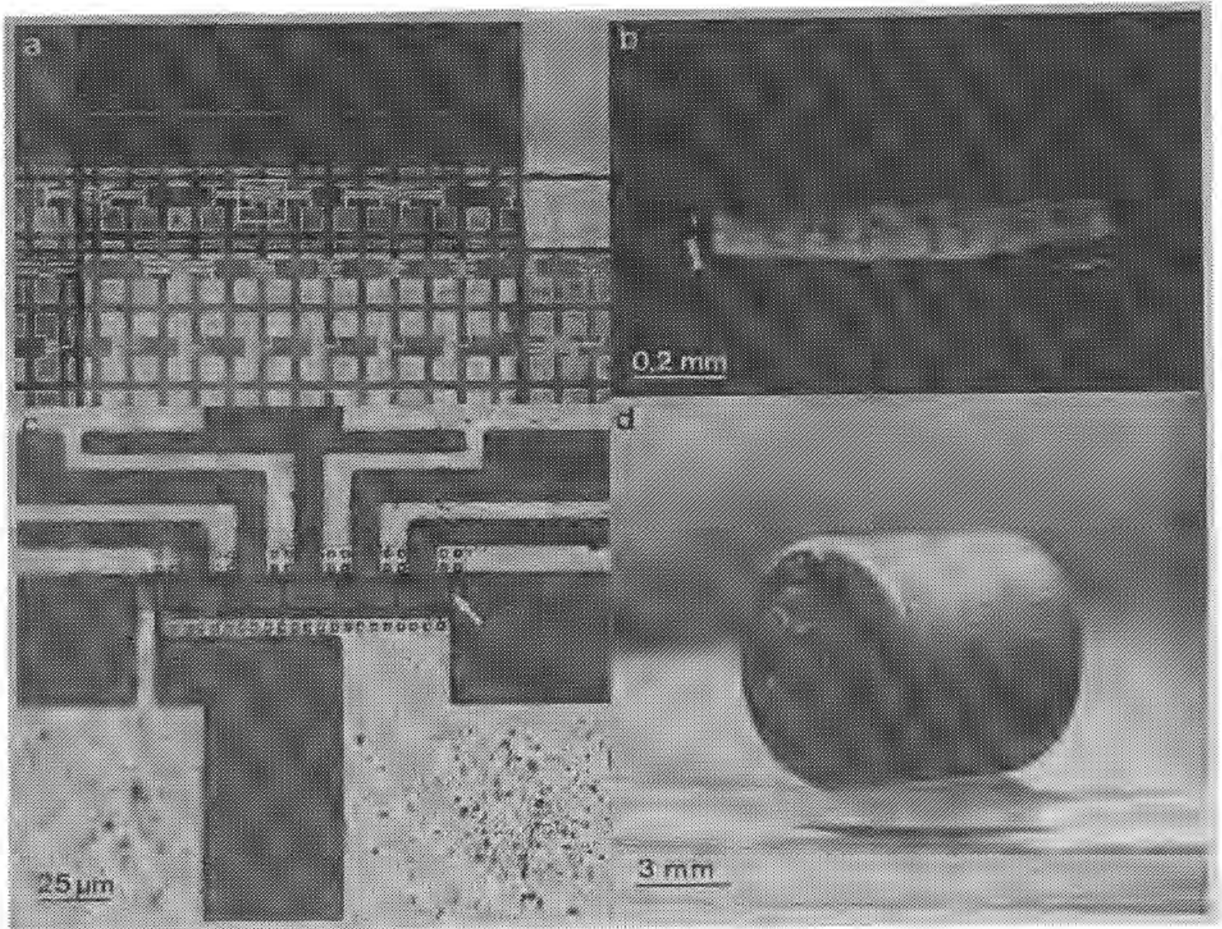


Figure 2.7: a) Cutting of the wafer; b) observation of the desired structure through the glass stripe (top view); c) magnification of the box of figure a, showing the feature which has to be investigated (arrow): a transistor 10 µm long by 0.6 µm wide; d) cross-section stack glued eccentrically to the cylinder in order to perform the first prethinning.

section). The quality of the glass stripe is also very important in the described preparation technique because of two reasons: first, it allows to observe the structure during the mechanical prethinning of the specimen and secondly, it also protects the active surface against undesired damage during thinning.

2.3.1.3 Mechanical prethinning and dimpling

The stack is set in the furnace at 150°C for 15 minutes. Once taken out of the furnace, the cross-section stack is glued to the support cylinder on one of the flat sides but very close to the perimeter of the cylinder and with the glass stripe closest to the perimeter, as can be seen in figure 2.7d. This approach allows to inspect the structure during the next mechanical prethinning steps using an optical microscope. The stack has to be glued in such a way that the side of the stack which is closer to the structure to be investigated is the one which will be ground first, i.e., is not glued to the cylinder (this means that the specimen of figure 2.7b will be glued to the cylinder with the side which is at the bottom of the figure against it). Once the whole cross-section stack-support cylinder is cooled to room temperature, it is introduced into the Disc Grinder to perform the prethinning.

The grinding procedure starts using, in order, abrasive paper of 40 μm, 5 μm and 0.3 μm grain size. First the rougher paper is used for grinding in 10 μm steps until the surface of the silicon stripe starts to be ground, so that the sawing damage has been partly eliminated. Then the specimen is taken out of the Grinder and put under the optical microscope (400 times magnification), so that the distance between the structure and the polished surface can be measured. For this the cylinder is placed under the microscope in the same way as it appears in figure 2.7d, so that the observation is done through the glass stripe. This gives a rough estimate of the amount of material that has to be removed. Next the whole cylinder-stack is put again in the Grinder in the same position as it was before (this is important because the cylinder is not always completely flat) and the grinding procedure is continued with the same paper and

following with the same procedure. Regularly the cylinder has to be inspected in the microscope. When the distance between the structure and the ground surface is about 20 μm the grinding is stopped and the 5 μm abrasive paper is used to continue. Now the amount of material removed during each step is of the order of 2 to 3 μm, until the distance is about 5 μm. Finally the finer paper is used to grind about 4 μm in two steps. The effect of the different abrasive papers on the surface quality of the cross-section is illustrated in figure 2.8, where a top and a cross-section views through an optical microscope are presented.

Because the finest abrasive still leaves some scratches, an additional polishing procedure has to be used in order to obtain a mirror like surface. This is performed with Syton. In our experience polishing with the Grinder and using Syton on a felt cloth leads to the destruction of the specimen. Probably the pressure which is applied to the specimen during the polishing is too high and difficult to control. However the dimpler allows to control the pressure quite accurately.

Next the specimen support with the ground specimen is taken out of the Disc Grinder and introduced in the dimpler. The specimen is carefully centred with respect to the rotation axis of the base of the dimpler. The polishing wheel with a felt stripe is used for about 10 minutes using Syton as polishing agent. The settings of the dimpler are the same as mentioned in section 2.1.4. This treatment removes practically all the surface damage introduced during the previous steps and also a few microns of material in the central part of the cross-section. The outer rim of the specimen is not well polished, but this is not important because only the very central part will be used for TEM investigation.

After this the cylinder is taken out and observed again in the optical microscope to be sure that the surface damage has been totally removed. The optical cross-section shows clearly the structure of interest, as can be seen in figure 2.8. Also looking through the glass plate allows to see if the structure has been reached. If scratches are still present, further polishing is used to obtain a mirror like surface.

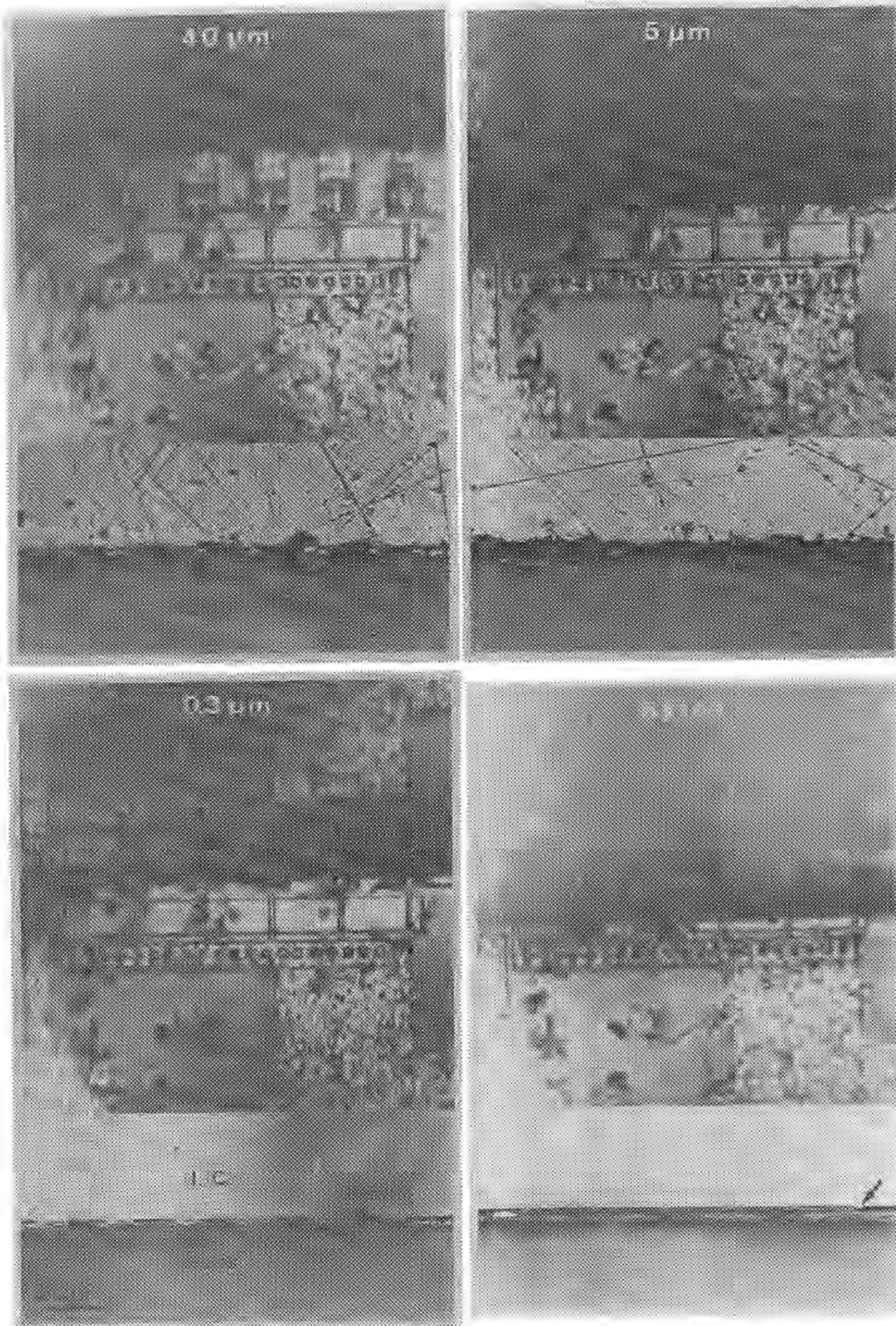


Figure 2.8: Optical top and cross-section views of the effect of the different prethinning steps: after the use of 40 µm, 5 µm and 0.3 µm abrasive paper and after polishing using Syton. Remark that the structure can be seen reflected in the glass stripe when the polishing system becomes finer. In the figure showing the effect of Syton the polishing has not been performed in the centre of the stripe and due to multiple reflection in the glass stripe it is not possible to see if the area of interest has been reached. This is clarified in the cross-section below. The 0.6 µm transistor is indicated with an arrow.

It is very important to know the topography of the cross-section (i.e., the different layers that form the specimen, the location of the structure that has to be studied related to certain indicative features, like contacts) because in the following steps it will be impossible to inspect the structure through the glass stripe, as dimpling will reduce its thickness and give rise to a spherical section where multiple reflection of the light will take place. Also the position of the structure related to the surface of the cross-section has to be measured.

All the mechanical prethinning steps have to be carried out very carefully, because otherwise pieces of the glass slide can be abraded and it will become impossible to look through it. For this reason always new abrasive paper has to be used for each specimen, the movements of the Grinder have to be very gentle and no pressure should be applied.

Next the specimen is unglued, turned around and mechanically thinned from the other side in the same way as indicated in section 2.1.3. Dimpling is also performed following the indications of section 2.1.4. Figure 2.9 shows the aspect of the cross-section stack after dimpling using the diamond compound (a) and after polishing with Syton (b). It is easily seen that practically all damage has been removed. If some small residual damage is still present it will be removed during the subsequent ion milling step and it will not affect the quality of the final TEM specimen. If the residual damage is important, it will be enhanced during ion milling, leading in a bad specimen.

2.3.1.4 Ion milling

The specimen is first unglued from the cylinder, cleaned thoroughly with chloroform and methanol, glued to a copper grid using again M-610 Bond and cured for 10 minutes at 150°C.

On the specimen holder first a very thin glass disc (cut from a cover-glass of optical microscope using the Gatan Ultrasonic Disc Cutter) is placed and on top of it the grid with the cross-section. Next the cover of the holder is placed. The glass disc protects the first polished surface against the rede-

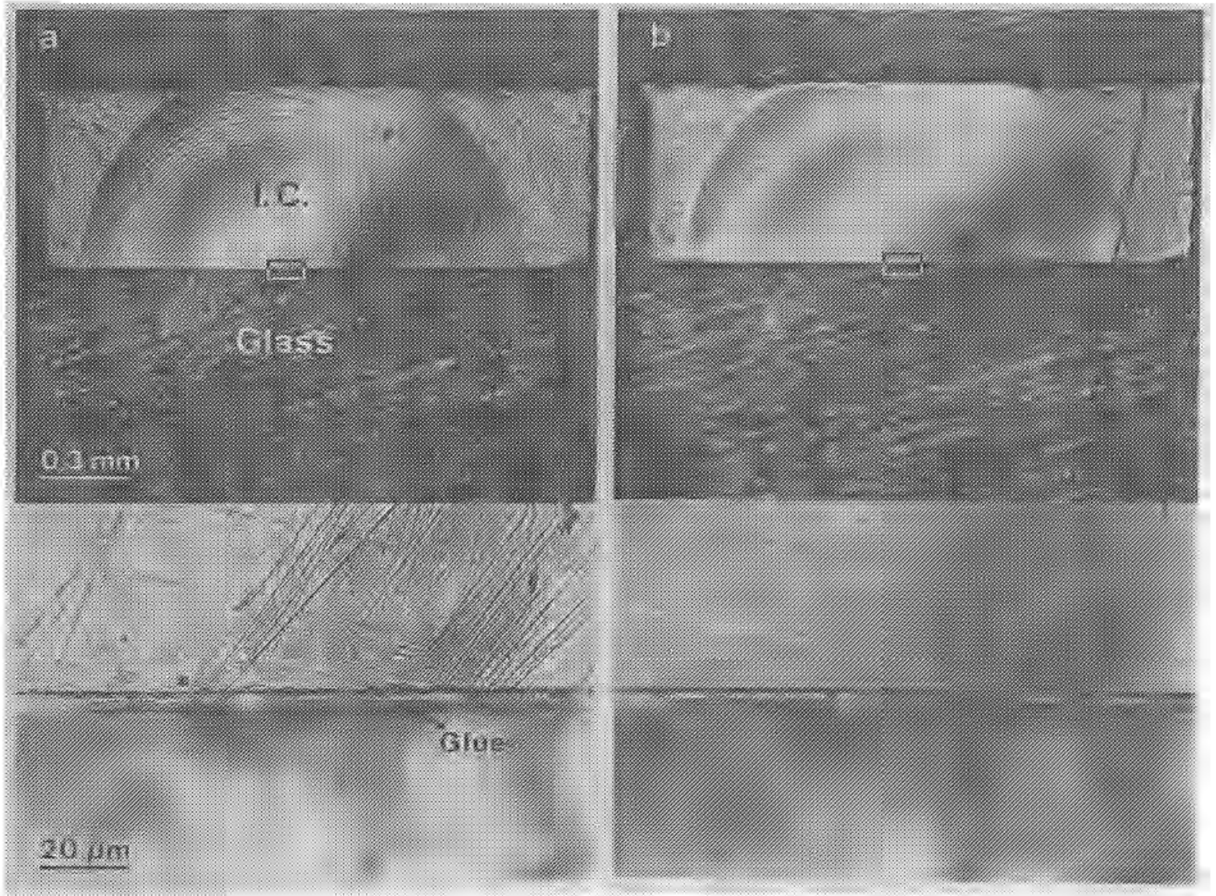


Figure 2.9: Optical top and cross-section views of the dimpled specimen after a) dimpling with the 3 μm diamond compound and b) polishing using Syton.

position of sputtered material (see section 2.2.3.2 for one sided ion milling).

Next the specimen is ion milled from one side with the settings 0.3 mA, 6 kV, an angle of incidence of 75° with the normal to the surface of the specimen and using argon ions. Typical times are about 3.5 to 4 hours. The specimen has to be controlled optically regularly to assess that no hole or only a small one has been formed. The Autoterminator can not be used as the glass stripe used for the cross-section is transparent to the red light of the laser. Next the holder is taken out of the ion miller, put under the optical microscope and the distance between the edge of the hole and the structure to be investigated and the thickness at the position of the structure measured (of course, this is done in cross-section). Based on the colour of the transmitted light through the specimen one can deduce quite accurately the thickness of the specimen at the position of the structure of interest. If the area is still quite thick (more than 3 μm), further milling with the same settings is performed, taking into account that the etching rate is about 70 nm/min in our ion miller. When the specimen is thinner than 3 μm, the thickness has to be measured accurately in order to decide the time that the specimen will be milled further. For this the interference fringes obtained with the optical microscope are very useful. By comparing their position with the one of the thickness fringes obtained from a TEM picture of the same area, it can be deduced that each interference fringe corresponds, approximately, to a thickness of 125 nm.

Next the glass plate is removed and milling from both sides is performed, keeping the same settings. The milling time is estimated by dividing the approximate thickness by the etching rate. Its value is about 150 nm/min because both guns are used, so that the etching rate is approximately double of the one corresponding to one gun. After reaching the neighbourhood of the area of interest, further iterative milling with the settings 0.1 mA and 5 kV is performed. The etching rate for these low settings is about 25 nm/min. This allows to well control the etching process. Two sided ion milling is required to eliminate

all remains of damage of the first polished side of the specimen.

2.3.1.5 Finished specimen

Figure 2.10 shows the aspect of the finished specimen. The arrow indicates the position of the 0.6 μm transistor. The figure shows the same area as the bottom right of figure 2.8, but now observed from the opposite surface and rotated 180°. This picture has been taken using reflected light. Some dark lines can be seen below the transistor, which correspond to interference of the optical light when crossing the silicon substrate.

Figure 2.11a is the TEM image corresponding to the optical picture of figure 2.10. The different layers which form the transistor can be seen. HREM images are possible in the edge of the specimen.

Figures 2.11b to d show the evolution of the specimen after further ion milling for 3, 6 and 9 minutes with the settings 0.1 mA and 5 kV. This illustrates that a scan over the complete device can be performed, allowing HREM imaging of the specimen in different areas.

2.3.2 Plan view preparation

The procedure to prepare plan view specimens of prespecified regions on a silicon substrate is mainly the same as indicated in section 2.3 and using always ion milling for the final thinning.

In this procedure it is necessary, similarly to the cross-sections, that some structures are present, allowing to recognize the area to be studied. These features need to have a size of at least 5 μm long by 0.6 μm wide, so that they can be recognized during the preparation.

The most important difference to the conventional plan view preparation is that the structure to be studied has to be located close to the centre of the disc that will be cut. A good compromise is that the structure is displaced about 200 μm from the centre. This is difficult to achieve using the Ultrasonic

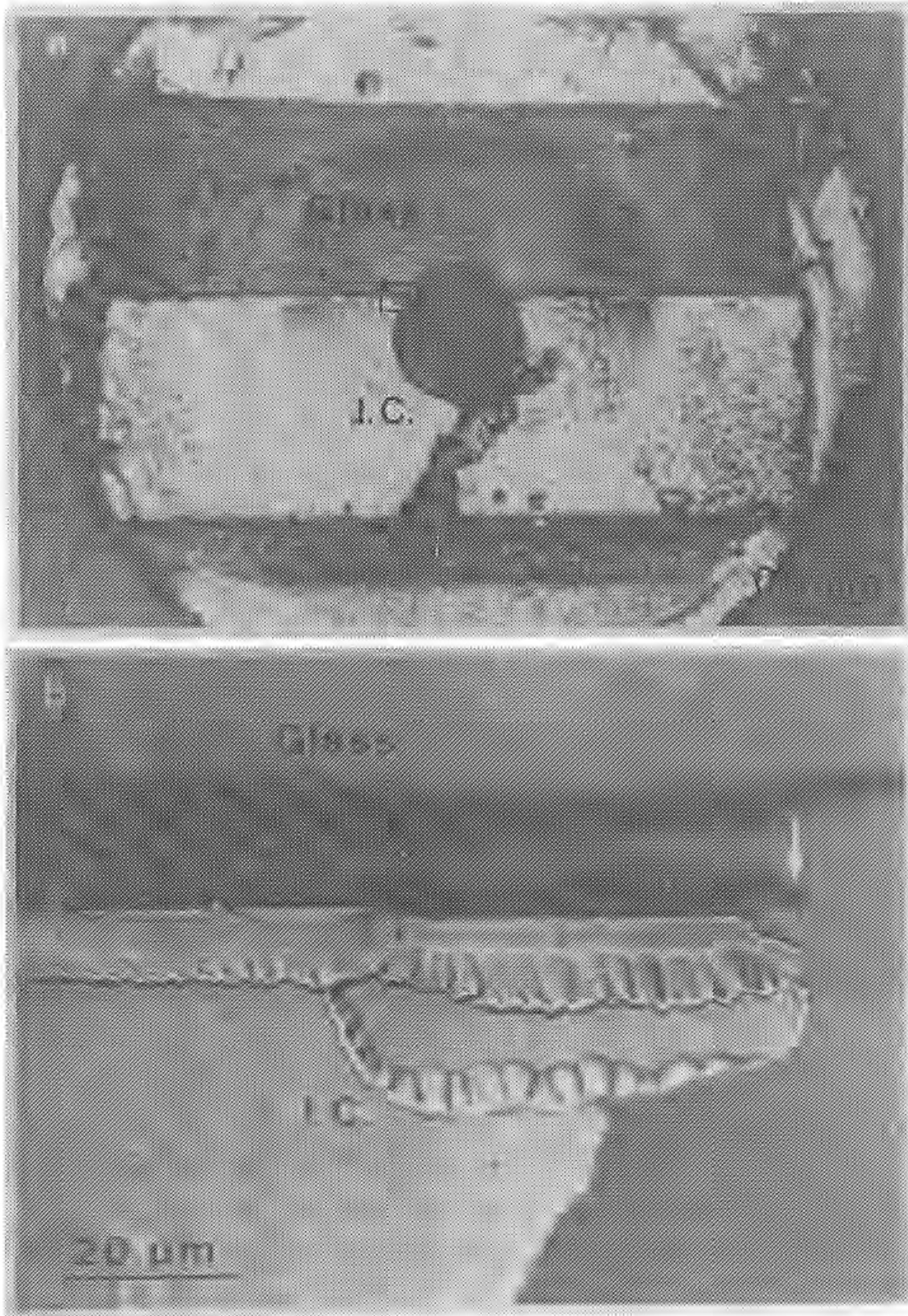


Figure 2.10: a) Finished cross-section specimen; b) enlargement of the area of figure a, showing the structure to be studied (arrow). The peculiar shape below the structure of interest is due to the shadowing effect of the different layers which form the transistor.

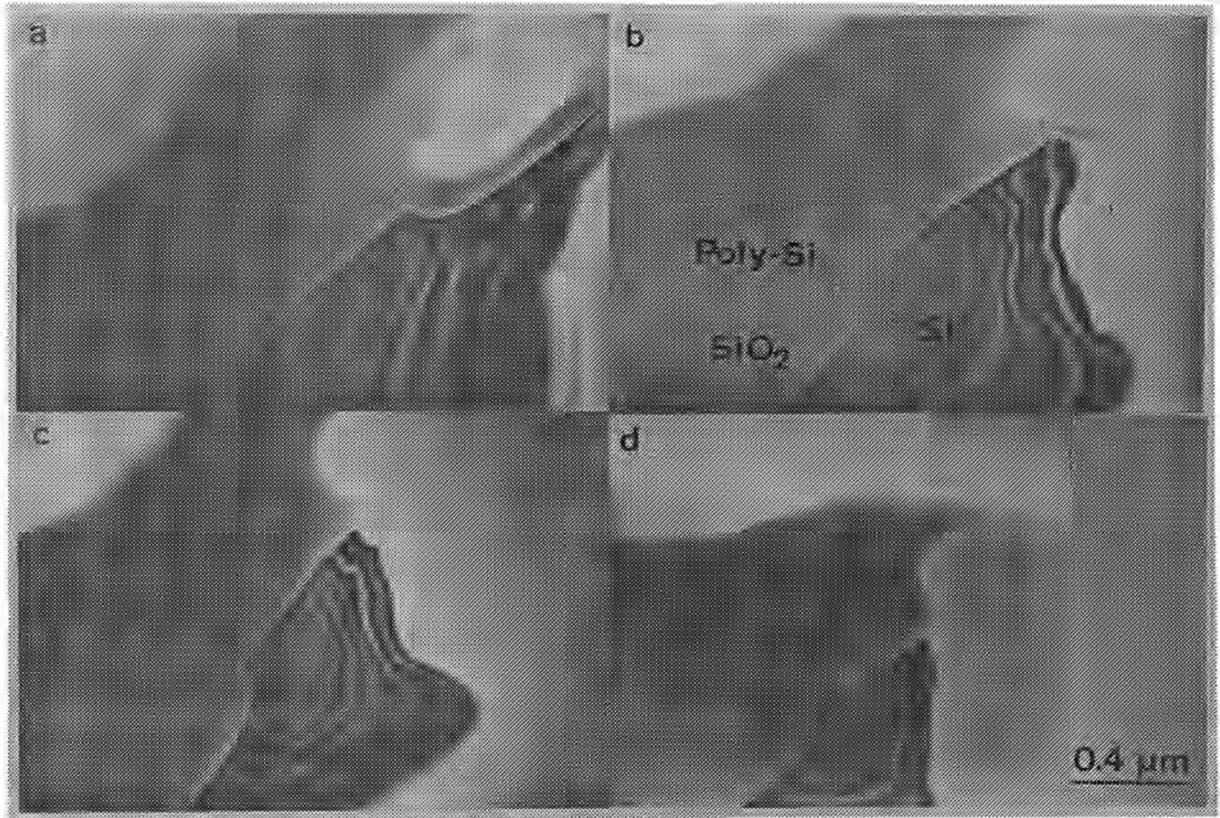


Figure 2.11: TEM observations: a) specimen of figure 2.10; b) same area after 3 minutes ion milling using the settings 0.1 mA and 5 kV; c) after 6 minutes; d) after 9 minutes.

Disc Cutter from Gatan, although a microscope is supplied which allows to view which will be the area that will be cut.

The mechanical prethinning step is performed exactly in the same way as for a normal plan view specimen. Also the ion milling is performed in the same way, using the same settings and using the Autoterminator to stop the milling. In this case the Autoterminator is set at its maximum sensitivity in order to stop the milling just when a small hole is formed. Normally the Autoterminator stops later, because the deposition of sputtered material on the glass plate makes it less transparent and a larger hole is obtained. After the hole is formed, the specimen is observed with an optical microscope, the thickness near the structure estimated and, as before, the specimen is milled further from one side using lower settings (0.3 mA and 6 kV or 0.1 mA and 5 kV) with periodic inspection to monitor the progress of the thinning process.

2.4 OTHER SEMICONDUCTOR MATERIALS

As indicated before this preparation technique with ion milling can be applied, with small modifications, to the preparation of "difficult" semiconductor materials, as for example, InP and ternary and quaternary compounds containing In and P. It is known that these materials, when ion milled using conventional milling techniques, i.e., using argon ions and working at room temperature, strongly develop indium islands due to the evaporation of phosphorus. Normally this problem is circumvented using iodine ions [2.18] instead of argon. But not all laboratories have yet this option and also the contamination of the milling chamber is quite important. The use of a cooled milling stage can reduce strongly the evaporation of phosphorus.

An additional difficulty of these materials is that they are quite fragile, so that special care has to be taken during the mechanical prethinning.

The technique described above can also be used for preparing these specimens, only introducing some small changes in the

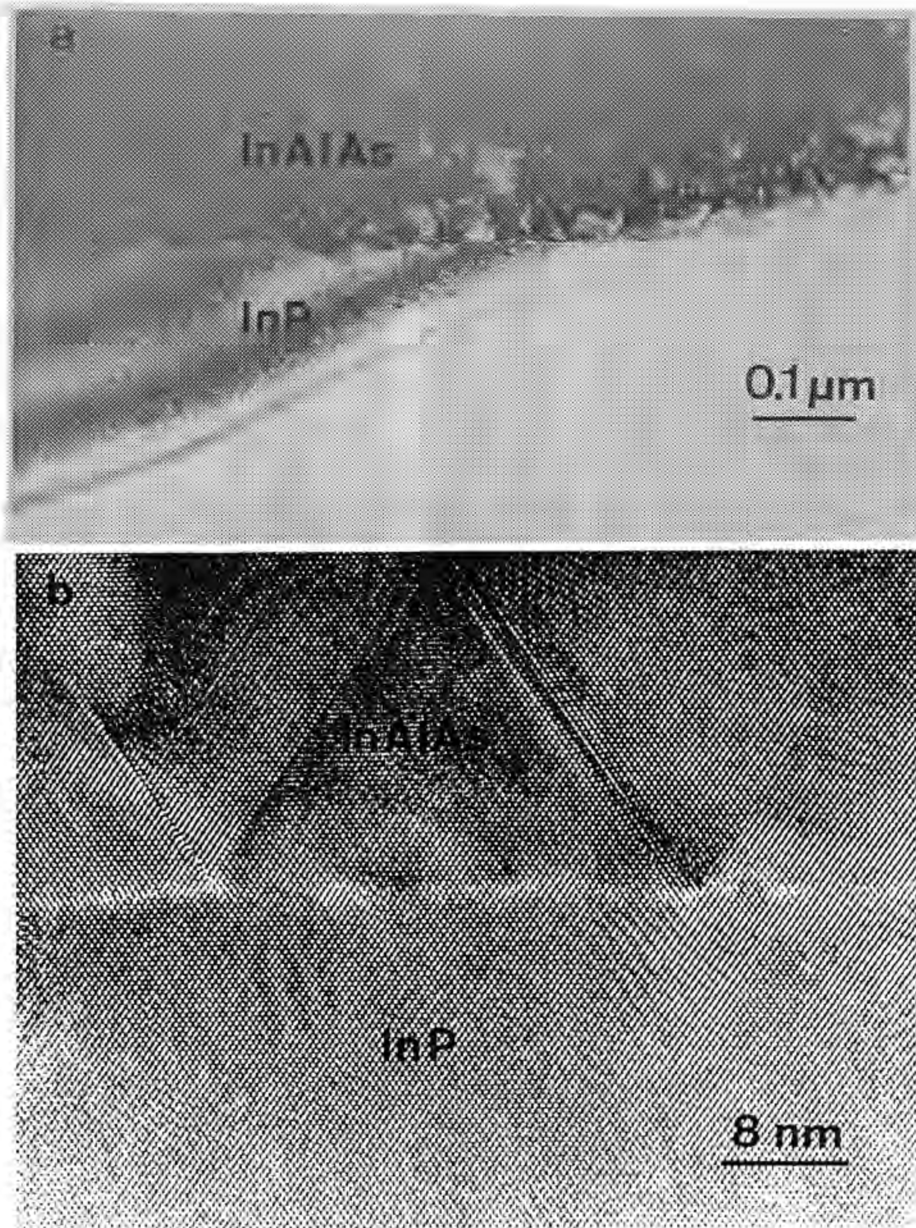


Figure 2.12: InGaAs grown by MBE on an InP substrate. a) Low magnification image of the interface and b) HREM of the same area, showing three stacking faults.

cross-section procedure (for the plan view there are not so much problems).

These changes are summarized below:

- The steps used for the flat grinding are reduced to 10 μm when using the paper with a grain size of 40 μm. The grinding of the second side of the cross-section with the 40 μm paper is stopped at 110 μm and with the 5 μm paper, at 100 μm.
- During the dimpling the weight applied to the dimpling wheel is reduced to 25 g and dimpled to a thickness of 35 μm. For the polishing the weight is also reduced, now to 20 g.
- Finally the ion milling is performed using a cooling stage and the current and the tension lowered to 0.2 mA and 4 kV, respectively. In this case the Autoterminator is very useful, because these materials are not transparent to the red light and also because the milling time increases to about 7 hours.

Figure 2.12 illustrates the power of this technique. The upper image corresponds to a low magnification of the interface between a 2 μm thick MBE grown InGaAs layer on top of an InP substrate. As can be seen from this image the top layer has been milled slower than the substrate and small dark dots on the substrate occur, which correspond to indium islands. The lower image is a HREM image of the same area, showing clearly three stacking faults generated at the interface during the growth of the layer. Just below this faults a small indium island is visible but it does not affect the quality of the image.

2.5 CONCLUSIONS

A fast preparation technique for high quality cross-section and plan view TEM specimens of semiconducting materials has been presented. The whole procedure can be performed in less than 4 hours, allowing in urgent cases preparation and observation of semiconducting materials to be realized in one day. This ability is highly important in the microelectronics industry, where the

results of the TEM characterisation are often urgently needed by the process engineers in case of sudden problems with the processing. A further advantage of the presented preparation procedure is that no special equipment, except for the commonly available apparatus used in TEM laboratories, is required.

A second major advantage of this technique is the use of a fine blade diamond saw, which allows to cut the wafer in any desired direction when the observation of the specimens has to be realized in specific crystallographic directions.

Furthermore this preparation technique has been implemented in order to thin localized regions of semiconductor devices with a feature size of 10 μm long by 0.6 μm wide. This procedure can be performed in about 6 hours and is important when a problem appears in the production line and has to be solved on-line without having to prepare special test structures.

Finally, the same procedure can be used to thin other semiconducting materials which are normally difficult, as InP and related compounds. The only difference is that a cooling holder for the ion milling step is used to avoid the formation of indium islands due to the evaporation of the phosphorus atoms. Although the procedure is much slower, it is not necessary to have an ion miller with iodine as milling specie and the quality of the specimens is reasonably.

REFERENCES CHAPTER 2

[**2.1**] T. Boone and S. Nakahara, Mater. Res. Soc. Symp. Proc. 115, 81 (1988).

[**2.2**] C.J.D. Hetherington, Mater. Res. Soc. Symp. Proc. 115, 143 (1988).

[**2.3**] J.C. Bravman and R. Sinclair, J. Electron Microsc. Techn. 1, 53 (1984).

[**2.4**] M.S. Abrahams and C.J. Buiocchi, J. Appl. Phys. 45, 3315 (1974).

[**2.5**] R.B. Marcus and T.T. Sheng, Transmission Electron Microscopy of Silicon VLSI Circuits and Structures, Wiley-Interscience, 1983.

[**2.6**] A. Romano, J. Vanhellemont, H. Bender and J.R. Morante, Ultramicroscopy 31, 183 (1989).

[**2.7**] J. Vanhellemont, H. Bender, C. Claeys, J. Van Landuyt, G. Declerck, S. Amelinckx and R. Van Overstraeten, Ultramicroscopy 11, 303 (1983).

[**2.8**] J. Vanhellemont, H. Bender and L. Rossou, Mater. Res. Soc. Symp. Proc. 115, 247 (1988).

[**2.9**] A. Romano, J. Vanhellemont and H. Bender, Mater. Res. Soc. Symp. Proc. 199, 167 (1990).

[**2.10**] A. De Veirman, J. Eysermans, H. Bender, J. Vanhellemont and J. Van Landuyt, Mater. Res. Soc. Symp. Proc. 115, 241 (1988).

[**2.11**] B.O. Kolbesen, K.R. Mayer and G.E. Schuh, J. Phys. E 8, 197 (1975).

[**2.12**] M.B. Ellington, Mater. Res. Soc. Symp. Proc. 115, 265 (1988).

[**2.13**] R. Anderson, S. Klepeis, J. Benedict, W.G. Vandygrift and M. Orndorff, Inst. Phys. Conf. Ser. 100, 491 (1989).

[2.14] J.T. Wetzel and D.A. Danner, Mater. Res. Soc. Symp. Proc. 115, 253 (1987).

[2.15] E.C.G. Kirk, D.A. Williams and H. Ahmed, Inst. Phys. Conf. Ser. 100, 501 (1989).

[2.16] M.C. Madden and P. Crafard, J. Electron Microsc. Techn. 11, 161 (1989).

[2.17] H. Bender, Inst. Phys. Conf. Ser. 76, 17 (1985).

[2.18] A.G. Cullis and N.G. Chew, Appl. Phys. Lett. 44, 142 (1984).

3. ISOLATION TECHNIQUES

3.0 INTRODUCTION

Device isolation in integrated circuits is required to operate the devices so that they only interact with each other in a well controlled manner through the designed interconnects. Therefore it is desirable to eliminate or, when not possible, to limit to the minimum the leakage current between the devices, because they degrade their performances.

Since the early 1970's extensive research has been performed in order to develop appropriate isolation schemes. Device isolation techniques are a key issue, since they have become one of the major facts limiting the device packing density in VLSI circuits. For this reason it is desirable to minimize the surface of the wafer that is used for isolation, still maintaining sufficiently good isolation performances.

For the development of good isolation schemes for their use in ULSI technology, when the geometries scale down, a tight requirement on the analytical techniques is imposed. Especially for submicron devices the dimensions of the components become so small that only cross-sectional TEM has sufficient resolution to study the structural aspects in full detail. More and more the application of HREM is becoming an indispensable routine tool for process development.

LOCOS (LOCal Oxidation of Silicon) [3.1] is the major isolation technique employed in the microelectronics industry, due to its simplicity. For submicron devices LOCOS cannot be used and other isolation techniques have to be developed. In this decade extensive efforts have been devoted to this topic and different methods have been found, mostly involving quite complicated designs, including several mask steps. Modifications of LOCOS have also been tested and some of them seem, still, to be useful for submicron CMOS technology.

In this chapter we will present a study of the geometry and defect generation corresponding to some of the modifications of the LOCOS technique that give rise to a shorter bird's beak (BB). TEM has been used, preparing the specimens both in plan view and cross-section. Sometimes other techniques have been employed, which give complementary information. These techniques are RAMAN microprobe spectroscopy and Scanning Electron Microscopy (SEM).

In section 3.1 the experimental conditions employed for this study are presented. In the second section a brief introduction to the LOCOS process is presented, showing which are the main features of the technique, its advantages and drawbacks. In sections 3.3 and 3.4 modifications of LOCOS are presented, which have been developed in order to circumvent its specific problems. The improved steps will be commented, as well as the main results which can be obtained, their drawbacks and limitations.

In the last section the formation of LOCOS structures on SIMOX substrates is discussed. For thin film SIMOX the presence of the buried amorphous layer modifies slightly the behaviour of the silicon top layer towards the oxidation.

3.1 EXPERIMENTAL DETAILS

The starting wafers for sections 3.2 to 3.4 are 4 and 5 inch Czochralski grown, (001) oriented p-type silicon wafers, with a resistivity between 16 and 40 Ωcm. Most of the wafers have been chosen so that the interstitial oxygen content was of an intermediate value, about $5\cdot10^{17}$ cm^{-3}. This is important because it has been reported that for oxygen concentrations above $8\cdot10^{17}$ cm^{-3} the probability of film edge induced dislocation generation increases significantly [3.2].

For the experiment presented in section 3.3.2, 6 inch wafers were employed having, again, a medium oxygen content, in the range of $6-7.5\cdot10^{17}$ cm^{-3}.

For the study described in section 3.4.1 the wafers were covered by an epitaxial silicon layer about 11 μm thick on top. In this way the oxygen content in the area where devices should be fabricated will be lower than the above mentioned.

For the last section, where the formation of LOCOS structures on SIMOX substrates is presented, the material used were commercial 4 inch SIMOX wafers, supplied by IBIS Co.

The fabrication of the structures involving these two last types of substrates were performed at the Swedish Institute of Microelectronics.

A test structure consisting of parallel lines with a variable width and separation is used. This allows to study the influence of the oxidation and oxidation mask parameters, as well as the influence of the width and spacing of the lines on the final geometry and on the defect generation in the silicon substrate. The lines are oriented along [110] directions, as is the commonly used in integrated circuits processing.

Different layer thickness combinations have been used. The field oxidation temperature was varied between 900 and 1050°C. The field oxide (FOX) thickness was usually 600 nm, which is the value employed commonly for submicron CMOS processing. Other FOX thicknesses were also studied in order to see the different steps of the oxidation process and the feasibility of its use for smaller device sizes. Especially in the case of SIMOX wafers, a much thinner field oxide can be employed, because the top silicon layer is quite thin and it is only necessary to oxidize it down to the buried oxide layer in order to achieve full isolation of the devices.

3.2 LOCAL OXIDATION OF SILICON

3.2.1 Principles of the process

Since the first publication by Appels et al. [3.1] in 1970 on the use of Si_3N_4 for selective oxidation processes to isolate

the electrically active device regions, the LOCOS technique has been studied extensively and is nowadays the most commonly used isolation technique for bipolar and MOS technology. A very exhaustive study of the LOCOS technique can be found in reference [3.3]. The main steps of this process are represented schematically in figure 3.1, and are listed below:

a) Growth of a thin silicon oxide layer, the "pad oxide", over the complete wafer.

b) Deposition of a layer of Si_3N_4 by low pressure chemical vapour deposition (LPCVD), which constitutes the oxidation mask. It will prevent the selected areas of the silicon surface to be oxidized.

c) Deposition of the photoresist.

d) Definition, using photolithography, of the areas where the FOX has to be grown.

e) Removal of the photoresist and of the underlying nitride in the exposed areas, followed by removal of the unexposed photoresist.

f) Oxidation of the opened windows in wet (containing water vapour) or dry ambient, typically at temperatures between 850 and 1150°C. The thickness of the grown field oxide ranges between 400 nm and 1 μm.

If the wafers have to be further processed, the nitride and pad oxide need to be chemically removed after the field oxidation. If cross-sectional TEM or SEM analyses have to be performed, after step f) it is convenient to deposit a layer that enhances the image contrast at the wafer surface and that, at the same time, protects the surface of the wafer. Mostly polysilicon is used or some other contrasting layer which can easily be deposited.

For plan view investigations of substrate defect all the layers on top of the substrate are removed. In this way no diffuse scattering of the electrons will take place, giving a better image quality.

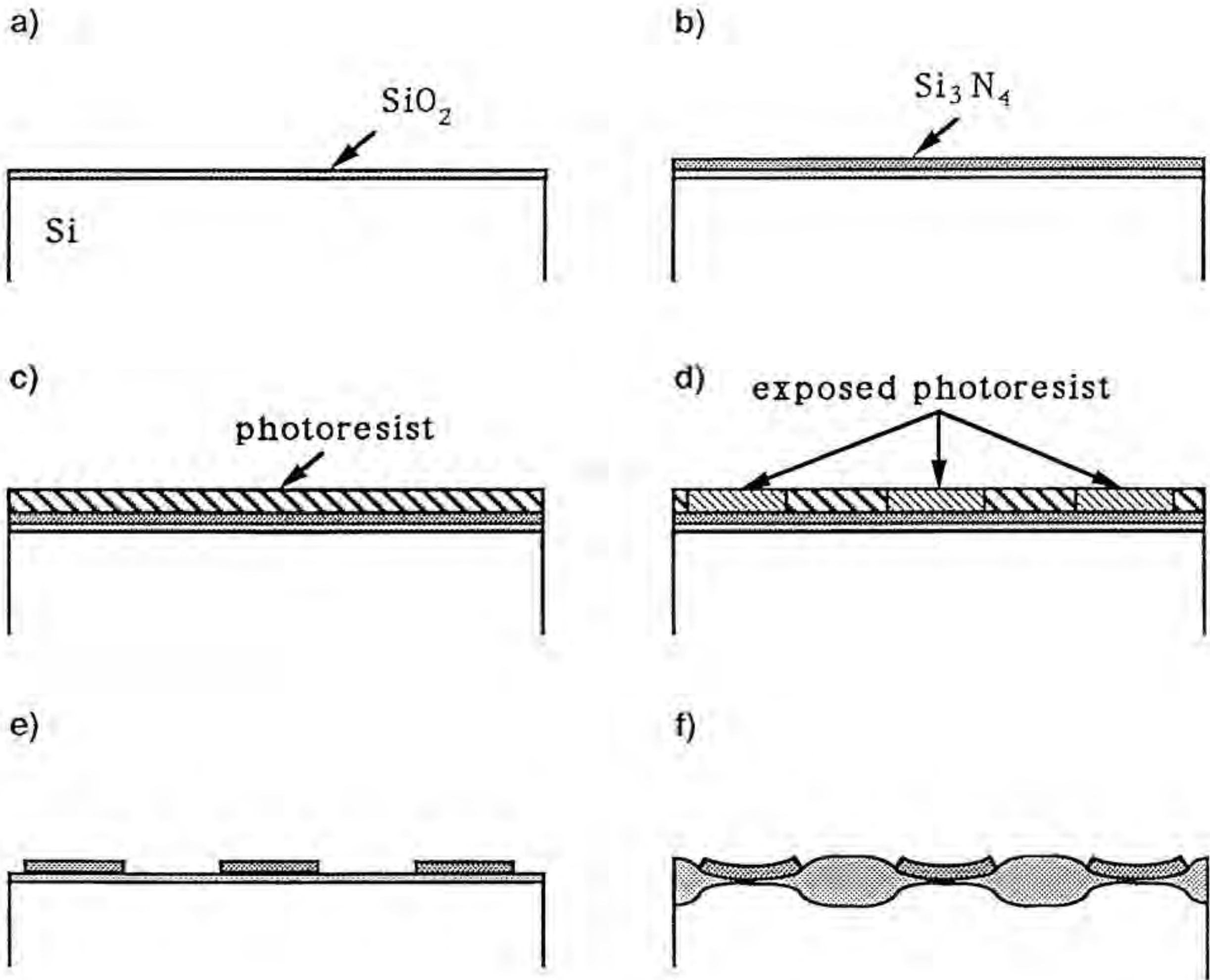


Figure 3.1: Schematic representation of the LOCOS process: a) growth of the pad oxide, b) deposition of the nitride, c) deposition of the photoresist, d) definition of the field areas, e) removal of the photoresist and of the nitride, f) growth of the field oxide.

3.2.2 Geometry effects

During this field oxidation step not only the silicon substrate from opened windows is oxidized but also part of the oxygen can penetrate laterally below the mask through the pad oxide and oxidize the underlying silicon. This effect gives rise to an elevation of the mask edge, which has a characteristic shape and which is known as bird's beak (BB) (figure 3.2). Its main drawback is the reduction of the active area that can be used afterwards for device fabrication. The BB is one of the most important limiting factors for the further increase of the packing density.

Figure 3.2 shows a cross-sectional TEM micrograph of a typical LOCOS structure, for which the different layer thicknesses have been optimized in order to get the shortest bird's beak without defect generation in the substrate. The pad oxide is 10 nm thick, grown in dry oxygen, and the nitride, 150 nm. The wet field oxidation is performed at 975°C. In this case the length of the BB is of the order of 500 nm for a FOX thickness of, nominally, 600 nm.

A field oxidation without pad oxide produces a much shorter BB, as illustrated in figure 3.3, but an extremely dense dislocation network is also present, making this approach useless.

This example shows that the use of the pad oxide decreases the interfacial stresses at the silicon surface, thus strongly decreasing the number of substrate defects. A second advantage is that the pad oxide acts as a protective layer for the silicon surface during the removal of the nitride. This process is performed using boiling orthophosphoric acid, which also attacks the silicon substrate. In absence of the pad oxide this process would give rise to the creation of a rough silicon surface.

A semi-empirical study of the influence of the processing parameters on the shape of the bird's beak structure can be found in reference [3.4]. The main conclusions of the model developed by these authors are:

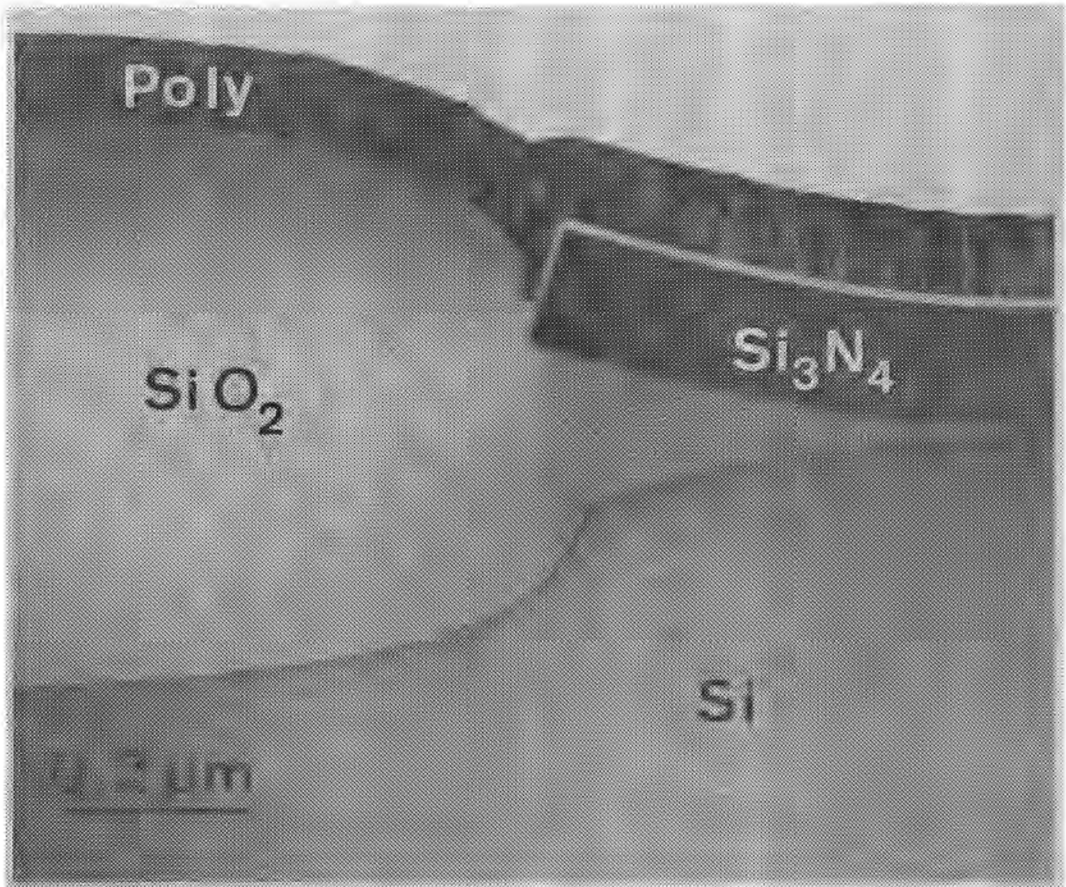


Figure 3.2: Cross-section image of the BB resulting from an optimized LOCOS process.

a) A decrease in the pad oxide thickness reduces the BB length, but increases the stresses in the substrate.

b) Increasing the processing temperature results in a decrease of the BB length.

c) At the initial stage of the LOCOS process, the ratio between the BB length, L_{BB}, and the FOX thickness, t_{FOX}, is very high. It decreases as the FOX increases, reaching a constant value.

3.2.3 Gate oxide thinning

The presence of the BB is not the only drawback of LOCOS. During the growth of the field oxide, which is performed in a wet oxygen ambient (steam), some water reacts with the silicon nitride, forming ammonia. This ammonia can diffuse towards the tip of the bird's beak and form a layer of oxynitride at the interface between the silicon and the pad oxide. This phenomenon is shown in figure 3.4, which is a combination of the models proposed by Kooi [3.5] and Nakajima [3.6]. This nitrided layer will act as oxidation barrier and cannot be removed by the HF etching of the pad oxide, performed prior to the growth of the gate oxide. The direct growth gives rise to gate oxide thinning [3.7] (the "white ribbon" effect) close to the BB edge and, consequently, to electrical problems in the processed devices, because the electrical field in these areas will be too high.

A way to circumvent this problem is to use a sacrificial oxidation prior to the gate oxide growth to consume the oxynitride. The sacrificial oxide is then etched away, after which the gate oxide is grown. In order to assure the complete removal of the pad and sacrificial oxides, some overetching is always performed, reducing the length of the BB. This BB reduction is visible in the resulting devices (figure 3.5).

3.2.4 Drawbacks and alternatives to the LOCOS process

As indicated in the previous sections, LOCOS has an important drawback which is the occurrence of an encroachment of the oxide below the oxidation mask. For the use of LOCOS isolation in ULSI technology it is essential to reduce the BB to values of

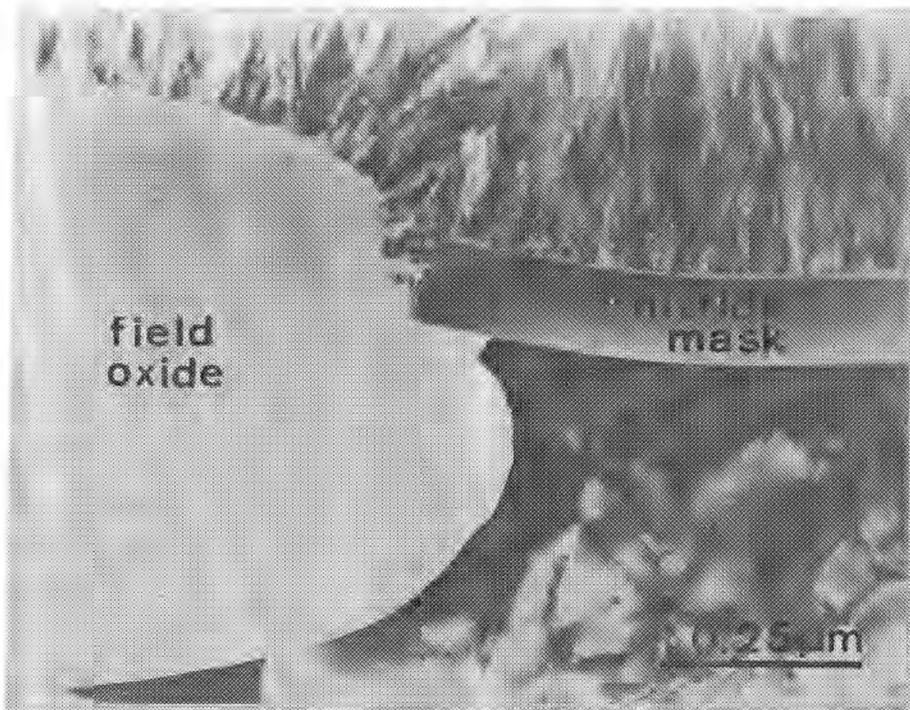


Figure 3.3: LOCOS process without the pad oxide. Observe the dense dislocation network due to the plastic deformation of the substrate (Courtesy J. Vanhellemont).

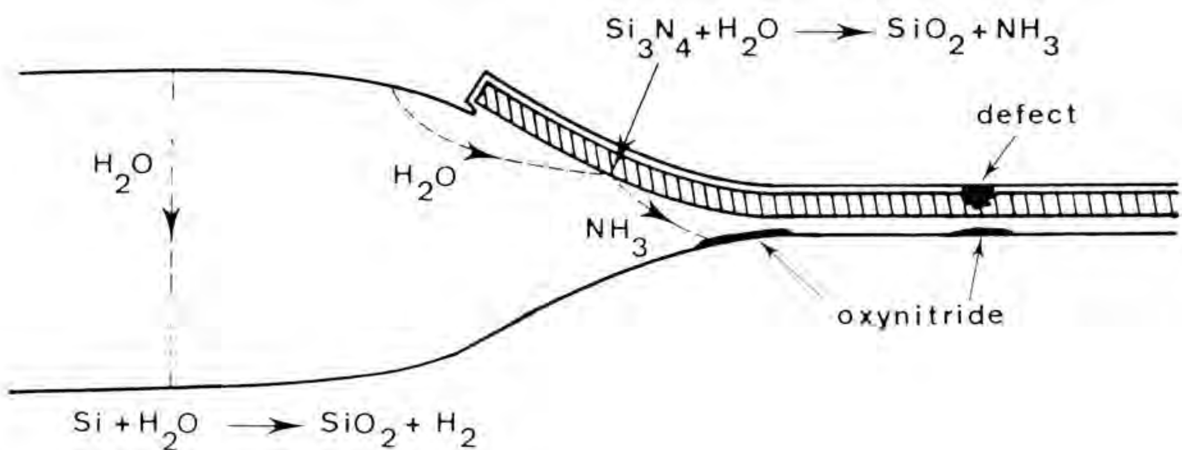


Figure 3.4: Schematic representation of the combination of the models proposed by Kooi [3.6] and Nakajima [3.7] for the cause of the gate oxide thinning (Courtesy J. Vanhellemont).

about 0.25 μm, while, at the same time, keeping a FOX thickness large enough to obtain adequate isolation.

Since the early 80's a large effort is devoted to this issue and alternative isolation schemes have been proposed. Mostly they involve the use of several sophisticated process steps. Some of the most promising methods are:

- Use of deep trenches refilled with oxide, nitride and/or polysilicon [3.8, 3.9].
- SWAMI (Side WAll Masked Isolation) [3.10, 3.11].
- Direct moat [3.12].
- BOX (Buried OXide), in which the thermal oxide is replaced by a deposited one [3.13].

In order to reduce the BB while still keeping the simplicity of the LOCOS process, other approaches have been investigated and have been reported to be suitable for submicron technologies. In the next section two of these modifications have been studied. Their advantage is that they can be performed using conventional processing steps.

- SILO [3.14, 3.15] (Sealed Interface Local Oxidation) technology, where a layer that inhibits the oxidation is put in direct contact with the silicon substrate.
- Use of a polysilicon buffer layer between the pad oxide and the nitride, giving rise to the so-called poly-buffer LOCOS [3.16] or LOPOS (Local Oxidation of POlysilicon over Silicon) [3.17].

3.3 SEALED INTERFACE LOCAL OXIDATION

3.3.0 General remarks

SILO consists, basically, in masking the silicon surface against oxidation by creating or depositing a layer that is poorly or absolutely not oxidized. This layer is typically ni-

tride or oxynitride. As the lateral diffusion of oxygen below the mask is dominant in the initial stages of the oxidation, the sealing layer will strongly inhibit this process. In this way only the edges of the sealing layer will poorly oxidize, giving rise to extremely short BBs. For the formation of this layer different procedures can be followed. All of them effectively eliminate the native oxide by transforming it into a nitride or oxynitride layer. The main three methods that have been proposed for the formation of this sealing film are:

a) A shallow implantation of nitrogen into the silicon where the active areas will be formed [3.14].

b) Thermal nitridation of the pad oxide [3.18], in order to convert it into an oxynitride film.

c) Deposition of a nitride film by LPCVD in direct contact with the silicon substrate [3.15].

As is already pointed out in section 3.2, during the growth of the field oxide and formation of the bird's beak, an oxynitride rich area can be created close to the BB tip. In the case of the methods listed above, the probability of forming the oxynitride is quite high, because of the high nitrogen presence at the silicon surface. This imposes the use of a sacrificial oxidation to avoid gate thinning effects. In addition, the problem already mentioned of etching of these nitrogen rich layers without etching the underlying silicon substrate can easily arise. It has been proposed that the use of a consumable nitride layer might solve these problems [3.19]. A further problem is to avoid any nitrogen contamination in the active areas, which may degrade the electrical performance of the devices.

In the next sections methods b) and c) will be discussed in more detail.

3.3.1 Nitridation of the pad oxide

3.3.1.1 Experimental procedure

As mentioned in the previous section the nitridation of a thin oxide to form oxynitride is a promising way of inhibiting

the oxidation of the silicon substrate in the protected areas. In reference [3.18] the use of rapid thermal nitridation at temperatures in the order of 900 to 1250°C is proposed as satisfactory method to achieve it.

In our experiment the pad oxide was 10 nm thick and the nitridation has been carried out with ammonia gas flow at two different temperatures, 1050 and 1100°C, for times ranging between 1 and 10 minutes. These times are long enough, according to [3.18], to strongly inhibit the oxidation. Next the conventional LOCOS process has been used as described starting at point b) of section 3.2.1.

3.3.1.2 Results

Using this approach a very small reduction of the BB length is obtained and the geometrical aspect of the structure is very similar to the one of figure 3.2. Probably the reduction is only effective during the very initial stages of the FOX growth, while at longer times the behaviour is very similar to conventional LOCOS.

3.3.1.3 Fully processed transistors

In order to test the feasibility of this isolation method to real devices, twin well CMOS transistors have been processed. A cross section TEM micrograph obtained in the gate area of a 0.7 μm wide PMOS transistor is shown in figure 3.5a [3.20]. In this figure the BB length is about 250 nm, much shorter than for conventional LOCOS (compare with figure 3.2). The length of the BB is defined as the distance between the area where no increase of the gate oxide thickness takes place and the position where it is supposed that the edge of the nitride was. It has to be mentioned that this BB length, which is obtained also from electrical measurements, does not correspond to the one defined when the nitride mask is still present, because of the two oxide etching steps that are used, as explained in section 3.2.3.

In the same figure it can be observed that the silicon/gate oxide interface is not flat but presents a certain slope. This slope increases starting from the FOX as it approaches the cen-

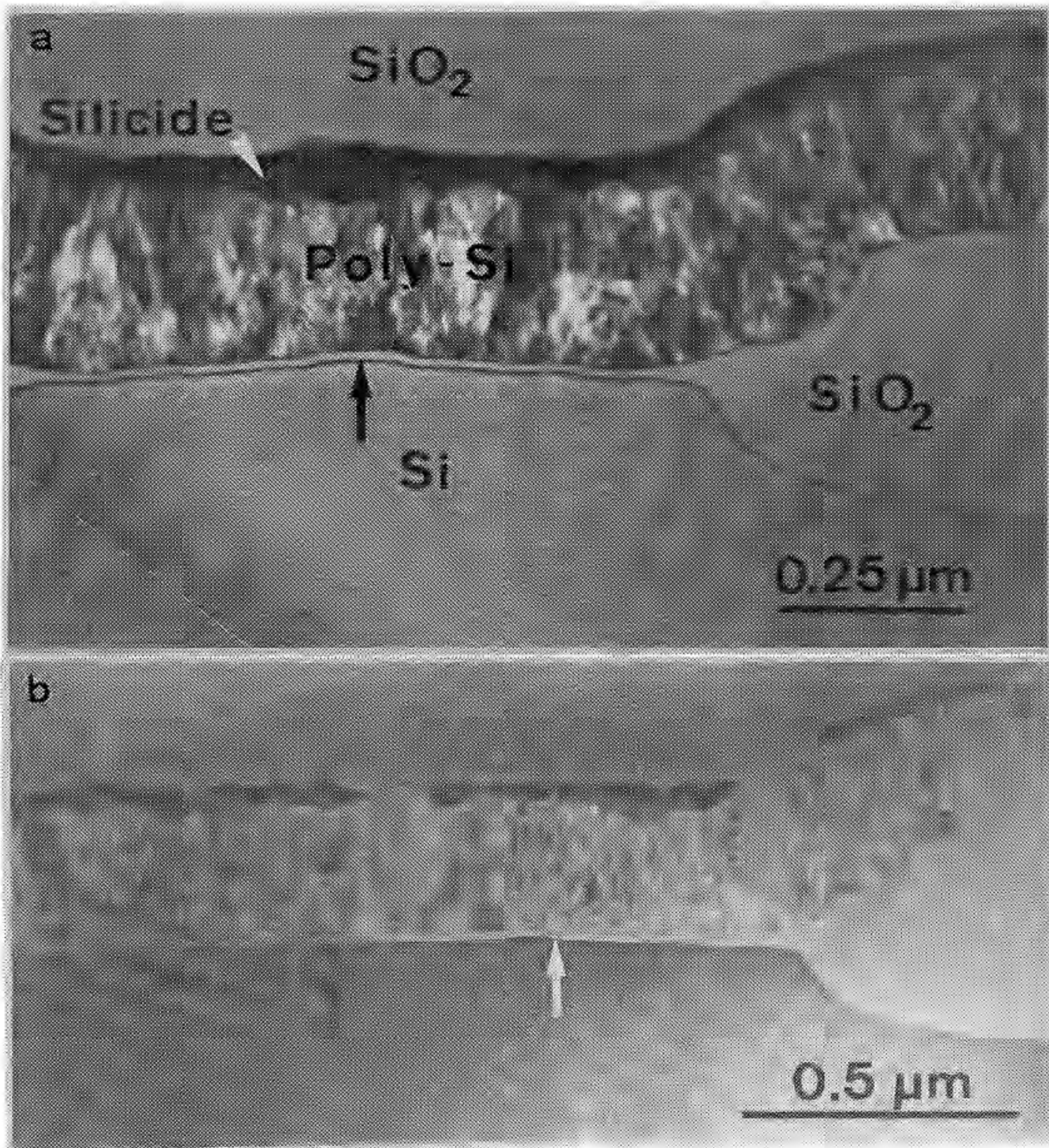


Figure 3.5: a) Gate area of a 0.7 µm PMOS transistor whose isolation has been obtained using nitridation of the pad oxide. The black arrow indicates the centre of the gate and the white, a bump in the gate oxide. b) Gate area of a 1.5 µm transistor. Observe that the bump is at the same distance from the edge of the FOX as for figure a.

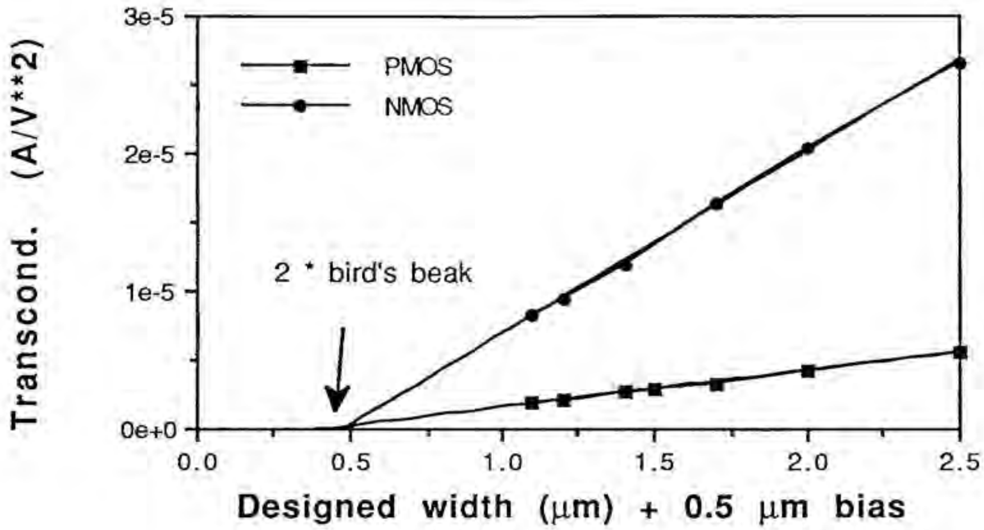
tre of the structure (black arrow). Very close to the centre a bump in the gate oxide is visible, indicated with a white arrow. Figure 3.5b corresponds to a 1.5 μm transistor from the same wafer. The bump is present at the same distance from the FOX as in the previous picture. Comparing these figures with a LOCOS structure with the nitride mask still on top, it has been deduced that the bump corresponds to the edge of the bird's beak tip. The observed short BB after complete processing is due to severe overetching of the different oxide etching steps. However, it is very dangerous to rely on overetching in order to produce isolation structures for submicron technology, as the etching rate is strongly influenced by the age, lot number, temperature and other parameters of the etchants, which are difficult to control in practice.

Anyhow, the devices realized using this isolation method present reasonable electrical characteristics, as can be deduced from figure 3.6a. This figure indicates that the electrical bird's beak for NMOS and PMOS transistors is similar to the values obtained in figure 3.5. Although the most narrow gates (0.6-0.7 μm) show a very undulated interface due to the very small region between the two bird's beak tips, the CMOS transistors still have quite satisfactory electrical characteristics and the threshold change is still small (figure 3.6b).

3.3.1.4 Asymmetry in closely packed structures

In addition to the disadvantage mentioned in the previous section, the described process presents another drawback. It is presented in figure 3.7, showing a cross-section corresponding to two areas with different field oxide window width and separation. Figure 3.7a corresponds to oxidation windows of 0.5 μm, while the windows of figure 3.7b are 0.7 μm wide. The masked areas are in both pictures 0.6 μm wide.

In these figures two surprising effects can be observed. The first is the evident asymmetry of the BB for the different field oxide windows. It can clearly be seen that the BB from the FOX region 1 is much longer than the left BB from area 2. At its time, the right BB from 2 is shorter than the left BB from win-



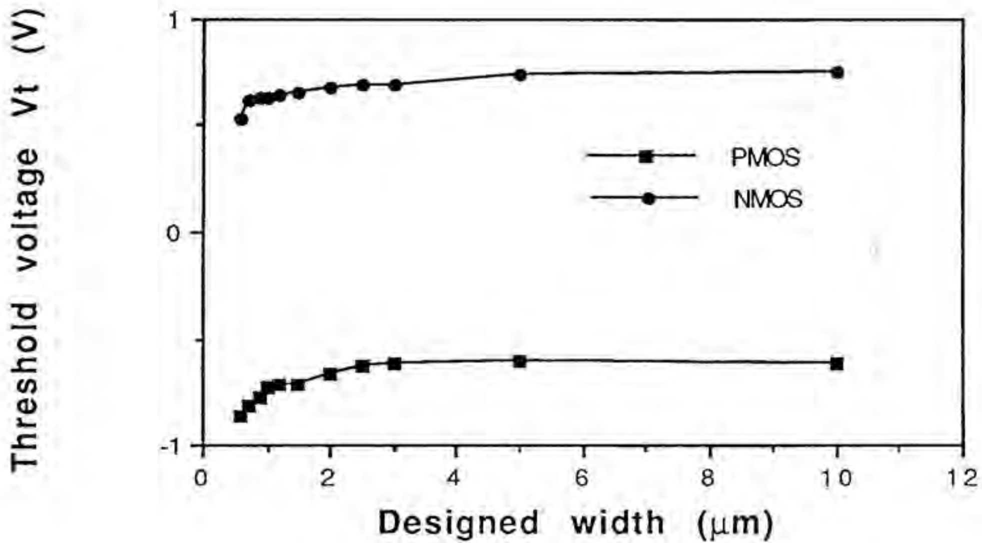


Figure 3.6: a) The transconductance as a function of the designed transistor width. b) The influence of the transistor width on the threshold behaviour.

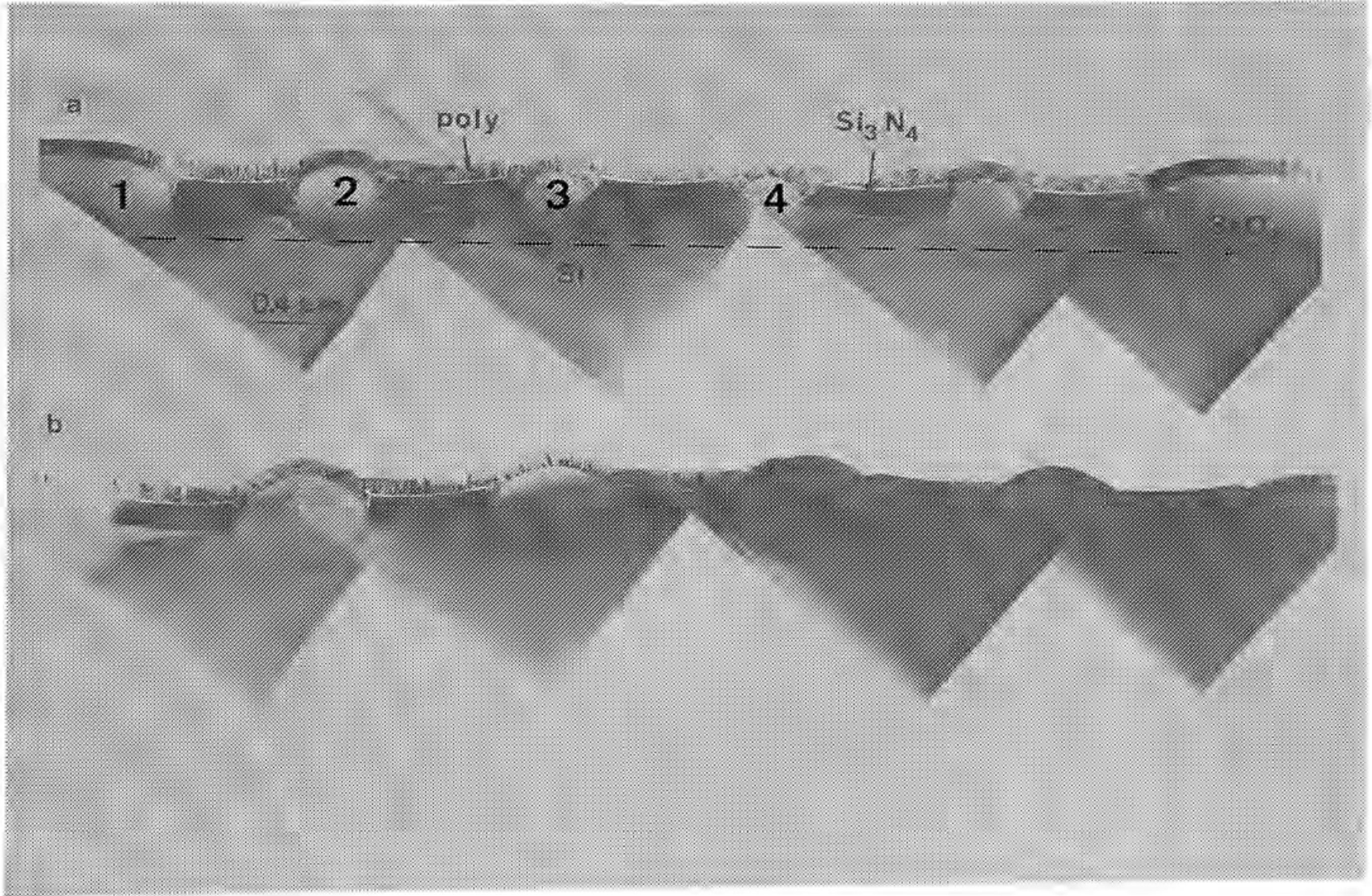


Figure 3.7: Asymmetry of the BB in a highly packed structure. Also the lifting of the field oxides is visible. The drawn line is parallel to the original surface of the wafer. The isolation has been obtained using nitridation of the pad oxide.

dow 3. Its right BB is equally long as the left of area 4 because of the symmetry of this structure: the area between windows 3 and 4 is the centre and the right part of the figure is a mirror image of the left one.

The second effect is shown by the line which is drawn parallel to the original silicon surface. It shows that the distance between this line and the bottom of the different field oxide windows is not the same throughout the structure. The more to the centre, the more the field oxide is lifted. The field oxide thickness remains the same for windows of the same size, indicating that it is not an effect of the FOX thinning (see section 3.4.2.3).

3.3.2 Nitride layer in direct contact with the silicon substrate

3.3.2.1 Processing conditions

As mentioned in section 3.3.1, the use of a nitride layer in direct contact with the silicon substrate inhibits the oxidation, thus reducing the BB length. Thin nitride layers, less than 30 nm thick, have to be used in order to avoid the nucleation of defects in the silicon substrate as a consequence of the stresses that appear at the interface [3.15].

In this experiment a NONO structure (nitride-oxide-nitride-oxide) has been used to test the feasibility of this SILO approach. The lowest nitride layer is composed of two different ones: the first has been created by rapid thermal nitridation of the silicon surface, as explained in the previous section, to a thickness of about 2 nm. The second step consisted on the LPCVD deposition of a nitride layer 10 nm thick. Next a 30 nm thick oxide layer was grown, followed by another nitride of 50 nm, and finally an oxide of 90 nm. Photolithography has been used to define field regions and all the deposited or grown layers have been removed in these windows until the substrate was reached. In the masked areas the last oxide layer was partially removed, leaving about 20 to 30 nm.

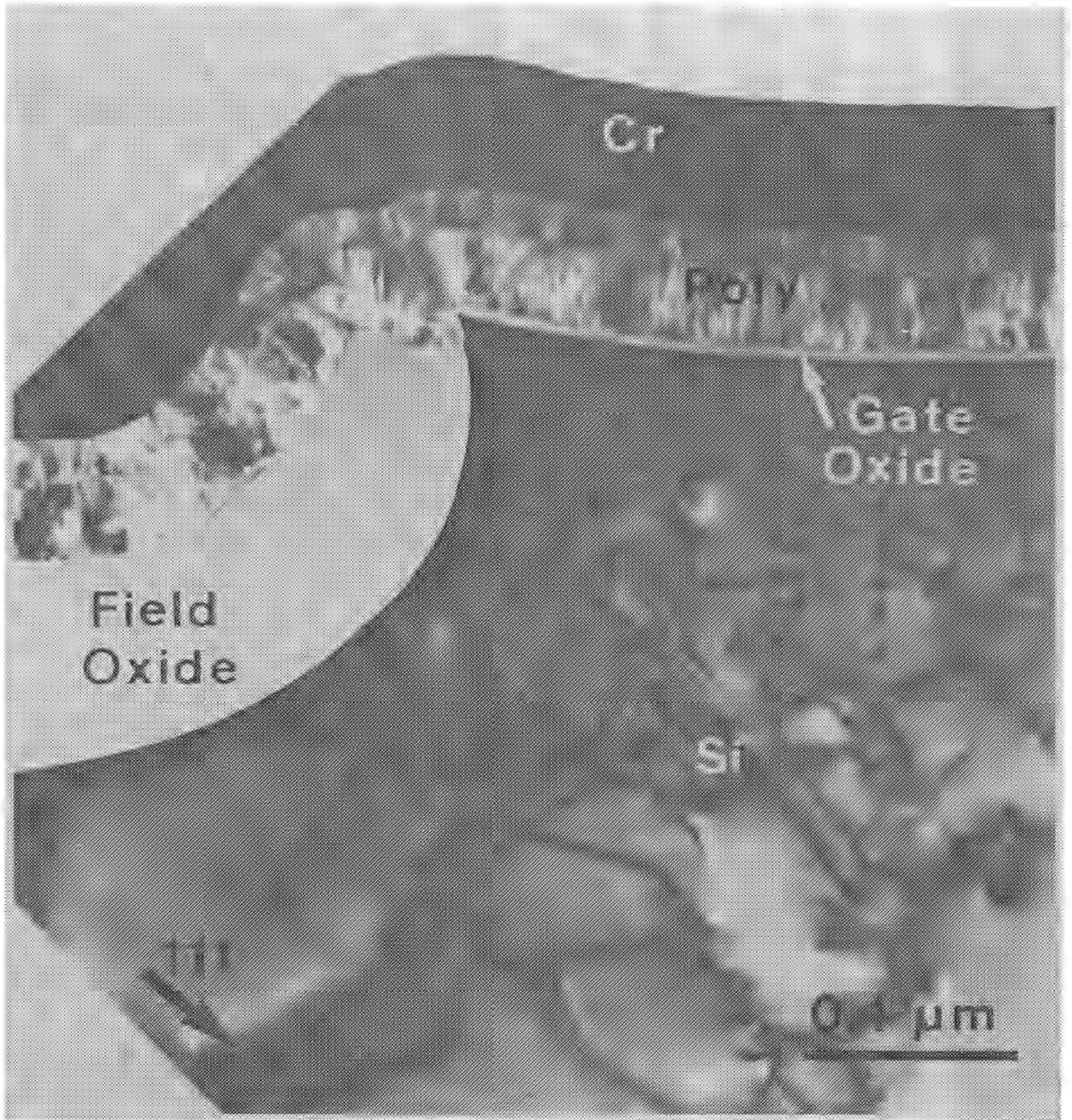


Figure 3.8: SILO structure obtained with a NONO stack. Observe the dense dislocation network below the gate oxide, the severe plastic deformation of the Si/SiO_2 interface area and the thinning of the lateral gate oxide.

Finally a 700 nm thick field oxide is grown in wet oxygen at 950°C. The processing was further continued, including the removal of the mask, the sacrificial oxidation, the growth of a gate oxide and the deposition of polysilicon and chromium to protect the surface.

3.3.2.2 TEM observations

Figure 3.8 shows a cross-section of a sample processed under these conditions. It can clearly be seen that an extremely dense dislocation network is present, making this approach not feasible for ULSI application. We believe that, although the thickness of the first nitride layer is about 12 nm and a stress relief oxide is deposited between the two nitride layers, the second strongly contributes to an increase of the stresses in the substrate.

A second drawback is that the gate oxide close to the field oxide is not flat, but strongly bent upwards, with a slope of about 10 to 15°. This is due to a plastic deformation of the substrate, caused by the high stresses. During the growth of the FOX this strongly pushes laterally the gate region, due to the volume increase of the oxide. As a consequence the silicon substrate tends to be lifted close to the FOX.

A third important drawback of this isolation method is the strong thinning of the lateral gate oxide. Processing devices in this area would give rise to a very strong electrical field and, thus, to a potential failure site.

3.4 LOCAL OXIDATION OF POLYSILICON OVER SILICON

3.4.0 Introduction

LOPOS (Local Oxidation of POlysilicon over Silicon) [3.17] or poly-buffer LOCOS [3.16], is one of the most promising modifications of the LOCOS process for application in submicron isolation technology. In this process, the main difference compared to LOCOS is the addition of a thin polysilicon layer between the

pad oxide and the nitride. It acts as a buffer layer and relieves the stresses that appear as a consequence of the technological process. No additional photolithography step is required and the problems are reduced to the removal of the polysilicon layer without attacking the silicon substrate.

In our experiment two different LOPOS structures have been fabricated. One involved the use of the conventional polysilicon layer, while in the second one a microcrystalline silicon layer was used. This microcrystalline or amorphous layer transforms into polycrystalline silicon during the field oxide growth.

3.4.1 LOPOS with polycrystalline silicon

The different layer thicknesses employed in this section are: pad oxide 15 nm, 50 nm of polysilicon, 120 or 155 nm of nitride and field oxide thicknesses varying between 700 and 1000 nm.

Figure 3.9 shows a typical structure with a FOX of 700 nm. The BB is about 450 nm long, giving a L_{BB}/t_{FOX} ratio of 0.67, which is better than for optimized LOCOS. No defects were observed in these conditions. Only close to the rim of the wafer pile-up of 60° dislocations has been observed for the very narrow oxidation windows, but not for wider ones. We believe that the source of the dislocations is warpage of the wafer.

In figure 3.9 a strong bending of the nitride is seen, associated with a strong consumption of the polysilicon buffer layer. If the nitride thickness is increased, the bending should be much smaller, but the structure would be prone to extended defect nucleation. This experiment will be performed in the near future.

3.4.2 LOPOS with microcrystalline silicon

3.4.2.1 Experimental conditions

This is a new approach to the LOPOS process and is studied in this section in full detail, because we believe that it presents some advantages compared to the use of polysilicon as buffer layer. For this experiment the following parameters have

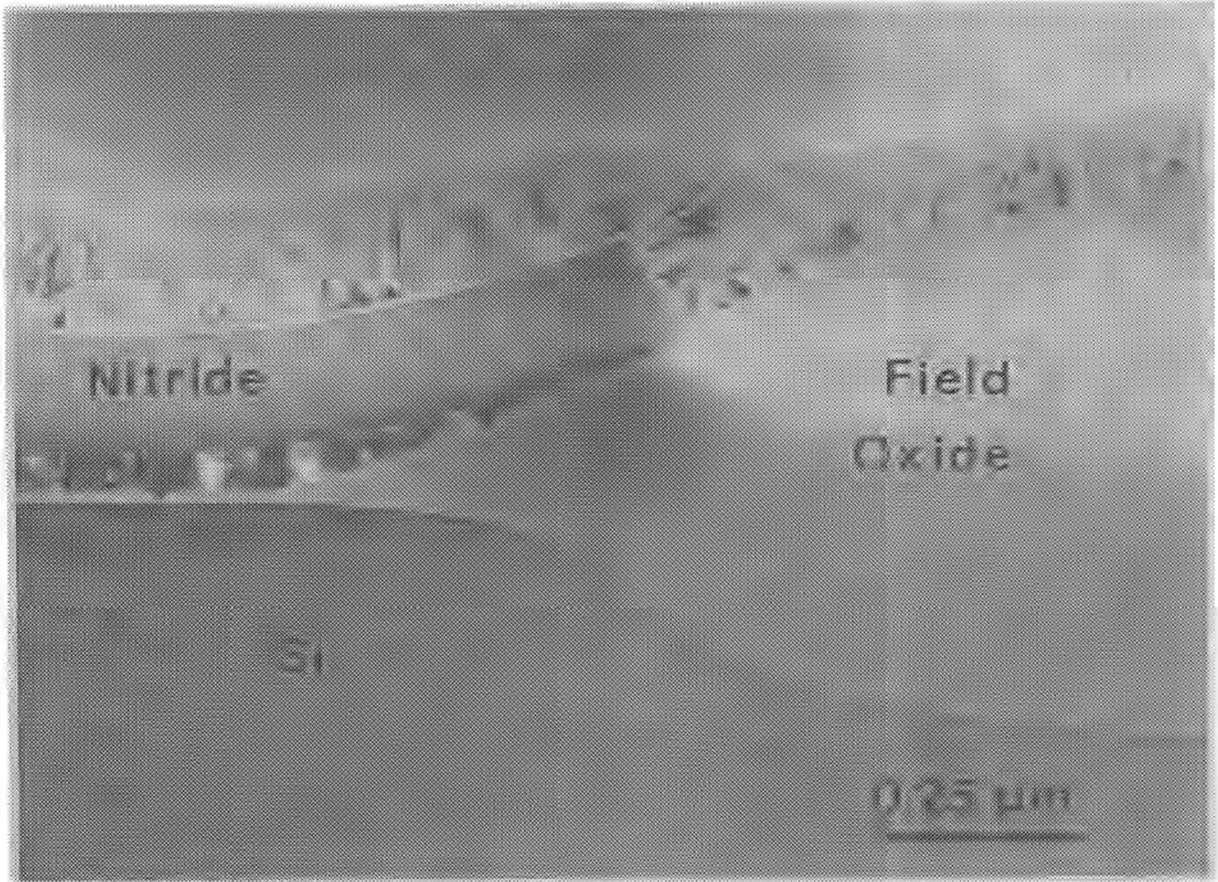


Figure 3.9: LOPOS structure processed using polysilicon as buffer layer. The nitride was 155 nm thick and the 0.7 µm thick field oxide has been grown at 950°C.

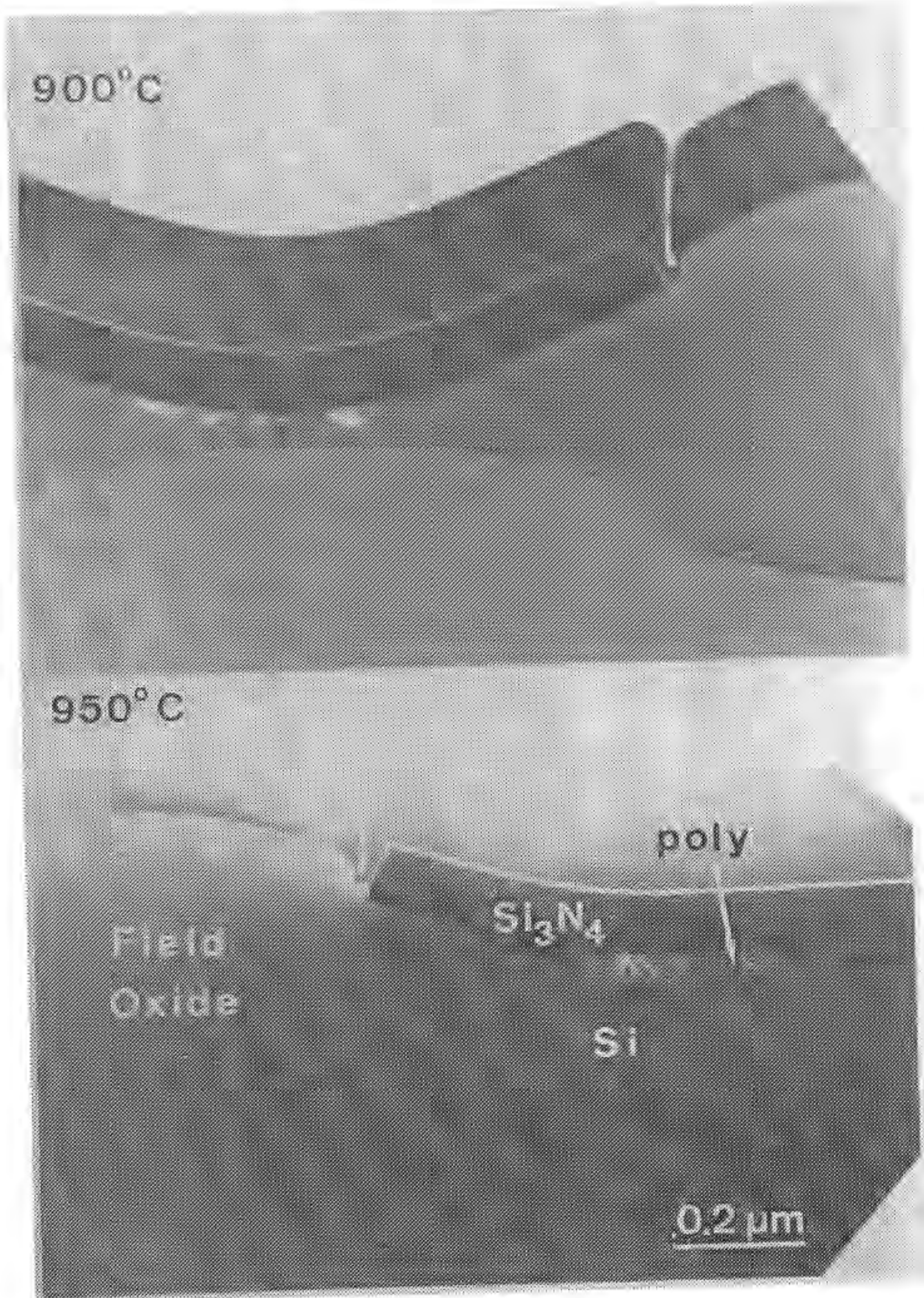


Figure 3.10: Influence of the temperature on the BB length. The pad oxide was 10 nm thick and the nitride, 100 nm. The field oxide is 0.6 µm thick.

been used. The pad oxide was either 10 or 15 nm in thickness, grown by dry oxidation. A 50 nm microcrystalline silicon layer was deposited at 620°C. The LPCVD nitride mask thickness varied between 100 and 150 nm. Finally, the temperature of the wet oxidation ranged between 900 and 1050°C and the FOX thicknesses, between 400 and 800 nm. This last process was carried out in wet ambient.

3.4.2.2 Geometrical aspects

Figure 3.10 shows two cross-sections of typical LOPOS structures, processed using a thin nitride, of 100 nm. Comparing both pictures the shorter BB for the higher temperature, 950°C, is evident, which is in good agreement with the predictions of the model by Wu et al. [3.4].

In figure 3.11 two SEM pictures, corresponding to wafers oxidized at 900°C, are presented, showing the effect of the nitride thickness on the final geometry of the BB. Figure 3.11a corresponds to a nitride of about 100 nm, while 3.11b shows the result for a value of about 150 nm. In order to give contrast in the SEM, the cross-sections, which have been obtained by cleavage, have been etched in buffer HF for about 1 minute. The amount of silicon oxide removed is of the order of 30 nm for this particular mixture. In figure 3.11b it can be seen that close to the nitride edge the oxide has been removed much faster than in the rest (indicated by arrows). In these areas the stresses in the silicon oxide are much higher and it is well known that buffer HF etches much faster stressed SiO_2 regions. This etching technique allows us to reveal highly stressed areas in the oxide. A similar etching technique has been tried on TEM specimens, but was unsuccessful. Reason for this could be the stress relaxation in the very thin foils required for TEM observation.

Figure 3.12 shows a cross-section view of a wafer oxidized at 950°C, using the thicker nitride. The BB length is about 250 nm. The picture only shows the area of the BB and a special feature of LOPOS can be observed: the appearance of a second BB between the polysilicon layer and the nitride due to the consump-

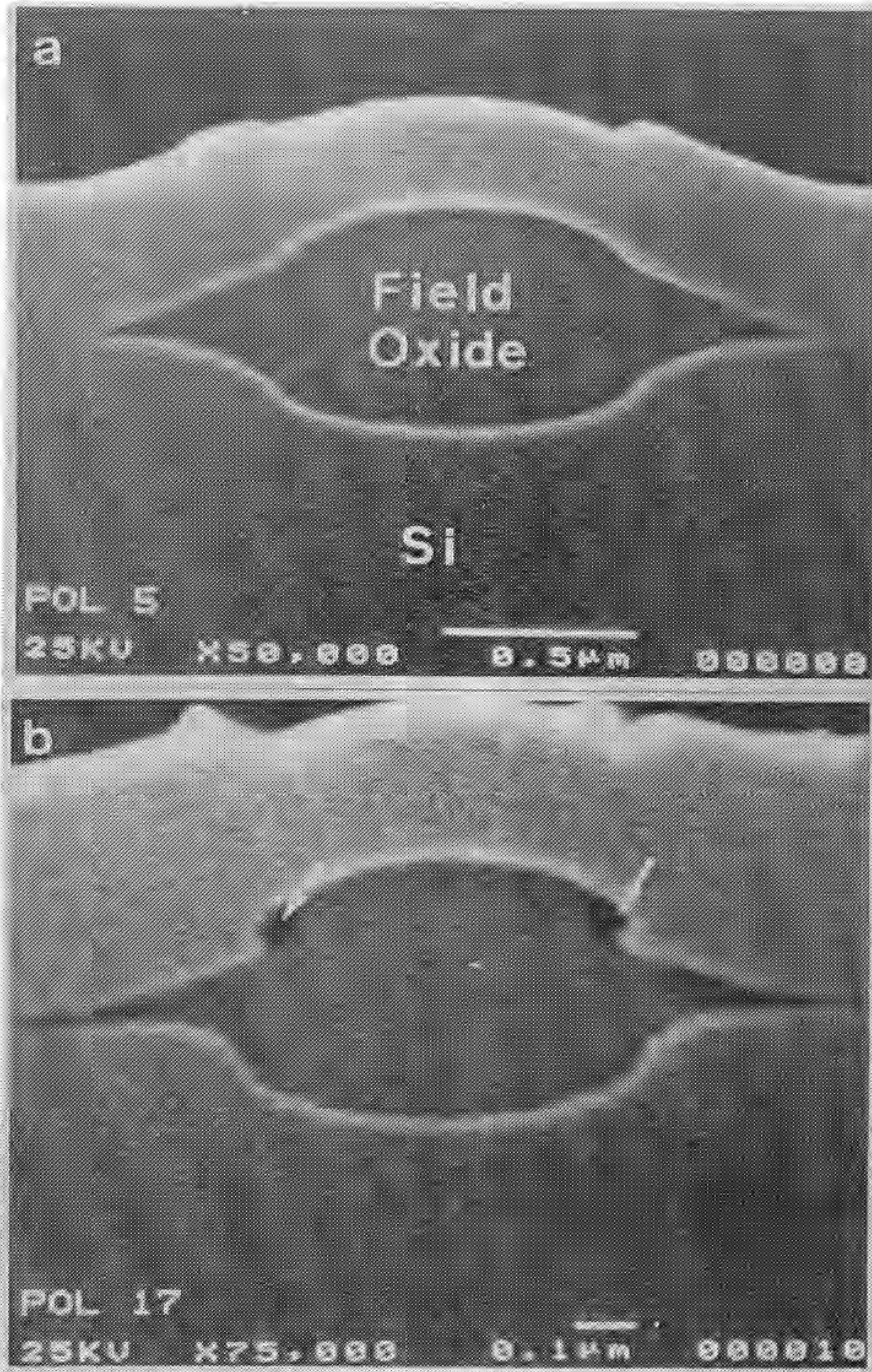


Figure 3.11: Cross-section SEM image showing the influence of the nitride thickness on the BB geometry. a) 100 nm, b) 150 nm, 1000°C. These specimens have been etched for 1 minute in buffer HF in order to introduce a contrast difference between the different layers. In addition the stresses in the silicon oxide close to the nitride edge have been revealed.

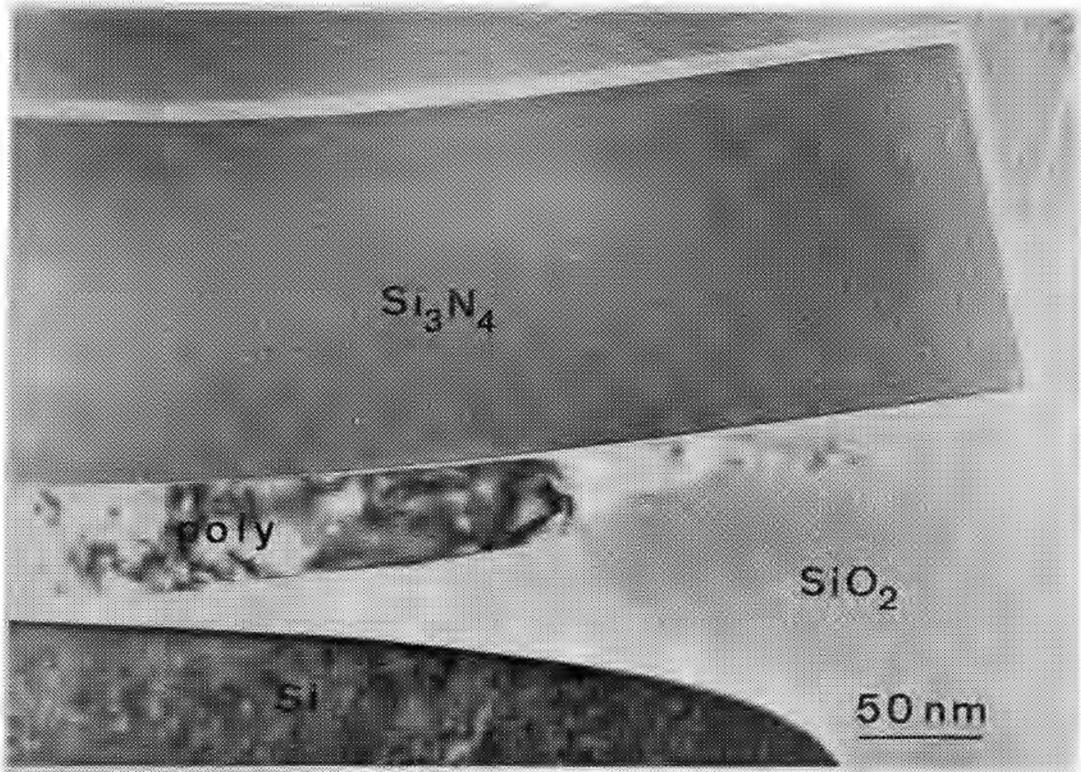


Figure 3.12: BB of a LOPOS wafer oxidized at 950°C. The appearance of a second BB between the polysilicon and the nitride is clearly visible.

tion of part of the polysilicon. The oxidation of the polysilicon buffer layer is, in part, responsible for the reduction of the BB. From the oxygen that is flowing through the oxidation window, one part is used in oxidizing the substrate, another diffuses through the pad oxide and oxidizes the underlying silicon. A third part is used to oxidize the polysilicon edge. If during the formation of the LOPOS stack some time passed between the deposition of the microcrystalline silicon and the covering nitride, a thin native oxide will be grown at its top. During the field oxidation part of the oxygen can also penetrate through this native oxide and oxidize the polysilicon. This could explain why a second BB appears.

Due to this consumption of the polysilicon, the amount of oxygen that is available to consume the substrate below the pad oxide is smaller, giving rise to a shorter BB. As a consequence, the silicon-pad oxide interface will remain more horizontal than for conventional LOCOS and the slope below the edge of the mask will be steeper. This should lead to devices with better electrical characteristics.

3.4.2.3 Stresses in LOPOS structures

The reduction of the BB length leads to a strong increase of the stresses at the nitride edge. It is known that these stresses can lead to the generation of defects (mainly dislocations) in the substrate [3.2]. This is one of the possible mechanisms, but also other have been found.

The specimen of figure 3.13a shows a peculiar effect, which is the breaking of the nitride and separation from the polysilicon layer. This effect has only been observed at temperatures below 950°C. The length of the BB is the same as mentioned above. We believe that the breaking phenomenon is a consequence of the increase of the stresses and, thus, it is a stress relief mechanism. In the same specimen another stress relief mechanism has been found, which can be seen in figure 3.13b. In this case the stress relaxation takes place in the pad oxide, which breaks up. The second mechanism is dominant, and appears at nearly all oxidation temperatures. We believe that in both cases the driv-

ing force is the volume change of the microcrystalline silicon during its transformation into polysilicon.

At higher temperatures, always at 1050°C and, sometimes, also at 1000°C, a third relaxation mechanism takes place. It is the creation of extended lattice defects in the substrate, as can be seen in the plan view of figure 3.14a (using a 10 nm pad oxide and a nitride of 150 nm) and in the cross-section of figure 3.14b (pad oxide 15 nm and the same nitride thickness). The dominant defect type are dislocations running parallel to the nitride lines, although also a relatively high density of surface stacking faults is observed in the field oxide area. The fact that these defects appear at the higher temperatures is not surprising. The oxidation rate is larger than at lower temperatures, leading to a higher generation rate and injection of silicon self-interstitials. This higher supersaturation of self-interstitials leads to a lower yield stress for defect nucleation [3.21].

For the determination of Burgers vectors of dislocations (figure 1.4), one of these samples has been used in the first chapter. In that figure it can be seen that the 60° dislocations have alternating the two possible Burgers vectors in the same glide plane. This is in agreement with the model by Vanhellemont et al. [3.22]. According to this model the alternating Burgers vectors give rise to a more efficient stress relief at the film edges.

In figure 3.14b it can also be seen that the generation of defects is accompanied by severe plastic deformation of the substrate. This is evidenced by comparing the BB that appears at this temperature with the one in figure 3.11b. The BB shape is quite different in both cases.

In certain areas of the wafer it has been observed that the generation of defects and the breaking of the pad oxide coexist. In addition, if the first mechanism takes place, however, the field oxide areas close to it do not present defects nor breaking of the pad oxide, indicating a medium range relaxation.

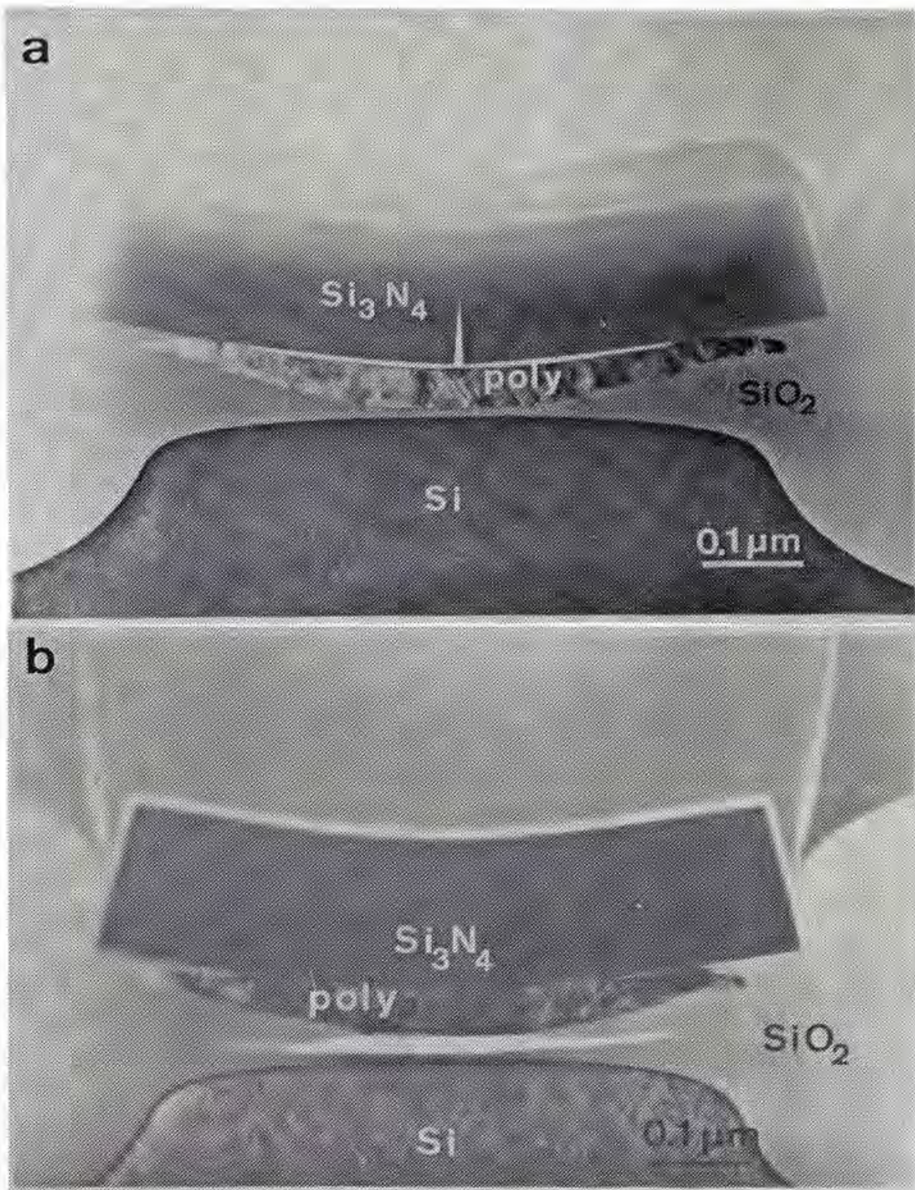


Figure 3.13: a) Breaking of the nitride and separation from the polysilicon as consequence of the high stresses of this process. b) A second relaxation mechanism, consisting in the breaking of the pad oxide. Both mechanisms coexist in the same wafer (10 nm pad oxide, 150 nm Si_3N_4, 950°C).

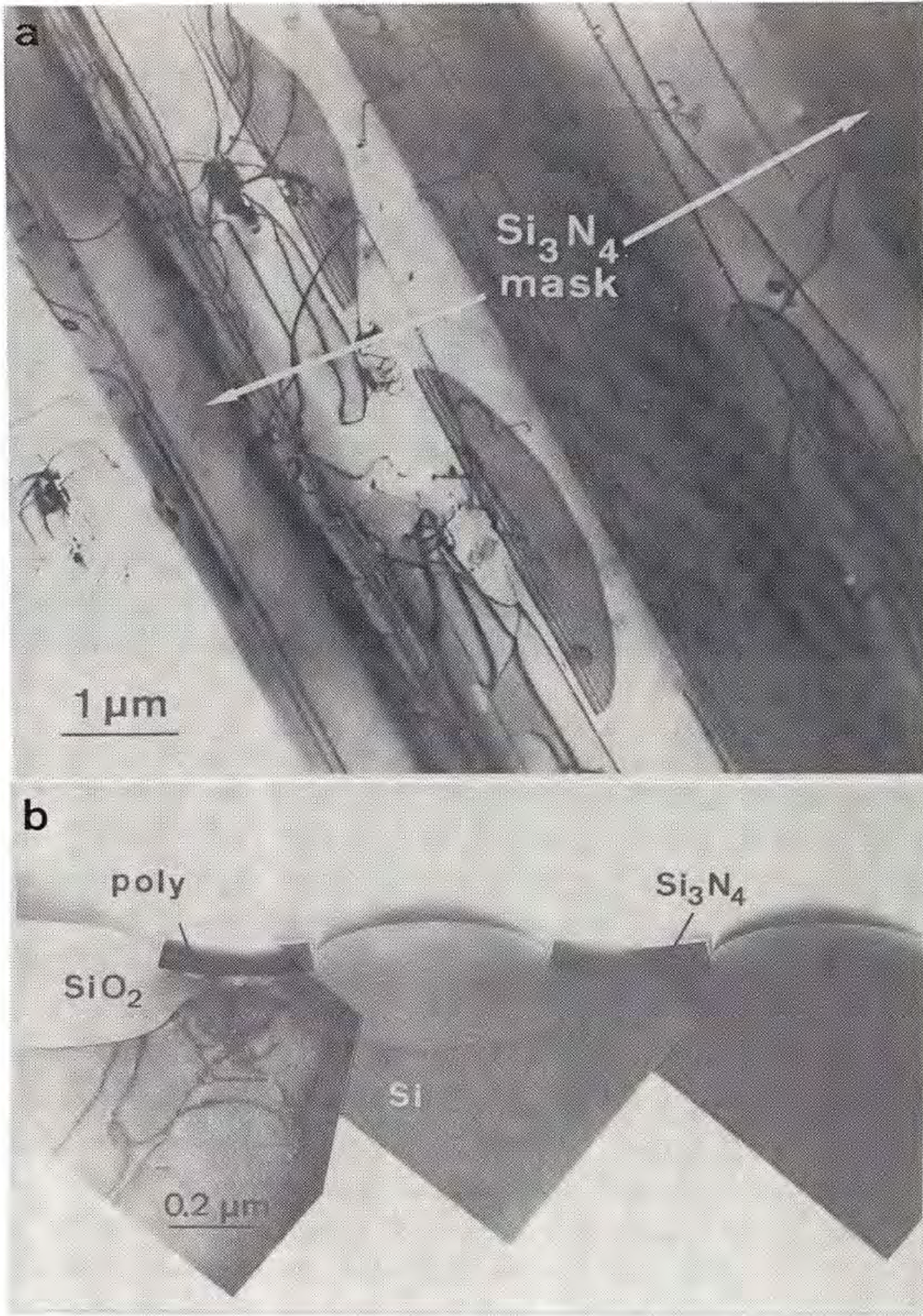


Figure 3.14: a) Plan view TEM micrograph showing the defects generated between the lines (10 nm pad oxide, 150 nm Si_3N_4, 1050°C). b) Cross-section of a specimen with a thicker pad oxide, 15 nm, showing generation of defects and plastic deformation of the substrate close to the BB.

TEM can also provide some information about stresses present in crystalline materials. A simple technique has been proposed, that allows to directly image the stresses induced by thin film edges [3.23]. The authors propose the use of the diffraction vectors $\bar{g}_{200}$ or $\bar{g}_{220}$, because the interpretation of the image is more simple. However, good two beam conditions using these diffraction vectors and that allow to image the stresses are difficult to obtain. Much simpler, because of the higher contrast, is to use a vector of the type $\bar{g}_{113}$. This is presented in figure 3.15. As predicted also in the work by Vanhellemont et al. [3.22], the mechanical strains are concentrated at the BB tip. This stress concentration can be observed by the contrast contours close to the tip. It has to be mentioned that in this case relaxation of the stresses in depth takes place, because of the free surfaces of the cross-section specimen.

Raman microprobe spectroscopy has been reported to be a good technique for the local measurement of stresses in crystalline materials [3.24] and recently application to LOCOS structures has been reported [3.25]. In this last work the nitride and pad oxide were removed in order to obtain the Raman spectra. This has the drawback that relaxation in the vertical axis takes place, so that not the total stress components are measured.

In our experiment the measurements of the stresses induced in the silicon substrate by the LOPOS process have been measured [3.26, 3.27]. All the spectra have been recorded without removing any of the top layers to avoid the above mentioned relaxation and to be able to measure the total stress.

Figure 3.16 shows Raman spectra corresponding to different line widths for a LOPOS sample oxidized at 1000°C to a field oxide thickness of 600 nm, with a 10 nm pad oxide and a 150 nm thick nitride. The spectra are plotted in such a way that the vertical axis indicates the shift of the Raman frequency from the stress free value, while the horizontal axis gives the position along the width of the line, with the origin taken at its

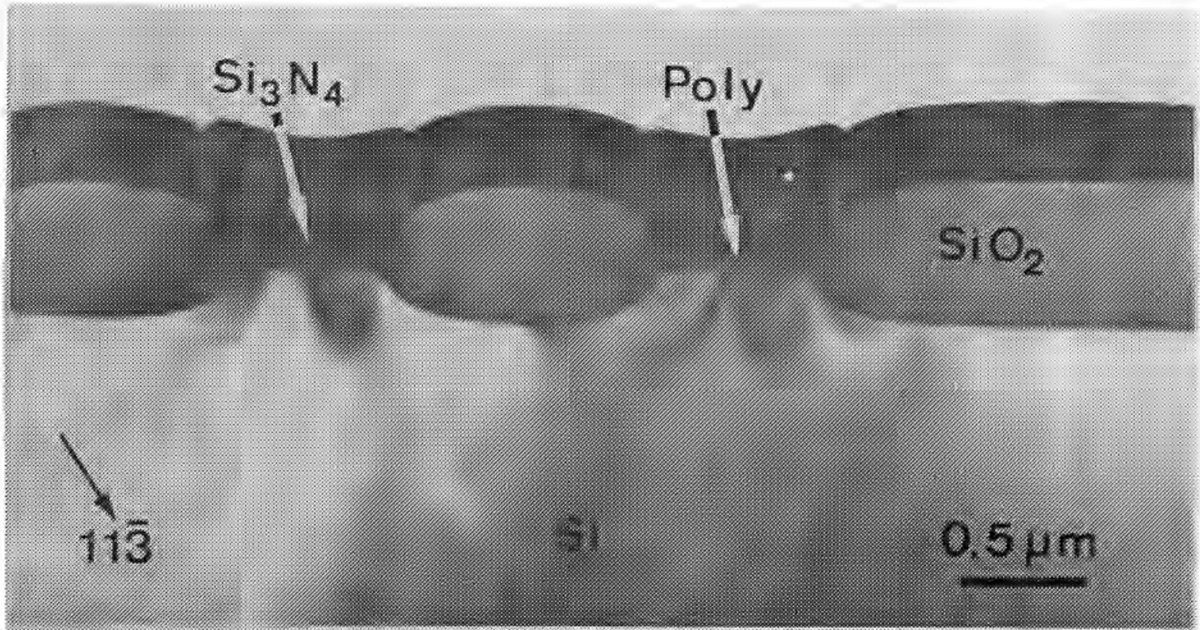


Figure 3.15: Cross-section micrograph taken using a two beam condition $\bar{g}_{11\bar{3}}$. The contours below the BB tip can be correlated with strain profiles.

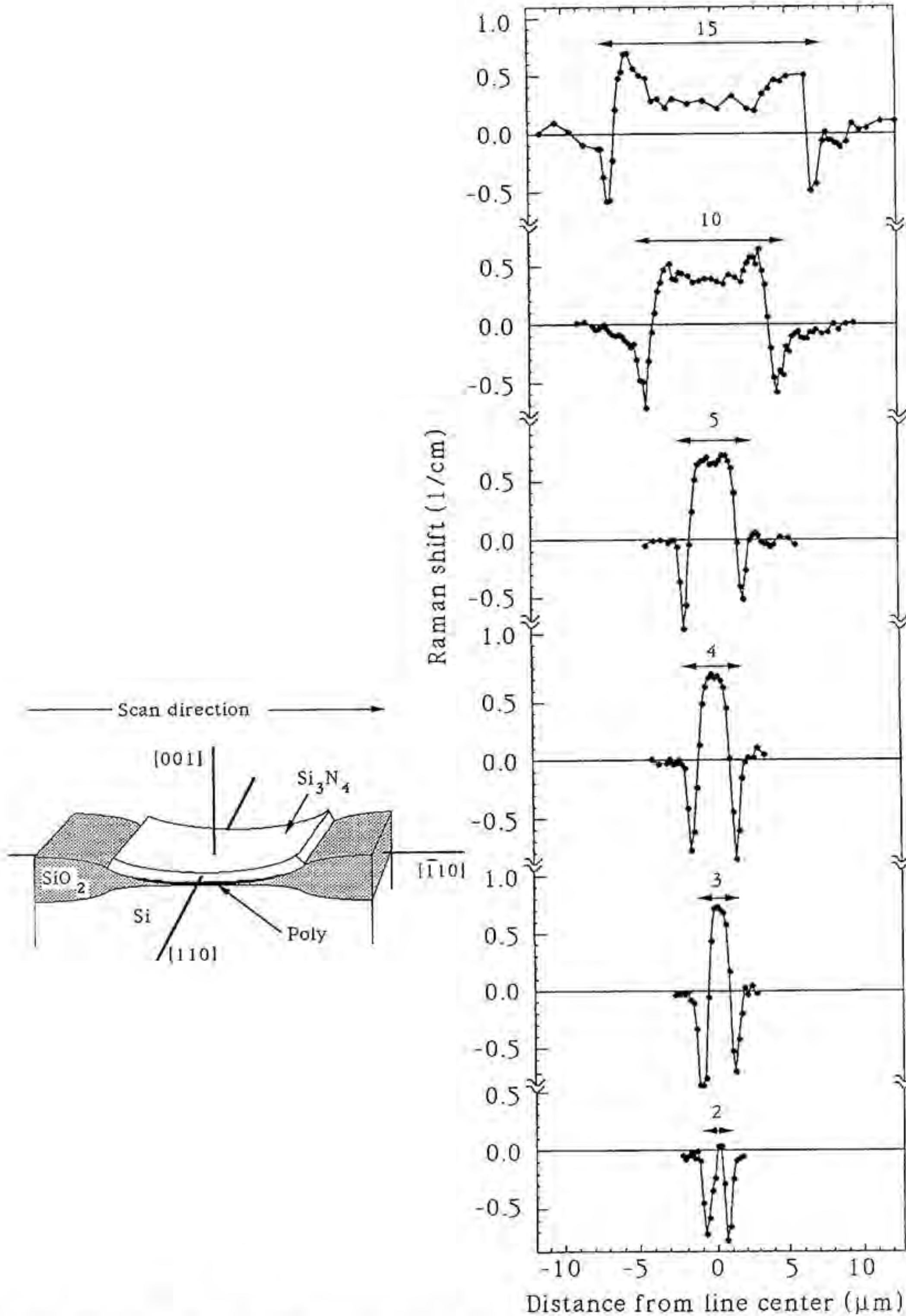


Figure 3.16: Raman microprobe spectra corresponding to different LOPOS (10 nm pad oxide, 150 Si_3N_4, 1000°C) line widths. Below the BB the Raman shift is negative, corresponding to a tensile stress. In the centre of the lines a compressive stress is observed. The value of this stress decreases with increasing line width.

centre. The positive or negative shift of the Raman frequency can be associated, respectively, to a compressive or tensile stress in the silicon substrate [3.28].

In our experiment the following is observed:

a) Below the BB a strong tensile stress is present. It is observed that the shift is independent of the line width, indicating that it is only consequence of the BB geometry. The measured Raman shift is double the value of the obtained for an optimized LOCOS structure.

b) A large compressive stress is observed in the line centre, decreasing with increasing line width.

c) For lines smaller than 4 μm the compressive stress at the centre of the line decreased and becomes even tensile. This can be due to the increasing influence of the tensile stress at the BB tip.

3.4.2.3 Field oxide thinning

As briefly mentioned in section 3.3.1.4, with the reduction of the oxidation windows a thinning of the field oxide occurs [3.29]. This thinning is more severe the smaller the windows. Intuitively this can be explained because the amount of oxygen that penetrates through a window is proportional to the window size. The consumption of oxygen to form the BB is the same for each window, independently from its size [3.30]. As a consequence the total amount of oxygen remaining to form the field oxide is reduced and its thickness will be smaller.

Figure 3.17 shows the results obtained for this LOPOS approach.

Two models have been proposed to explain the observed field oxide thinning in small apertures. Mizuno et al. [3.30] propose a two dimensional oxidation model. In their model, as no reduction of the BB length is observed, the thinning is simply correlated with the fact that the oxidation window size is reduced. In this way the amount of oxygen that is needed for the BB for-

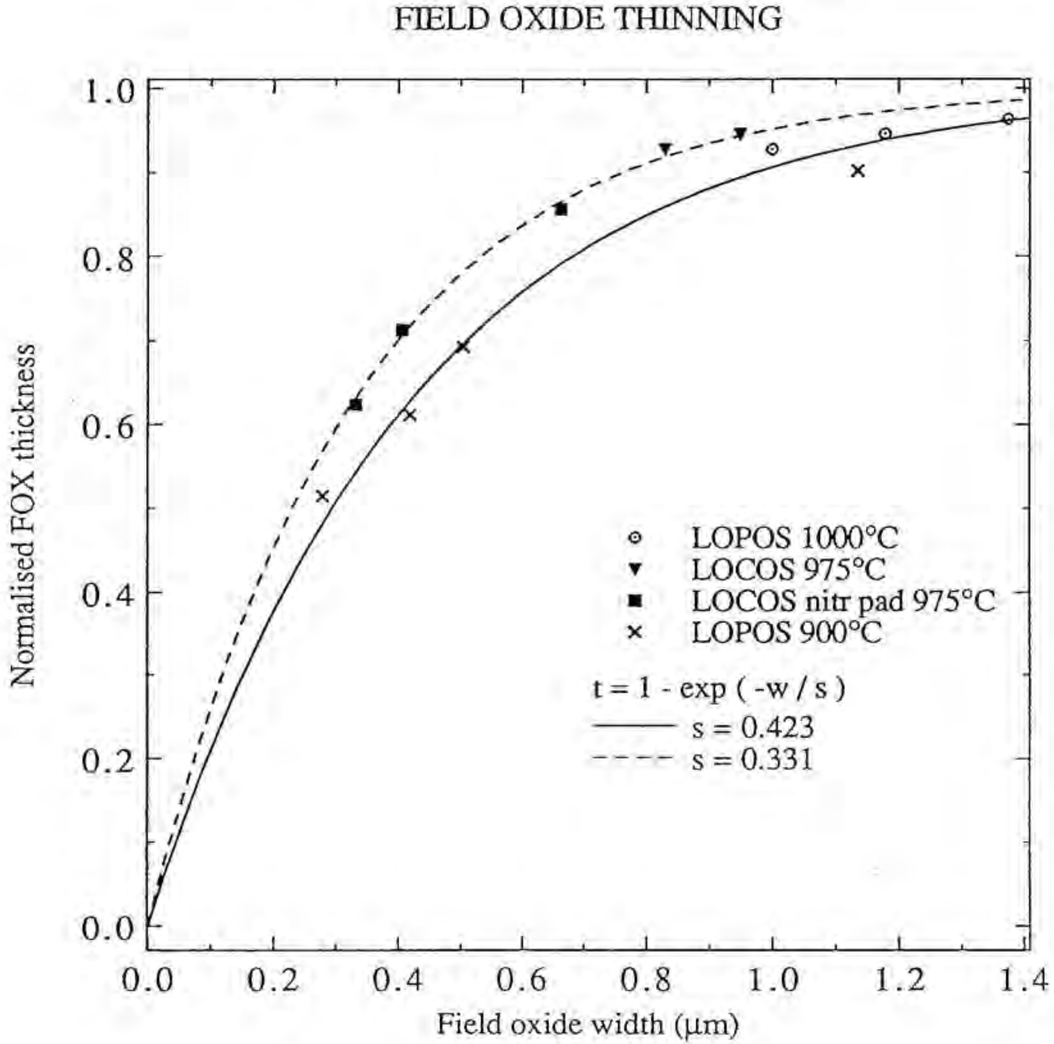


Figure 3.17: Field oxide thinning for small oxidation windows in the LOCOS and LOPOS processes. For LOPOS the thinning is more severe.

mation is fixed and the remaining amount of oxygen for the growth of the FOX will decrease with decreasing window size.

The second model presented by Lutze et al. [3.31], fits the experimental observations with a formula of the form:

$$t = 1 - \exp(-w / s)$$

where t is the normalized oxide thickness, w is the oxidation windows width and s is a characteristic length that takes into account the oxidation temperature and which can be correlated to the viscosity. According to them s can be calculated for each temperature. In their work no dependence on the nitride thickness nor on the pad oxide thickness was proposed, as these parameters were always constant. In our case various oxide and nitride thicknesses have been used, slightly changing the s values for a given temperature.

3.5 LOCOS ON SIMOX SUBSTRATES

3.5.0 Introduction

Up to now all the results on isolation that have been presented in this work were related to conventional silicon substrates. Nowadays large efforts are devoted to the processing on SOI (Silicon On Insulator) substrates. These substrates consist of a thin single crystal silicon film on top of the wafer, separated from the bulk by an isolating layer, commonly silicon oxide. These substrates present some interesting advantages:

1) The total-dose radiation hardness of MOS devices is, roughly, inversely proportional to the square of the thickness of the oxide layers in contact with the active silicon. In SIMOX devices the silicon layer is separated from the substrate by an oxide layer, which is typically 400 nm thick. For bulk silicon the field oxide is thicker. For this reason these substrates are more suitable for military and space applications.

2) The latch-up phenomenon, which is typical for integrated circuits on bulk silicon, does not occur. Latch-up is an interaction of two closely located but separated devices through the underlying bulk silicon. Due to the presence of the isolating buried layer in SOI wafers, full isolation is achieved.

3) For very thin top silicon films, MOS transistors fabricated on these substrates can be fully depleted [3.32]. This makes their speed much higher than the corresponding transistors on bulk silicon.

4) Increase of the packing density compared to similar processing performed on bulk silicon wafers.

Little information on the formation of selected isolated areas in SIMOX wafers can be found in the literature. Probably this is due to the fact that similar conditions as used for bulk wafers can be used and that no specific problems appear. Three major isolation techniques have been reported:

1) LOCOS [3.33], that has the advantage of simplicity, without covering problems. Drawback is, however, the rounded top surface close to the BB.

2) MESA [3.33, 3.34], consisting in the etching of the top silicon layer in the selected areas. For this it is necessary to use a highly anisotropic etchant, i.e., reactive ion etching. A drawback is that a spacer has to be grown to avoid the problems in covering of the steep sides.

3) Ion implantation of oxygen through polysilicon masks to form the buried isolating layer [3.35]. When removing the mask, in the masked areas the implanted volume reaches the surface, achieving complete isolation. In order to be able to use this approach it is necessary to have a high current ion implanter, only available in few laboratories.

3.5.1 Experimental conditions

In this section the bird's beak formation for a LOCOS process on thin film SIMOX substrates is studied as a function of the pad oxide and field oxide thicknesses. Bulk silicon wafers have been processed in the same conditions, so that they could serve as reference for the process on SIMOX and to determine the oxide thickness.

The wafers used for our experiment had a buried oxide layer 380 nm thick and a top silicon layer thinned down to 155 nm. This thinning has been obtained by oxidizing the silicon top layer, followed by removal of the formed oxide with HF. To study the LOCOS on SIMOX process, two different pad oxides, 22.5 and 42 nm, and two silicon nitrides, 120 and 150 nm, have been used. In all cases the field oxide was grown at 950°C.

The aspects that are of interest in this study are the BB length and geometry, the possible generation of defects or deformation of the top thin film due to the stresses involved in the process, as well as the effect of the overoxidation of the active layer and compare it with the result after just consuming the top film. For this reason two different field oxides have been grown, the first corresponding to the nominal value employed for submicron CMOS processing, about 600 nm thick, while the second was thinner, 340 nm. This thickness corresponds to stopping the oxidation process short after the complete the top silicon film has been oxidized.

3.5.2 TEM observations

For the discussion of the obtained results, the top silicon layer will be called active silicon, while the substrate below the buried oxide will be referred as bulk silicon.

All the observations that will be presented have been performed on cross-section specimens. The method used commonly for plan view preparation of SOI substrates, consisting in dissolving the buried layer in HF [3.36], cannot be used, because this would remove also the FOX, leaving floating silicon lines. These

lines would change their orientation during observation. Furthermore, even in cross-section, the orientation of the top silicon film is not exactly the same as for the bulk silicon and during prolonged exposure to the electron beam changes of orientation have been observed. This effect makes cross-section geometry studies of these wafers quite difficult. Commonly the bulk silicon was oriented in the electron microscope, assuming that the deviation of the top layer from the [110] zone axis was not too large.

The first result that has to be mentioned is the absence of any crystalline defect in the top silicon layer. This is as expected, because the LOCOS conditions used were very close to the ones optimized for conventional LOCOS on bulk silicon.

Figure 3.18a shows a cross-section view of a SIMOX wafer after the field oxide has been grown. The thicknesses of the different layers are: 42 nm pad oxide, 120 nm nitride and 340 nm field oxide. The length of the BB is about 400 nm. This long BB is due to the thin nitride that does not hamper the penetration of oxygen below it. As a consequence the top surface of the active silicon has been consumed by the FOX.

In this figure it is evident that the bulk silicon has not been oxidized much. The silicon islands, typical for a SIMOX wafer obtained by single implant, are always at the same distance from the bulk silicon throughout the complete length of the picture. The lower interface of the active silicon, however, is not flat close to the FOX. The drawn line indicates its original position. The lower interface has been oxidized once the field oxide reached the buried oxide. This process is performed with an increase of the volume of the buried oxide. Under this influence the active silicon bends upwards. This is clearly observed in the figure.

When oxidizing for longer times, the oxygen reaches, first, the silicon islands and, afterwards, the bulk silicon. In both cases oxidation of the silicon takes place. This is shown n figure 3.18b. Below the mask little oxidation of the bulk silicon took place, which can be seen by the presence of the islands.

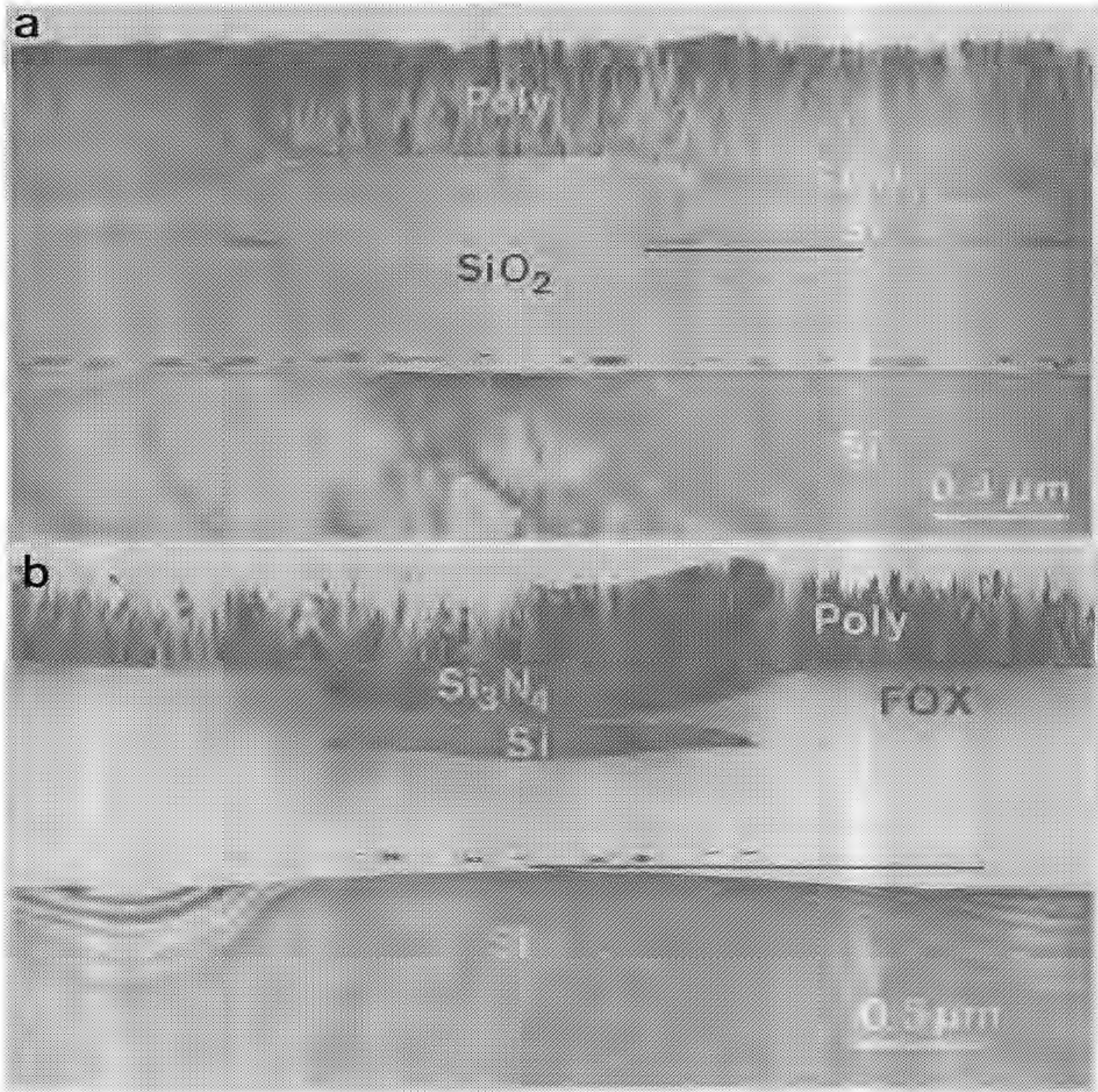


Figure 3.18: LOCOS process on SIMOX substrates. a) The employed parameters are: 42 nm pad oxide, 120 nm nitride and 340 nm field oxide. b) Oxidizing the top silicon for long times leads to the oxidation of the bulk silicon and of the lower interface of the active silicon. As a consequence the active silicon will be shifted upwards.

Below the opened windows the islands have been completely consumed, as well as part of the bulk silicon. It can clearly be seen that due to the volume increase the islands are shifted compared to their original position. The black line corresponds to the original bulk silicon-buried oxide interface. The distance between the active and the bulk silicon is larger than the original thickness of the buried layer, showing that the lower interface of the active silicon has also been consumed in the central part.

3.6 CONCLUSION

In this chapter one of the most important steps of silicon integrated circuit processing, the isolation, has been presented. Due to the problems that LOCOS presents for submicron technology, alternative isolation methods have been tested, all based in modifications of the standard LOCOS. It is important to test these methods because LOCOS is quite simple and gives reasonable results.

Three modifications have been presented. Two of them are a simplified version of the SILO process: nitridation of the pad oxide and a nitride layer in direct contact with the silicon substrate. The presented results show that the drawbacks of LOCOS are not solved by them.

In addition another drawback of LOCOS not reported in the literature, as far as we know, has been presented. It is the asymmetry of the bird's beak in a closely packed structure. The outer BB (and sometimes even the one close to it) being different from the others. This is extremely important for memories and other cells with a high number of transistors, because the electrical characteristics of the transistors at the edges of the structures will differ from the others, giving rise to problems in the performance of these devices.

The third modification of LOCOS, LOPOS, has been studied from a structural point of view, showing a quite good alternative to LOCOS. The use of the buffer layer, effectively, acts as

stress relief. Structures with two different buffer layers have been investigated. The conventional, using polysilicon, has shown that the BB slightly is reduced and that, for the studied conditions, no defects occur.

Furthermore a new LOPOS approach, obtained by inserting a microcrystalline layer between the pad oxide and the nitride, has been studied in more detail. During field oxide growth the microcrystalline silicon has been transformed into polysilicon. The obtained results show that during this process a strong reduction of the BB length takes place, coupled with a strong increase of the stresses at the edge of the BB takes place. This stresses are high enough to give rise to different relaxation mechanisms:

- Breaking of the pad oxide.

- Breaking of the nitride.

- Generation of defects in the substrate.

The appearance of one or the other effects has been observed to be dependent on the oxidation temperature. For low temperatures, about 950°C, breaking of the nitride occurs in combination with breaking of the pad oxide. At temperatures around 1000°C or higher, the third mechanism is the dominant, but breaking of the pad oxide can also take place. In some cases it has been observed that when generation of defects occur, medium range relaxation happens, dominating the two other stress relaxation mechanisms.

In the last section the conventional local oxidation on SIMOX wafers has been studied. The results show that this standard LOCOS process gives rise to a strong curvature of the top silicon layer at the bird's beak areas. This will give problems in the processing steps that follow the isolation. Oxidation for long times give rise to a lifting of the top silicon layer, and also to an overall orientation change with respect to the substrate.

REFERENCES CHAPTER 3

[**3.1**] J.A. Appels, E. Kooi, M.M. Paffen, J.J.H. Schatorje and W.H.G.G. Verkuylen, Philips Res. Repts. 25, 118 (1970).

[**3.2**] L. Jastrzebski, R. Soydan, J. McGinn, R. Kleppinger, M. Blumenfeld, G. Gillespie, N. Armour, B. Goldsmith, W. Henry and S. Vecrumba, in Reduced Temperature Processing for VLSI, eds. R. Reif and G.R. Srinivasan, The Electrochemical Society, Pennington, 405 (1986).

[**3.3**] J. Vanhellemont, Ph. D. thesis, University of Antwerpen (1990).

[**3.4**] T.C. Wu, W.T. Stacy and K.N. Ritz, J. Electrochem. Soc. 130, 1563 (1983).

[**3.5**] E. Kooi, J.G. van Lierop and J.A. Appels, J. Electrochem. Soc. 123, 1117 (1976).

[**3.6**] O. Nakajima, N. Shiono, S. Muramoto and C. Hashimoto, J. Appl. Phys. 18, 943 (1979).

[**3.7**] L.O. Wilson, J. Electrochem. Soc. 129, 831 (1982).

[**3.8**] H. Goto, T. Takada, R. Abe, K. Kawabe, K. Oami and M. Tanaka, IEDM Tech. Dig., 58 (1982).

[**3.9**] R.D. Rung, H. Momose and Y. Nagakubo, IEDM Tech. Dig., 237 (1982).

[**3.10**] K.Y. Chiu, J.L. Moll and J. Manoliu, IEEE Trans. Electron Dev. ED-29, 536 (1982).

[**3.11**] C. Claeys, J. Vanhellemont, T. Cavioni and F. Gualandris, J. Electrochem. Soc. 136, 2619 (1989).

[**3.12**] K.L. Wang, S.A. Saller, W.R. Hunter, P.K. Chatterjee and Y.P. Yang, IEDM Tech. Dig., 372 (1981).

[**3.13**] K. Kurosawa, T. Shibata and H. Iizuka, IEDM Tech. Dig., 384 (1981).

[**3.14**] J.C. Hui, T.Y. Chin, S.W. Wong and W.G. Oldham, IEEE Trans. Electron Dev. Lett. EDL-2, 244 (1981).

[**3.15**] J.C. Hui, T.Y. Chin, S.W. Wong and W.G. Oldham, IEEE Trans. Electron Dev. Lett. ED-29, 554 (1982).

[**3.16**] Y. Han and B. Ma, Electrochem. Soc. Extended Abst. 84-1, 98 (1984).

[**3.17**] M. Ghezzo, E. Kaminsky, Y. Nissan-Cohen, P. Frank and R. Saia, J. Electrochem. Soc. 136, 1992 (1989).

[**3.18**] C.A. Paz de Araujo, Y.P. Huang and R. Gallegos, J. Electrochem. Soc. 136, 2035 (1989).

[**3.19**] S.S. Lee and K.P. Fuchs, in Semiconductor Silicon 1990, eds. H.R. Huff and K.G. Barraclough, The Electrochemical Society, Pennington, 850 (1990).

[**3.20**] A. Romano, L. Deferm, J. Vanhellemont and C. Claeys, Electrochem. Soc. Extended Abst. 90-2, 427 (1990).

[**3.21**] J. Vanhellemont and C. Claeys, J. Electrochem. Soc. 135 (6), 1509 (1988).

[**3.22**] J. Vanhellemont, S. Amelinckx and C. Claeys, J. Appl. Phys. 61, 2176 (1987).

[**3.23**] J. Vanhellemont and L. Van den Hove, Inst. Phys. Conf. Ser. 93 (2), 423 (1988).

[**3.24**] E. Anastassakis, A. Pinczuk, E. Burstein, F.H. Pollak and M. Cardona, Solid State Commun. 8, 133 (1970).

[**3.25**] K. Kobayashi, Y. Inoue, T. Nishimura, M. Hirayama, Y. Akasaka, T. Kato and S. Ibuki, J. Electrochem. Soc. 137, 1987 (1990).

[3.26] I. De Wolf, J. Vanhellemont, A. Romano-Rodríguez, H. Norström and H.E. Maes, submitted for presentation at the Electrochemical Society Spring Meeting, 1991.

[3.27] A. Romano-Rodríguez, J. Vanhellemont, I. De Wolf, H. Norström and H.E. Maes, accepted for presentation at the "VII Oxford Conference on Microscopy of Semiconductor Materials", March 1991.

[3.28] S. Ganesan, A.A. Maradudin and I. Oitmaa, Annals of Phys. 56, 556 (1970).

[3.29] J. Hui, P. Vande Voorde and J. Moll, IEDM Tech. Dig., 392 (1985).

[3.30] T. Mizuno, S. Sawada, S. Maeda and S. Shinozaki, IEEE Trans. Electron Dev. ED-34, 2255 (1987).

[3.31] J.W. Lutze, A.H. Perera and J.P. Krusius, J. Electrochem. Soc. 137, 1867 (1990).

[3.32] J.P. Colinge, IEDM Tech. Dig., 817 (1989).

[3.33] M. Haond, Proc. IEEE SOS/SOI Technol. Conf. 117 (1990).

[3.34] O. Le Néel, M.D. Bruni, J. Galvier and M. Haond, Proc. ESSDERC 1990, 13 (1990).

[3.35] J.R. Davis, A. Robinson, K.J. Reeson and P.L.F. Hemment, Appl. Phys. Lett. 51 (18), 1419 (1987).

[3.36] A. De Veirman, J. Eysermans, H. Bender, J. Vanhellemont and J. Van Landuyt, Mater. Res. Soc. Symp. Proc. 115, 241 (1988).

4. JUNCTION DELINEATION

4.0 INTRODUCTION

The delineation of p-n junctions of electronic devices is one of the techniques routinely required in the microelectronics industry and development. For this purpose several characterisation techniques with high accuracy have been developed in the last years and are nowadays commonly used. The main two of them are SRP and SIMS. As mentioned in chapter 1, these two techniques yield complementary information: SIMS gives the total amount of a certain atom which is present in the substrate as a function of depth, while SRP gives the amount of free carriers in the specimen, i.e., the amount of dopant which is electrically active. A drawback is that both techniques have a poor lateral resolution. The probe used for the SRP measurements is relatively large (about 2 μm) and the separation between the two probes is about 20 μm. For SIMS the lateral resolution, in an optimized configuration, is of the order of 1 μm. As the lateral diffusion of the dopant species is of the same order as the diffusion in depth, a restriction exists in the available techniques for two-dimensional profiling of ULSI integrated circuits.

The techniques employed for two-dimensional junction delineation or dopant profiling have to fulfil some requirements. They are summarised in the following points:

- High spatial resolution in depth as well as laterally.

- High reproducibility rate.

- Independence of the morphology of the sample.

- Non destructive method.

During the last years some techniques have been designed and implemented, which accomplish some of the mentioned points. In the following paragraphs the most relevant of them are

briefly summarised, with a short explanation of their principles.

1) Angle lapping and staining of the device and observation in an optical microscope. This technique has been applied routinely in the last three decades and is very powerful for devices with large junction depths, typically over one micron. For submicron junctions, however, optical microscopy has an insufficient resolution power. Some improvements have been realized lately [4.1], but the results are not very convincing as special structures have to be used, the size of which should be of the order of some mm in one direction.

2) Capacitance-voltage measurements, which is a technique routinely used for one-dimensional carrier profiling and that has been extended recently to get also two-dimensional information [4.2].

3) Planar anodisation followed by measurement of the conductance variation [4.3], which, again, requires special test structures and, thus, cannot be applied to finished devices.

3) SEM (Scanning Electron Microscope) in combination with preferential chemical etching of the specimen, which has become a routine technique nowadays. This approach has two main advantages. First is the simplicity in the sample preparation, as it is only necessary to cleave the wafer. The second advantage is that the analysis can be performed on finished devices. Drawback is, however, the difficult calibration of the magnification of the SEM, which is critical for the steep dopant profiles used in ULSI technology. Consequently, only qualitative information about the dopant profile can be obtained, although lately some improvement has been made [4.4].

4) The previous etching technique can also be applied to TEM specimens, having the advantage of the very high spatial resolution of these microscopes [4.5-4.8]. Drawback is the still poor reproducibility.

5) TEM in the STEM mode in combination with an energy dispersive X-ray detection system (EDX) [4.9]. In this case the delineated level is about $5 \cdot 10^{18}$ cm^{-3}.

6) Use of a Scanning Tunneling Microscope (STM) on a cleaved cross-section [4.10]. According to the authors this method should give the position of the junction with an accuracy better than 50 nm.

Furthermore some attempts have been made in order to expand the two main one-dimensional techniques, SRP and SIMS, so that two-dimensional information could be obtained. In this moment preliminary results on a two-dimensional SRP technique have been obtained [4.11]. For these measurements special structures are required, making the approach not suited for device level. Similarly two-dimensional SIMS has been implemented [4.12], requiring again special structures.

In this work two different approaches have been used in order to study the use of TEM techniques for the visualisation of dopant profiles.

The first one has been briefly explained in point 4, i.e., use of chemical etching. Different chemical solutions have been studied in order to see the delineation capability of each of them. The procedures employed and the results are presented in section 4.2.

A second method which has been studied is based on the fact that during prolonged observation of the cross-section samples in the HVEM, radiation induced extended defects are generated. The distribution of these defects has been observed to be dopant concentration dependent. To our knowledge no work has been published reporting this effect. Some experiments have been carried out, using different electron currents and temperatures to see whether these parameters have any influence on the generation of defects. The results are presented in section 4.3.

In all the experiments presented in the following sections, comparison of the delineated depth has been made either with

SIMS or with SRP (when that depth corresponded to levels above the concentration where an important effect of the carrier spilling is expected, as mentioned in chapter 1). For some specific samples, where these measurements were not available, process simulation has been used for comparison. In this last case the obtained results have only been considered in a qualitative way.

4.1 STARTING MATERIAL

The starting substrates used for this study were Czochralski grown [001] silicon wafers, doped with phosphorus or boron to concentrations typically between $2 \cdot 10^{14}$ cm^{-3} and $1 \cdot 10^{15}$ cm^{-3}. Implantation of arsenic or phosphorus in p-type or boron in n-type substrates has been performed through a thin screening oxide, followed by annealing to recover the destroyed silicon lattice and to activated the dopants. Some wafers have been implanted through a polysilicon or oxide mask. After annealing diffusion of the doping species occurs both in depth and laterally. In this way the ability of two-dimensional dopant profiling of the methods that will be presented in the following sections can be tested.

Also some wafers with a well controlled and quite steep dopant profile, obtained by growth in an epitaxial reactor, were used.

For the calibration of the different etching solutions uniformly doped wafers were used.

These wafers allow to perform two types of analysis:

1) The wafers which did not receive any pattern have been used to calibrate the etching solutions. If a pattern is present the localized specimen preparation technique has to be employed, because of the low density of structures.

2) The patterned wafers, when thinned in the area close to the bird's beak tip or to the polysilicon edge, allow to study

the lateral diffusion of the doping species below the implantation masks.

4.2 TEM COMBINED WITH PREFERENTIAL ETCHING

Silicon of p- and n-types give the same contrast in the TEM if the dopant concentration is below 10^{19} cm^{-3}. Recently it has been reported that for concentrations of boron above 10^{19} cm^{-3} and for arsenic above 10^{20} cm^{-3} the change in the lattice parameter of the silicon can be appreciable. This gives rise to a contrast difference [4.13], when imaging the specimen in the adequate conditions. For lower doped regions, in order to distinguish between two different doping levels it is necessary to introduce a contrast difference. This should be dependent on the dopant concentration and should be sensitive to the presence of a junction.

As has been pointed out in the introduction to this chapter, a technique routinely used in SEM analysis is to introduce a difference of contrast by selectively etching the sample that has to be studied. A similar technique can be applied to the TEM. In the SEM the etching has to be quite severe in order to introduce a considerable height difference between differently doped areas. For TEM application this etching has to be quite light, because the thickness of the specimen which is useful for investigation is, in the best case, of the order of 5 µm, and commonly below 1 µm.

The characteristics which are required for "good" etchants for TEM cross-sections are summarised below:

1) The etching rate has to be relatively small, so that the amount of etched material can be controlled quite accurately.

2) The etching time should be long enough, let's say longer than 20 seconds, so that the procedure can be easily reproduced from one specimen to another.

3) The etchant should attack mainly the silicon substrate and very little the other layers that compose the TEM specimen,

in order to be able to see the complete structure after etching without any loss of information.

4) The etchant behaviour should depend little on the morphology of the specimen before etching, i.e., quite insensitive to surface roughness or to specimen aging.

In the next paragraphs two chemical systems, which have been used successfully, will be described, indicating their advantages and drawbacks. In both cases the etching method consists in preparing a TEM cross-section specimen (using the technique explained in chapter 2) and etching it with the desired solution. If the etching rate is dopant dependent, in the final geometry variations of the thickness fringes of the silicon will be seen as a consequence of variations in the thickness of the specimen. Commonly this behaviour is only obtained for highly doped areas, while for lowly doped areas the etching rate is constant. After etching the cross-section, a strong thinning will be found corresponding to the highly doped regions. For the rest no thickness difference will be found. The transition region between these two areas will be visible in the image. Comparing the position of this region with the corresponding SIMS or SRP profile it is possible to obtain the delineated level.

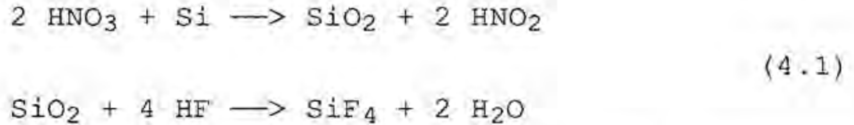
All the chemical solutions that have been used are based on the combined action of an oxidizing agent, that brings the elemental silicon to some higher oxidized state, and a complexing agent, that brings the oxidized silicon into solution. The most suitable oxidizing agents are: nitric acid, sodium dichromate, periodic acid, chloric acid, ... Some of the complexing agents are: hydrofluoric, oxalic, citric and tartaric acids, ... Sometimes also a motoring agent is used, that helps to control the rate of the chemical reactions. The two most common motoring agents are water and acetic acid.

In this work different oxidizing agents have been used, while only hydrofluoric acid as complexing agent and, if any, the motoring agent has been water.

4.2.1 HF-HNO_3 system

4.2.1.1 Etching behaviour

Combinations of HF and HNO_3 have been used for a long time to etch and polish silicon substrates. Formula 4.1 gives the chemical reactions that take place during the etching of silicon.



$$2\ HNO_3 + Si \longrightarrow SiO_2 + 2\ HNO_2$$

$$SiO_2 + 4\ HF \longrightarrow SiF_4 + 2\ H_2O \qquad (4.1)$$

An extended study of the etching behaviour of solutions containing HNO_3 and HF, using water or acetic acid as dissolvent, has been presented by Schwartz and Robbins [4.14]. Their results are presented in figure 4.1. The main conclusions that can be deduced from their study are:

- The etching rate is only dependent on the composition of the used solution and not on the dopant concentration.

- There is no difference between the etching behaviour of p- and n-type silicon.

- The difference in etching rate obtained when etching samples with a junction is only due to the presence of the junction itself.

The etching rates given in their work are above 10 µm/min, which is too high to be used to delineate junctions on TEM cross-section samples. If no motoring agent is used the only compositions that give low etching rates are located at the top and bottom-right edges of the triangle of figure 4.1. The top, corresponding to the very high concentrations of HF, should be avoided because HF etches the oxides very quickly. Thus, it is better to employ solutions with a low concentration of HF.

More recent studies by Roberts et al. [4.6] and Gold et al. [4.7] showed that the etching rates obtained by using solutions with concentrations of HF (40% in volume) below 0.5% are a clear

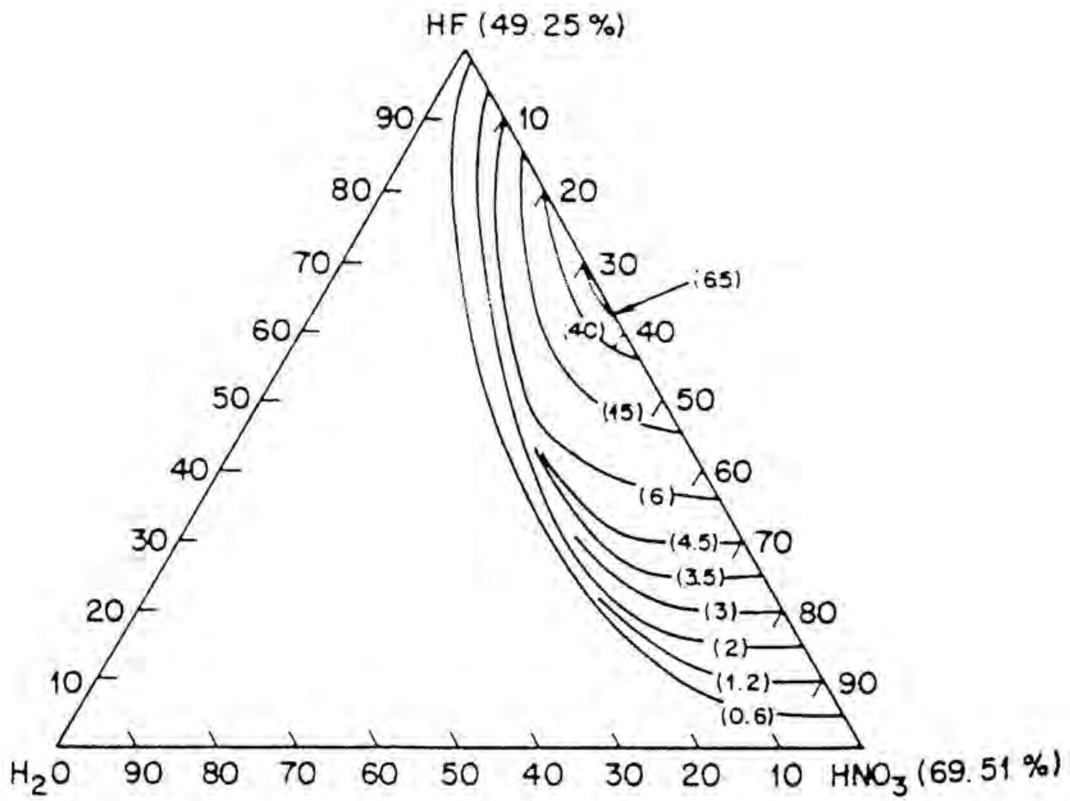


Figure 4.1: Etching behaviour of the $HF-HNO_3$ system [4.14].

function of the doping level for concentrations above 10^{17} cm^{-3}. Below this level the dependence is not clear.

In this work similar measurements of etching rates as a function of the doping level have been performed and have been compared with the ones of Gold et al. [4.7]. For this purpose silicon substrates [001] oriented, with uniform doping level, determined by SRP and four point probe, both p- and n-type have been used. First the silicon substrates have been cleaved into smaller squares, typically 2 x 2 mm^2. These pieces have been dipped in HF to remove the native oxide from the surface. The following step consists in glueing the silicon substrates to a polypropylene piece with flat sides using Lacomit. Both Lacomit and this plastic are not attacked by the etchant. Half of the polished surface is covered with Lacomit to protect it against etching. Afterwards this part will serve as reference point to measure the amount of removed material. This procedure is performed always with a piece of every wafer at the same time, so that changes due to variation of the experimental conditions, like temperature, concentration of the acids, ..., are minimised.

The solution employed in this work is the same as the one used by Roberts et al. [4.6], i.e., $HF:HNO_3=1:300$ and working at a temperature of 5°C. It seems that the use of lower temperatures allows to delineate a lower level. The calibration of the etchant has been performed as follows:

1) A large volume, typically 250 cm^3, of HNO_3 is put on a plastic beaker. This, at its turn, is placed in a larger one that contains water and ice, so that the temperature of the bath is about 0°C. Due to the bad thermal conductivity of the plastic beaker the temperature of the solution is always higher.

2) Using a thermometer the temperature of the HNO_3 as well as its variation with time are measured. After some experiments it has been deduced that close to the etching temperature the variation is of the order of 0.25°C per minute.

3) When the thermometer indicates 5.5°C, it is taken out of the bath (because the glass of which the thermometer is made is attacked by the HF), the HF is added to the HNO_3 and the solution is mixed with a plastic stick.

4) After two minutes the polypropylene piece, containing the wafer pieces that have to be etched, attached to a plastic stick, is introduced in the solution. The etching is performed for two more minutes using manual agitation by rotating the stick. The stirring is important because the solution is very diluted in HF. During the etching the concentration of HF close to the wafer surface diminishes very quickly, lowering the etching rate if no fresh HF is supplied.

5) The polypropylene piece is taken out of the solution and immediately introduced in distilled water to stop the etching process.

6) The plastic piece is then introduced in a beaker with acetone, which dissolves the Lacomit. The wafer pieces are cleaned with methanol and dried with nitrogen gas.

7) Using a fine style profilometer (in our case Dektak), the step difference between the etched and the protected parts is measured and from this value the etching rate can be deduced.

Figure 4.2 shows the results obtained in this work, performed at 5°C, compared with the ones by Gold et al. [4.7], which were performed at room temperature. The former does not indicate which was the orientation of the material employed. The etching behaviour of [110] oriented specimens should be the same as for [001], as these types of etchants are known to be isotropic. Only the etching rate is expected to be different. This result is confirmed by the etching of a cross-section prepared along a [110] direction (the normal observation direction) and one prepared along a [100] direction, i.e., 45° away.

4.2.1.2 Delineation with this system

The first work published about the application of this technique was en 1981 by Sheng and Marcus [4.5]. They used a so-

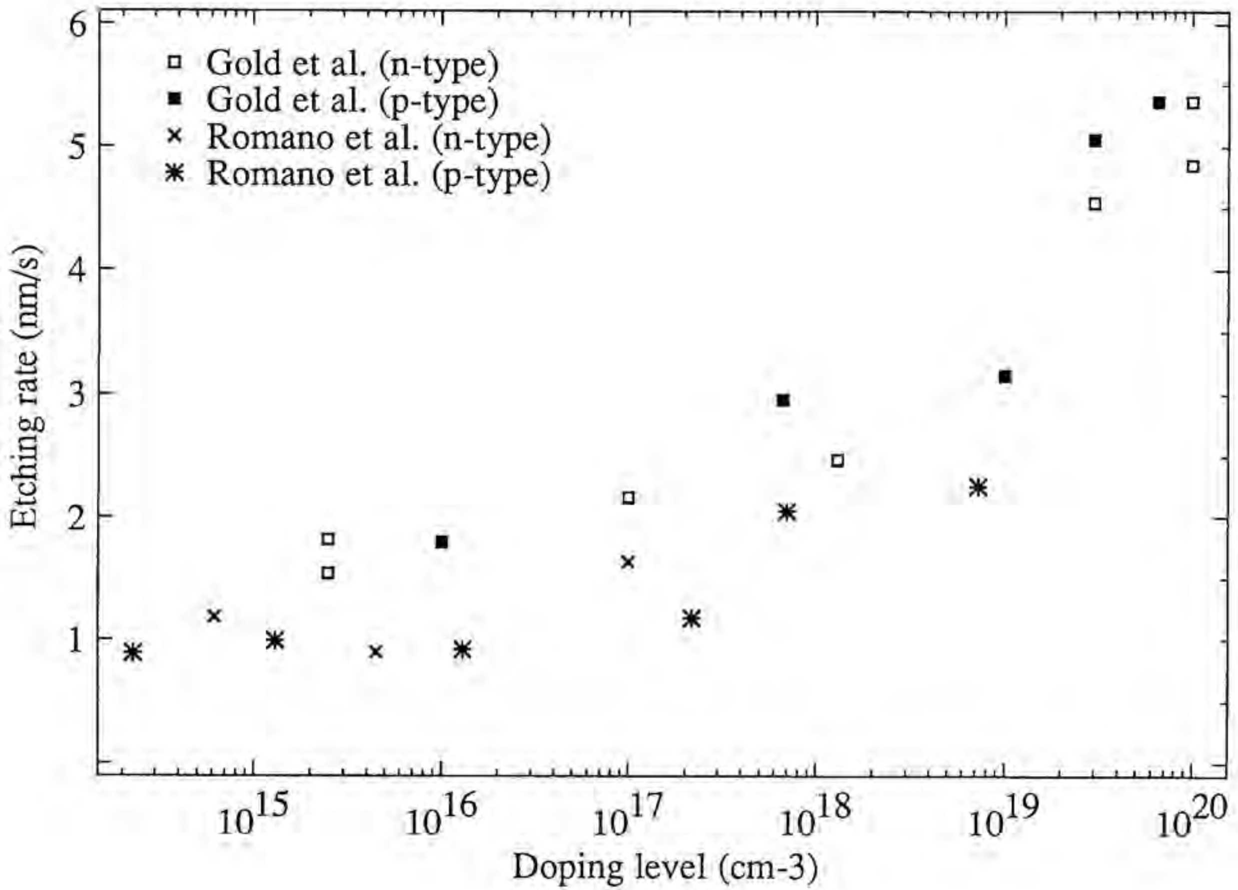


Figure 4.2: Etching rates for the "Roberts" solution obtained in this work, performed at 3°C [4.22]. For comparison the results of Gold et al. [4.5] are also plotted. These authors measured the etching rate at 21°C. It can be deduced that for doping concentrations below 10^{17} cm^{-3} the etching rate is not any more dopant dependent.

lution composed by 99.5% of HNO_3 (69.5% in volume) and 0.5% of HF (40% in volume) at room temperature. A cross-section specimen, arsenic implanted and annealed, was etched for about 5 seconds, leading to its delineation. By comparison the delineated depth with simulations of SUPREM II, a level of the order of 10^{19} cm^{-3} was assigned.

In the present work the delineation has been performed using the solution and conditions close to those proposed by Roberts et al. [4.6]. The first thing that has to be taken into account is that the HNO_3 dissolves strongly the copper grid that serves as support to the cross-section specimen. In addition, copper ions, dissolved from the grid, can strongly influence the delineation procedure, as copper based solutions are widely used for staining junctions. To avoid this problem the grid will be covered totally with Lacomit, leaving exposed only the cross-section specimen.

Once the Lacomit is dry (which takes about 20 minutes), the specimen is taken with anti-acid tweezers and dipped in HF for a few seconds, in order to remove any surface oxide. Next it is rinsed in deionized water. The cross-section specimen is cleaned with methanol and dried with nitrogen gas before the etching procedure begins. Finally it is introduced in the solution, which has been prepared in the same way as indicated in the previous section, and etch for times ranging between 80 and 120 seconds. The same procedure as before is used to stop the etching and to clean the specimen, which can be introduced in the microscope immediately after drying.

Figure 4.3 shows a cross-section specimen which has been etched for 80 seconds. This specimen has been prepared from an unpatterned wafer which was implanted with boron ions and annealed at 900°C for 30 minutes. The oxide through which the implantation has been performed (called here gate oxide) is visible and has not been etched away during the delineation thanks to the very low concentration of HF. A complex thickness fringes structure can be observed in this figure, which has been taken in the very thin areas of the specimen. The delineated level is

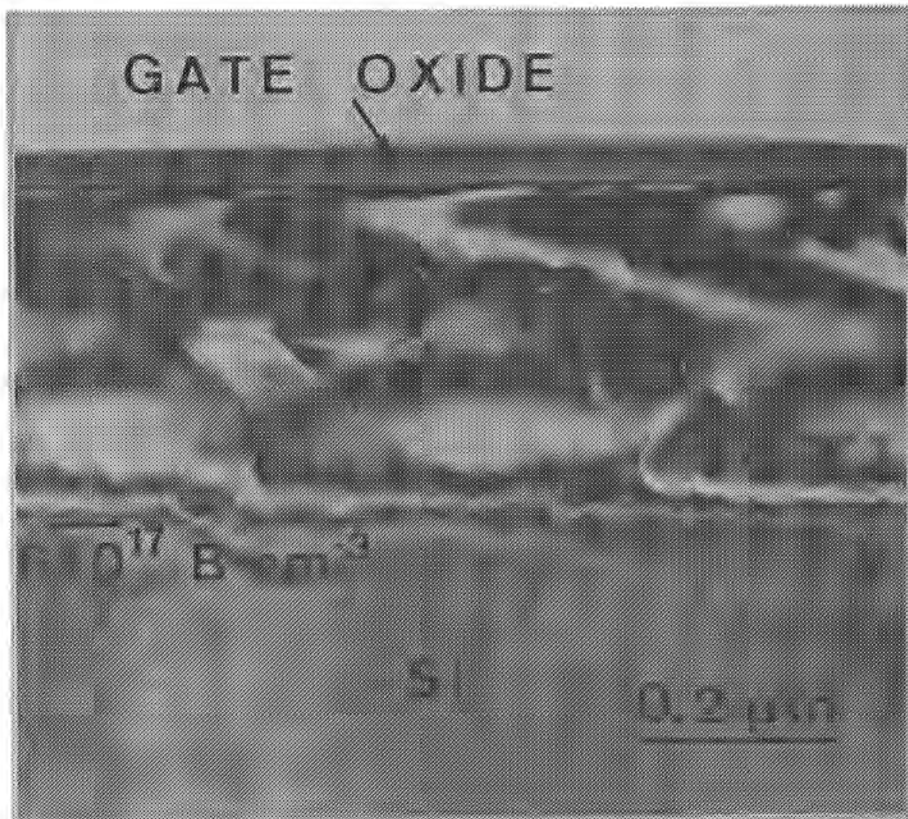


Figure 4.3: Cross-section specimen of a boron implanted wafer, etched for 80 seconds using the "Roberts" solution at 5°C. The delineated depth corresponds to a doping level of $6 \cdot 10^{17}$ cm^{-3}, deduced from comparison with SRP.

indicated in the picture, $6 \cdot 10^{17}$ cm^{-3}, and has been deduced by comparison with SIMS profile.

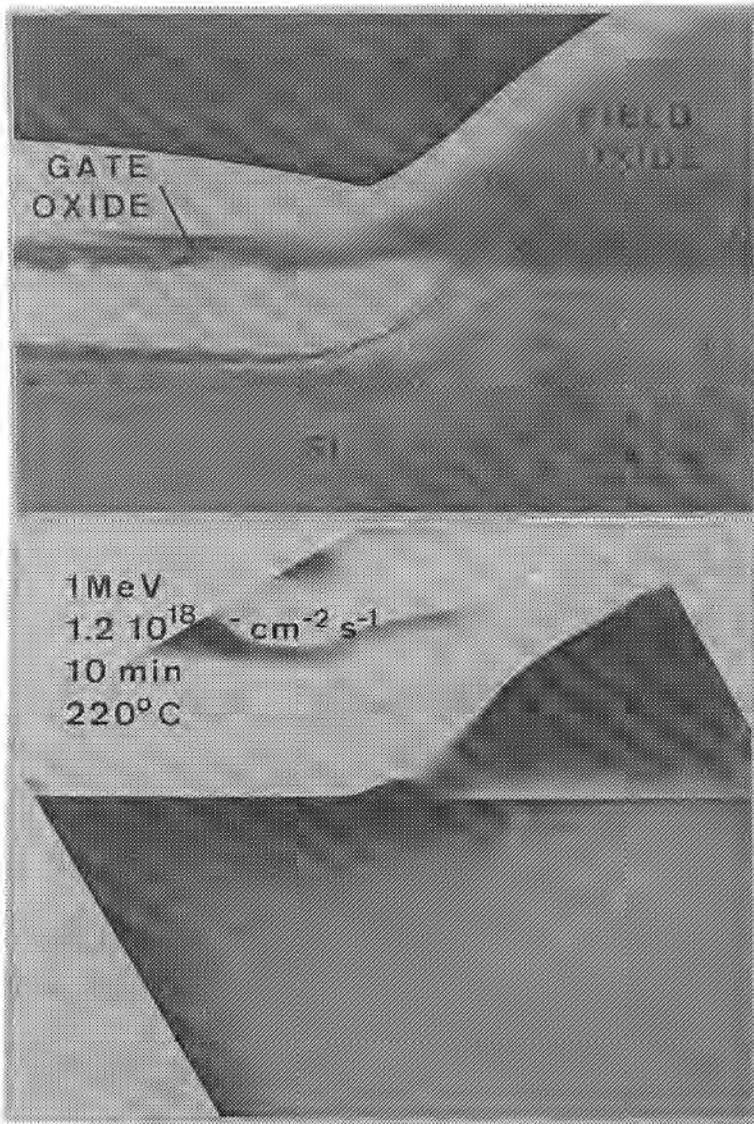
This figure has been taken approximately 2° off the [110] zone axis, tilting around an axis perpendicular to the surface of the specimen and using a $\bar{g}_{220}$ vector. In this way no overlapping of differently doped areas takes place and the contrast is enhanced compared to a symmetrically oriented specimen. Extrinsic dislocation loops, not removed during the anneal, are present but they do not influence the etching behaviour.

Roberts et al. [4.6] proposed a different method to obtain more information from an etched specimen. Their starting cross-section was a nearly perfect plan-parallel specimen, with only 1° inclination. The etching would give rise to the same thinning at levels with the same dopant concentration. These, at its time, would lead to thickness fringes and each fringe would be related to a certain doping level. This interpretation would be very interesting, but the reproducibility is very poor, as it is extremely difficult to control the slope of the cross-section.

Figure 4.4a shows the result of the use of this solution when a two-dimensional structure is present [4.8]. The wafer has been implanted with boron ions through an oxide mask. The specimen is perfectly oriented in the [110] zone axis. Two isoconcentration contours can be observed. The lowest corresponds, as before, to $6 \cdot 10^{17}$ cm^{-3}, while the second corresponds to $7 \cdot 10^{18}$ cm^{-3}, both obtained by comparison with the SRP profile of figure 4.4c. The second contour has been observed in this particular picture, but repeated experiments lead only to the delineation of the lowest level in a repetitive way. The higher one is strongly dependent on the specimen geometry prior to etching and changes, therefore, from specimen to specimen.

4.2.1.3 Limitations of this solution

From the results of figure 4.2 it can be deduced that this solution is only suited for delineation of doping levels above 10^{17} cm^{-3}. Below this value the difference in etching rate between differently doped regions is too small to produce a thick-



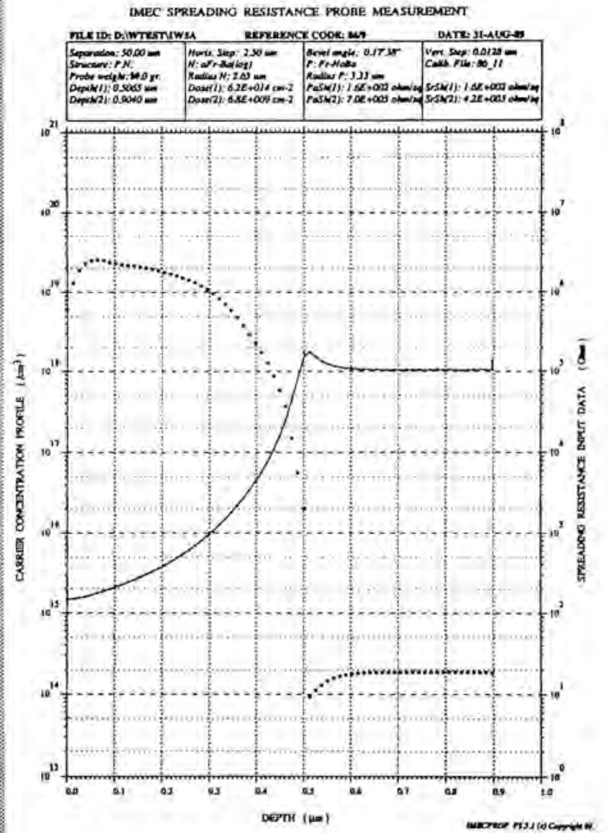


Figure 4.4: a) Delineation of a boron implanted cross-section specimen with a two-dimensional dopant profile. The Roberts solution has been used. Two levels can be seen, corresponding to $6\cdot10^{17}$ and $7\cdot10^{18}$ cm^{-3}. b) Result after 10 minutes electron irradiation in the HVEM at 220°C. The lower limit of the damage region corresponds to $6\cdot10^{17}$ cm^{-3}. c) SRP profile used for correlation of the delineated depth with the dopant concentration.

ness difference and to give rise to an appreciable contrast difference. This is confirmed by the TEM results, which show that the lowest possible delineated level is $6 \cdot 10^{17}$ cm^{-3}. This value is independent of the crystal orientation.

Similar results in the etching rate have been obtained for solutions with other HNO_3/HF concentration ratios.

Drawbacks of solutions of these type are that:

a) They age quite quickly, changing their etching behaviour when the solution is employed some time after the preparation.

b) As the concentration of the chemicals and the stabilising agents vary from lot to lot, changes in the etching rate can occur.

c) The copper grid has to be protected against the nitric acid or other type of grids have to be used. These are never as thin as the copper ones, increasing the ion milling time because of the shift away from the focus of the ion beams.

All these drawbacks force to look to a different chemical system or to completely different delineation methods in order to reach lower doping levels and the metallurgical junction.

4.2.2 CrO_3-HF system

4.2.2.1 Etching behaviour

Since the early sixties the use of solutions based on the combined action of the Cr(VI) ion and HF has been reported. Such solutions have been employed for the delineation of structural defects, like dislocations, stacking faults and others (Sirtl [4.15], Secco [4.16], Schimmel [4.17], Yang [4.18], ...).

It has been reported in the literature that solutions of this type are able to delineate silicon layers with different dopant concentrations (Chu and Gavaler [4.19], Maszara et al. [4.20]). Chu and Gavaler cleaved the specimens, followed by etching of the exposed surface and observation in a conventional optical microscope. Maszara et al. obtained their by bevelling

and etching the bevelled region and observing the etched specimen in an optical microscope using differential Nomarski contrast (this is the first delineation method described in introduction of this chapter). Advantage is the use of a Nomarski microscope, which gives a higher resolution than the conventional optical microscope. Both works report that the n-type regions are etched much faster than the p-type ones.

Relatively recently an extensive study about the etching behaviour of the CrO_3-HF system has been presented [4.21], investigating the etching rate as well as the anodical dissolution as a function of the concentration of the acids. Their results, again, have been performed on uniformly doped silicon [001] material, with doping levels of the order of 10^{14} cm^{-3}. No difference in etching rate between the two types of materials could be observed, which might be due to the absence of a junction and to the low doping levels.

The only information that is available is that the etching rate of the solutions mentioned above is of the order of 1.5 µm/min. Again this value is too high for our purposes. But it is a good starting point for the experiment, as it will indicate if the solution can be used or not.

4.2.2.2 Delineation experiments

The first experiment carried out consisted in repeating the same procedure as for the HF-HNO_3 system, i.e., etching a cross-section, with the advantage that in this case the etching is performed at room temperature and that the grid is not attacked by the acids of the solution. Using the Secco [4.16] and Schimmel [4.17] etches, cross-sections have been attacked for 5 seconds, giving as result no visible etching. Repeated experiments were performed in order to see whether this result was due to practical problems, but they revealed exactly the same.

These solutions are known to attack much more efficiently [001] oriented substrates, which means that our normal cross-section can not be used for these studies, because it is prepared along a [110] direction. For this reason a cross-section is prepared cutting the substrate along two perpendicular [100]

directions, i.e., 45° related to the conventional cutting direction. For the rest the preparation procedure continues exactly as described in chapter 2.

First standard solutions were used. Again Secco etch (was used for 5 seconds on a finished cross-section. Figure 4.5a shows a micrograph of a phosphorus implanted specimen prepared along a [100] direction, etched and observed along a [110] direction, i.e., rotated in the microscope 45° [4.22]. A level of $3 \cdot 10^{18}$ cm^{-3} is visible, as deduced from the SIMS profile. Deeper in the bulk, however, a faint line can be observed, which might correspond to the metallurgical junction when compared with the extrapolation of the SIMS profile. For comparison figure 4.5b shows a similar specimen, prepared and observed in the same conditions as previously but etched with the Roberts solution for 80 seconds. Comparing the delineated depth with the SIMS profile, a level of $6 \cdot 10^{17}$ cm^{-3} is obtained.

All these solutions present two drawbacks for delineation studies. First, the concentration of HF is too high, attacking rapidly the thin oxides. Second, the etching rate is too high. According to reference [4.21] a strong reduction of the amount of HF would lead to a substantial decrease of the etching speed. A compromise has been found to be a solution composed of $K_2Cr_2O_7$ 0.15M and HF 3M. The etching rate should be 0.85 nm/s. In this conditions etching for 2 minutes should remove the same amount of silicon than the Roberts etch.

The cross-section specimen of figure 4.6, implanted with a low dose of phosphorus, has been etched for 2 minutes using the mixture described in the previous paragraph. The peak of the dopant profile is below 10^{17} cm^{-3}, which is below the upper delineated level of figure 4.5a. A line is seen at a depth of 0.6 µm, which is beyond the resolution of SIMS, and which might correspond to the metallurgical junction.

Figure 4.7a shows the result of the use of the same solution on a cross-section of a boron doped wafer. 100 keV boron ions have been implanted and the sample annealed at 1000°C for 30 minutes. These conditions give rise to a buried dopant pro-

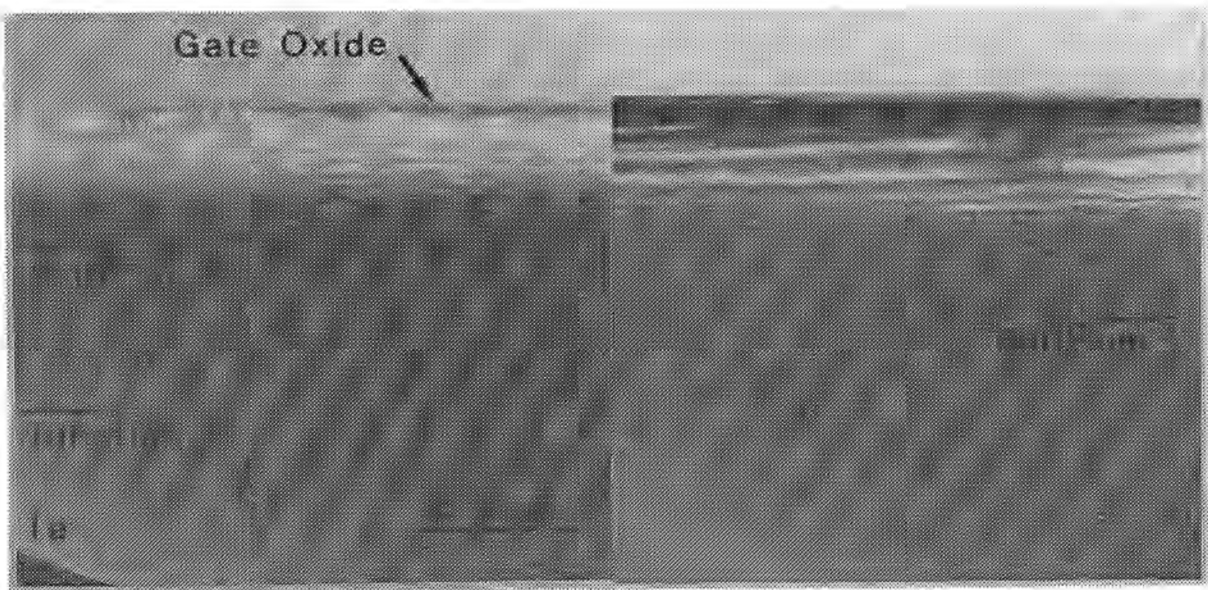


Figure 4.5: a) Phosphorus implanted cross-section specimen, prepared along a [100] direction, etched with Secco for 5 seconds and viewed along a [110] direction. A doping level of $3\cdot10^{18}$ cm^{-3} has been delineated. Deeper in the bulk a faint line can be seen, corresponding to the junction. b) Similar specimen etched with the "Roberts" solution for 80 seconds. The delineated level is again $6\cdot10^{17}$ cm^{-3}.

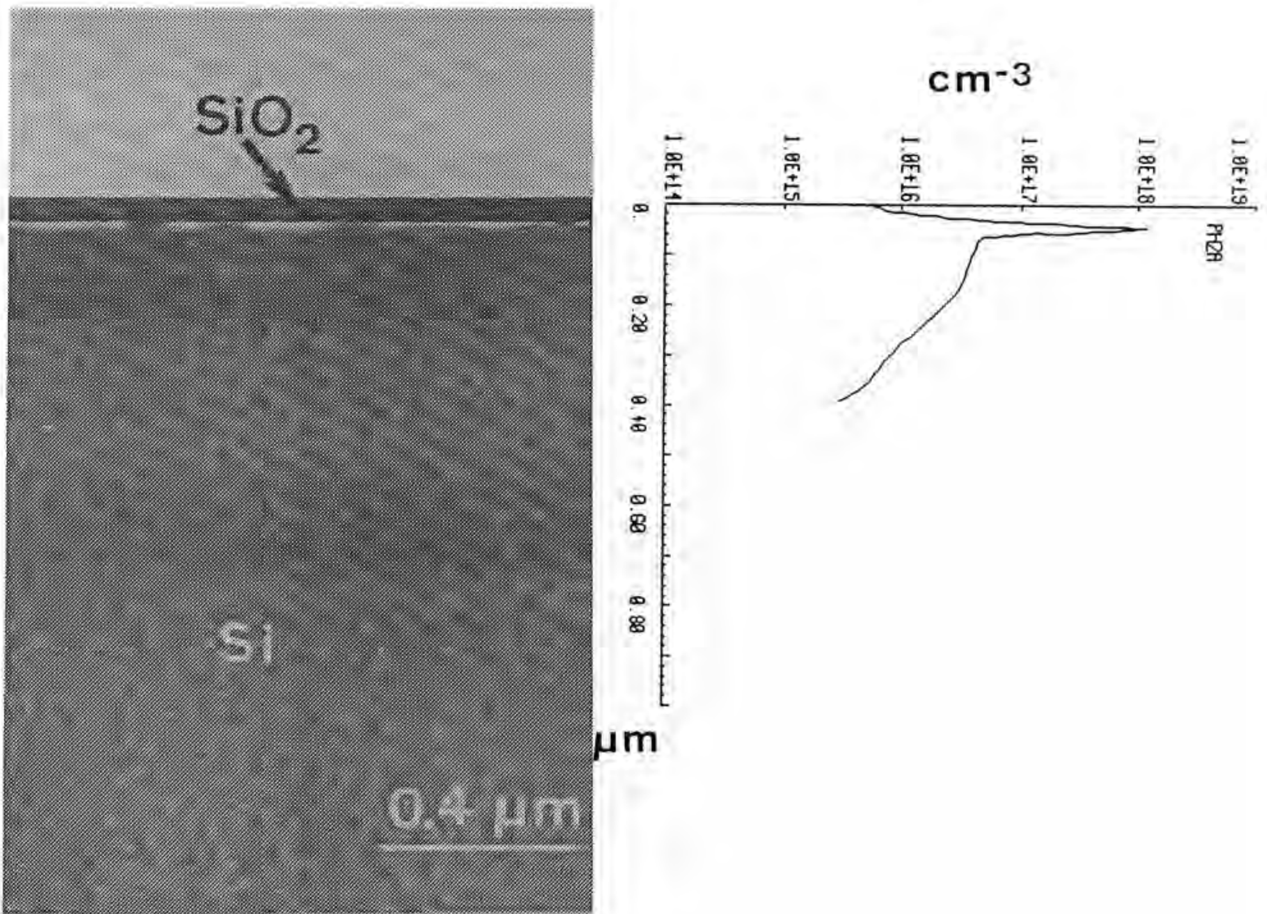


Figure 4.6: Cross-section specimen implanted with a low dose of phosphorus and etched with the solution $K_2Cr_2O_7$ 0.15M and HF 3M for 2 minutes. As the peak concentration is below $6 \cdot 10^{17}$ cm^{-3}, only a line can be seen, which might correspond to the junction.

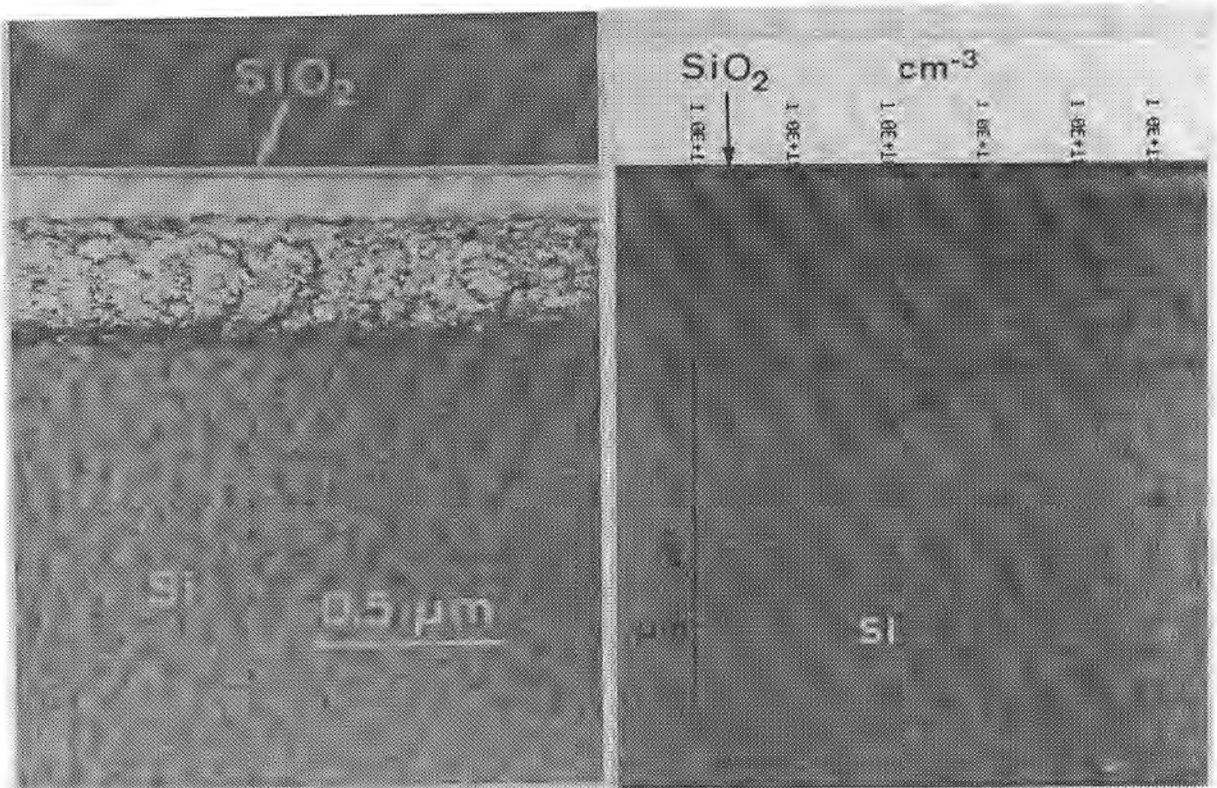


Figure 4.7: High energy boron implanted and annealed cross-section specimen, a) etched using the same solution as in figure 4.6. The buried profile has been delineated at a level of $3{\cdot}10^{17}$ cm^{-3}. b) Irradiated at 220°C for 22 minutes with an electron flux of $1.2{\cdot}10^{18}$ cm^{-2} s^{-1}. Preferential defect generation occurs at doping levels above $3{\cdot}10^{17}$ cm^{-3}.

file. After etching two lines are present, corresponding to a doping level of $3 \cdot 10^{17}$ cm^{-3}, obtained by comparison with SIMS.

Other experiments on boron doped wafers have been carried out, leading to the delineation of a high doping level, but no delineation of the junction has been observed.

4.2.3 Limitations of the chemical etching method

The use of chemical etching techniques combined with TEM has been presented, showing that delineation of specific doping levels can be achieved. However this method presents some drawbacks that limit the general applicability:

a) The reproducibility rate is still low.

b) Destruction of the specimen during the etching can occur, making the general procedure difficult.

c) The surface state of the specimen determines whether the etching will be successfully performed or not.

In the next section an alternative method, which does depend very little on the surface state of the specimen will be presented.

4.3 IN-SITU ELECTRON IRRADIATION IN THE HVEM

4.3.0 Introduction

It is well known that high energy electrons can generate point defects in crystalline materials. In the case of silicon and at room temperature the displacement threshold is about 11 to 22 eV [4.23]. According to this reference, to transfer this energy to the silicon atoms the incident electrons have to carry a kinetic energy above 140 keV. However, for the creation of extended defects energies above 280 keV and high electron fluxes, of the order of 10^{18} cm^{-2} s^{-1}, are needed. This means that no extended defect creation is expected to happen in a 200 keV electron microscope. With the actual intermediate voltage TEMs,

working at 300 or 400 kV, extended defect generation is observed during prolonged observation. Increasing the temperature leads to a decrease of the threshold energy. At 1000°C it is possible to create extended radiation defects even in a 100 kV electron microscope [4.24].

One of the first reports on extended defect creation in silicon crystals during electron irradiation in the HVEM, was published by Matthews and Ashby [4.25]. They observed elongated rod-like defects or loops lying in {113} planes. Later, Salisbury and Loretto [4.26] proved that these defects were of interstitial type. Since then extensive work has been performed on this subject, studying the influence of the substrate type and of the impurities present in it, and also of the irradiation conditions. The following general trends have been reported:

a) Irradiation at temperatures above 370°C leads to the creation of extended defects, typically rod-like defects, loops and stacking faults.

b) Irradiation at lower temperatures gives rise to small dot-like defects, of which the character is difficult to determine. It is supposed that they are of interstitial type and represent the initial stage of the extended defects created at higher temperatures.

4.3.1 Observations

During prolonged observation of a cross-section specimen in the HVEM, the creation of radiation induced defects can thus be expected. In this work the defect generation in cross-section specimens prepared from implanted and annealed wafers are studied. In our experiment the distribution of these defects has been noticed to be non uniform over the complete specimen. In certain areas clustering was enhanced, while in others no extended defects were visible. Comparing the defect distribution as a function of depth with the corresponding dopant profile, given by SIMS or SRP, it has been found that a one to one correlation between the dopant level and the defect generation exists [4.8].

In order to study this behaviour an extensive irradiation experiment performed. Using the cooling and heating holders of the HVEM, irradiations have been carried out on specimens at different temperatures and electron fluxes. The temperatures ranged between 77K and 643K. No higher values were used because already at this last temperature the diffusion length of the copper from the grid is about 600 µm for the irradiation times used in the experiment. Higher temperatures could lead to interaction between the copper and the generated point defects and might change the observed behaviour. The experiments performed at liquid nitrogen temperature gave no clear difference compared to room temperature irradiation. For this reason, only results will be given for the higher temperatures.

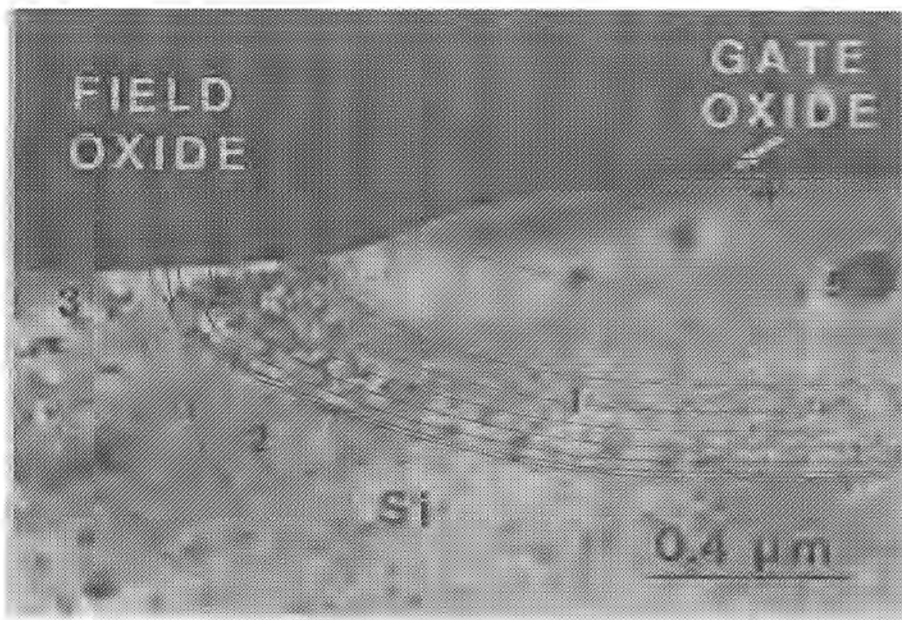
In most of the experiments discussed in the following the electron flux density has been kept constant at a value of $1.2\cdot10^{18}$ cm^{-2} s^{-1}. Irradiations have been performed in 0.25 to 0.5 µm thick specimen areas.

The investigations have framed mainly on boron doped material. Some samples doped with arsenic or phosphorus have also been studied. This allows to compare the defect generation both in acceptor and in donor doped substrates. The results of this comparative study give some first indications about the basic mechanisms involved in the lattice defect generation.

<u>4.3.1.1 Boron doped silicon</u>

Figure 4.8a shows the result of an electron irradiation experiment on a boron doped sample. The ion implantation (60keV, $5\cdot10^{14}$ cm^{-2}) has been performed in the gate area through an oxide mask. After implantation an anneal was performed at 1000°C for 30 minutes. The specimen has been irradiated at room temperature for about 30 minutes with an electron flux of $5\cdot10^{18}$ cm^{-2} s^{-1}. The electron flux is determined from a measurement using a Faraday cup located below the imaging system. In figure 4.8a different areas can be observed in the irradiated specimen:

a) An area free of irradiation induced defects (IID's), close to the gate oxide. Comparison with the SIMS profile (figure



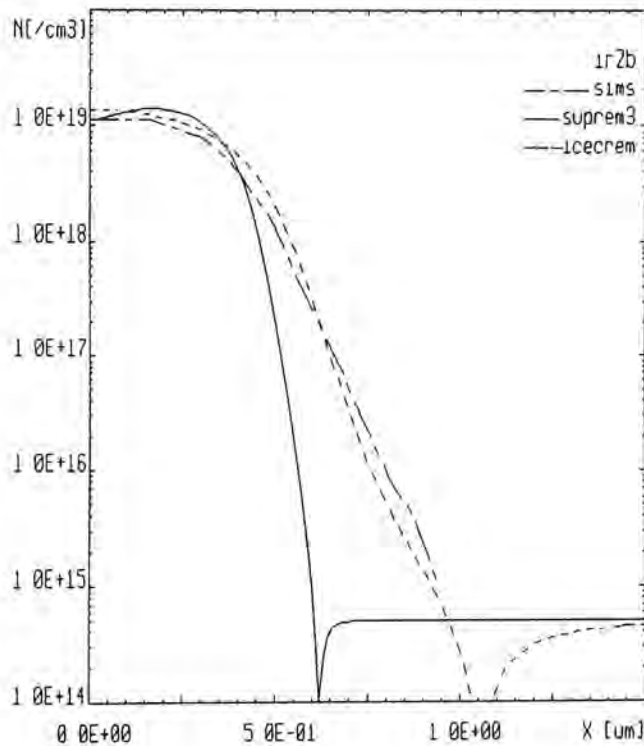


Figure 4.8: a) Irradiation of a boron implanted and annealed specimen at room temperature. The electron flux is $5 \cdot 10^{18}$ cm^{-2} s^{-1}. The simulation obtained using SUPRA is superimposed. It can be seen that the shape of the defects follows the isoconcentration contours. However, the lowest contour corresponds to 10^{14} cm^{-3}, according to SUPRA, while SIMS (b) gives $5 \cdot 10^{17}$ cm^{-3}. The area with the high density of defects corresponds to levels between $5 \cdot 10^{17}$ and $5 \cdot 10^{18}$ cm^{-3}. b) SIMS profile and simulations using SUPREM 2 and PREDICT, showing the discrepancy between them.

4.8b) shows that this layer corresponds to doping levels above $5 \cdot 10^{18}$ cm^{-3}. End of range defects due to the implantation which have not been annealed out can also be observed (5). Their presence seems not to influence the IID generation mechanism. Close to the gate oxide some small defects are also present.

b) A zone with a very high density of IID's (1); comparing the depth at which these defects are created with the SIMS profile of the same area (of course the SIMS profile has been obtained in the area where no structures were present) it can be determined that the doping level lies between $5 \cdot 10^{17}$ and $5 \cdot 10^{18}$ cm^{-3}.

c) In area (2) a uniform distribution of IID's is observed. The doping level at which this happens is below $5 \cdot 10^{17}$ cm^{-3} in the p doped region and also the bulk, which is doped with phosphorus.

d) Below the field oxide a small concentration of defects is visible (3), separated from the oxide by a defect free zone.

On the TEM image the simulated profile obtained with the program SUPRA (which is an extension to two dimensions of SUPREM 2) has been superimposed. It can be observed that the shape of the isoconcentration contours follows well the radiation induced defect profile. The concentration levels predicted by this simulator are, however, not correct. The lowest contour corresponds to 10^{14} cm^{-3}, while the level given by SIMS is $5 \cdot 10^{17}$ cm^{-3}. Also the predicted thicknesses of the gate and field oxides differ from the real ones. The predicted gate oxide is thicker than the measured one, while the predicted field oxide is thinner. This shows that the results of processing simulators have to be taken with some precautions.

On the right figure the predictions using one dimensional simulation programs SUPREM 3 and ICECREM are compared with SIMS, showing for the SUPREM 3 a big discrepancy.

Figure 4.9 shows the evolution and growth of the radiation induced extended defect generation at a temperature of 370°C as a function of the irradiation time. This is a specimen coming from the same wafer as the one in figure 4.8. However this area corresponds to the implantation through a polysilicon mask. A region with a high density of defects is observed between the same dopant concentration levels as in figure 4.8a. Some differences can however be observed. They are, mainly:

a) A zone denuded of defects is obtained at dopant concentration levels between the junction, deduced from the extrapolation of the SIMS profile and $5 \cdot 10^{17}$ cm^{-3}. This denuded zone was not visible in figure 4.8a.

b) The extended defects generated at this temperature appear already after 1 minute of irradiation and are larger than at room temperature.

c) There is a high density of defects just below the polysilicon mask to a depth of 10 nm. For irradiation at room temperature and 220°C, this area was defect poor. At 370°C the lateral extent of the defect poor zone close to the mask edge is larger than at the lower temperatures.

d) Below the gate oxide, small defects appear after 15 minutes irradiation.

An interesting observation is that after irradiation at high temperature, irradiating again at room temperature leads to shrinking and even annihilation of defects. Similar reverse annealing results have been reported in the literature [4.27]. This effect is illustrated in the section on the irradiation of arsenic doped material.

Initially it was thought that the denuded zone observed close to the surface at the highly doped areas could be strongly influenced by the gate oxide. The silicon/silicon oxide interface could act as a strong sink for the created point defects. To test this hypothesis irradiation of a specimen from the same wafer as the one of figure 4.7a was performed. The result after

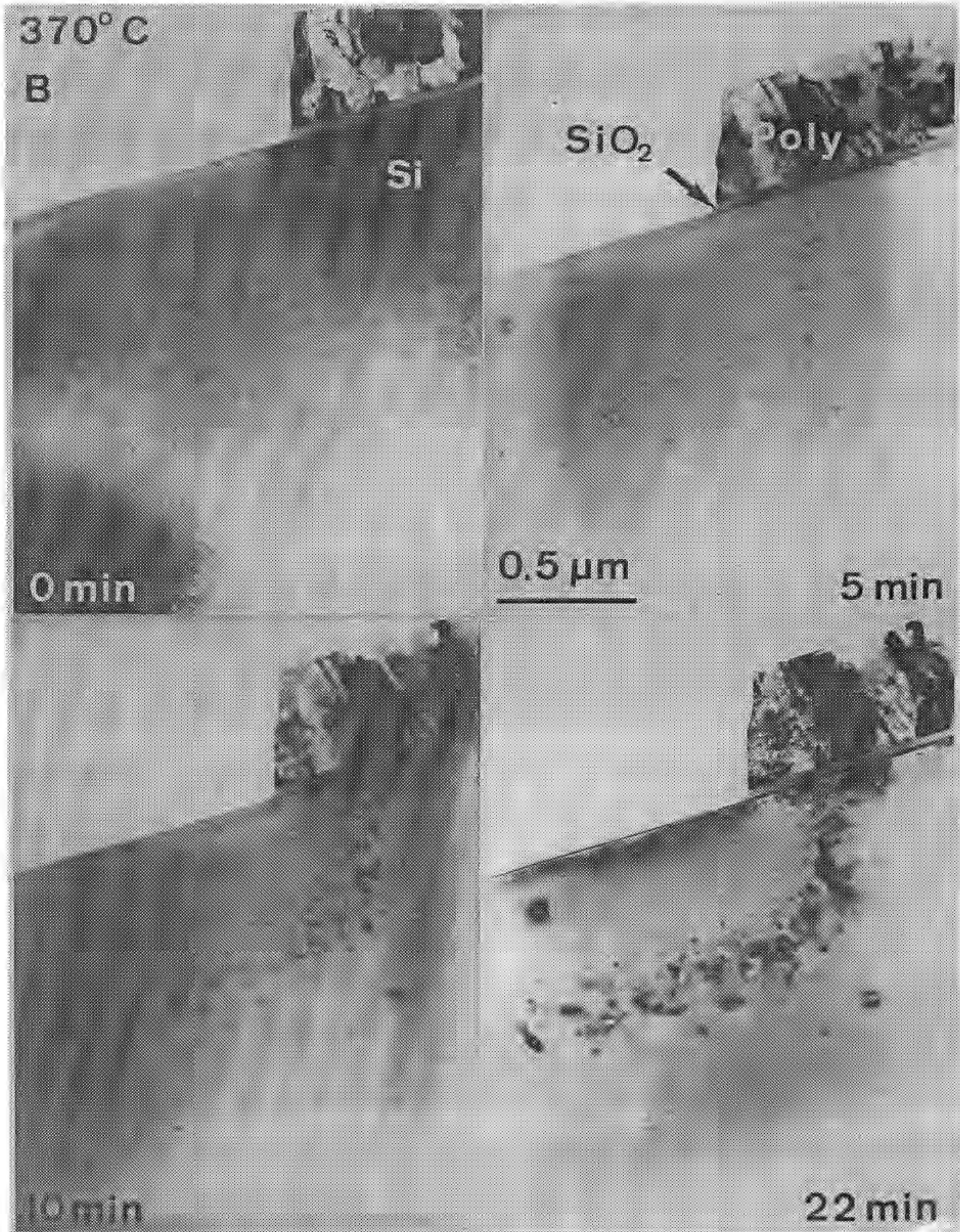


Figure 4.9: Irradiation at a temperature of 370°C of a similar specimen to the one in figure 4.8.

irradiation at 220°C for 22 minutes can be observed in figure 4.7b. Comparing the depth at which the defects are created with the superimposed SIMS profile, shows good agreement with the previously obtained results. Defects are indeed created in the areas with dopant concentration levels above $5 \cdot 10^{17}$ cm^{-3}. A denuded zone close to the surface is visible, which corresponds to doping levels below that value. This still does not give a real measure of the denuded zone, but an upper limit can be estimated to be 5 nm. As the mobility of the point defects decreases with decreasing temperature, this upper limit at room temperature should be even lower. From this argument it is deduced that the denuded zone observed for highly doped regions is not due to the presence of the gate oxide.

By irradiating a wafer with layers of uniform doping, grown by low temperature epitaxy and each with different levels, an additional interesting observation has been obtained. The substrate concentration was about 10^{19} cm^{-3} and an epitaxial silicon layer was grown. The growth has been achieved in a low temperature epitaxy system and the layer was boron doped, with intentional strong variations of the profile. Figure 4.10 shows the result of the irradiation of this sample at 220°C and with the measured SIMS profile superimposed. Effectively, preferential defect generation takes place in the areas with higher doping level, above about $2 \cdot 10^{17}$ cm^{-3}, while defect poor zones are obtained in the lower doped areas. The small discrepancy between this doping level and the one obtained for the first boron sample is within the experimental error both of the magnification of the microscope and of the depth accuracy of SIMS.

Figure 4.4b shows a similar specimen as the one of figure 4.4a, irradiated at 220°C for 10 minutes. The behaviour described in this section is visible, i.e., two denuded zones, one above $5 \cdot 10^{18}$ cm^{-3} and the other below $5 \cdot 10^{17}$ cm^{-3} are visible. The delineated level is the same as the achieved using chemical etching.

As the final objective of the described irradiation technique is its application to visualise dopant concentration pro-

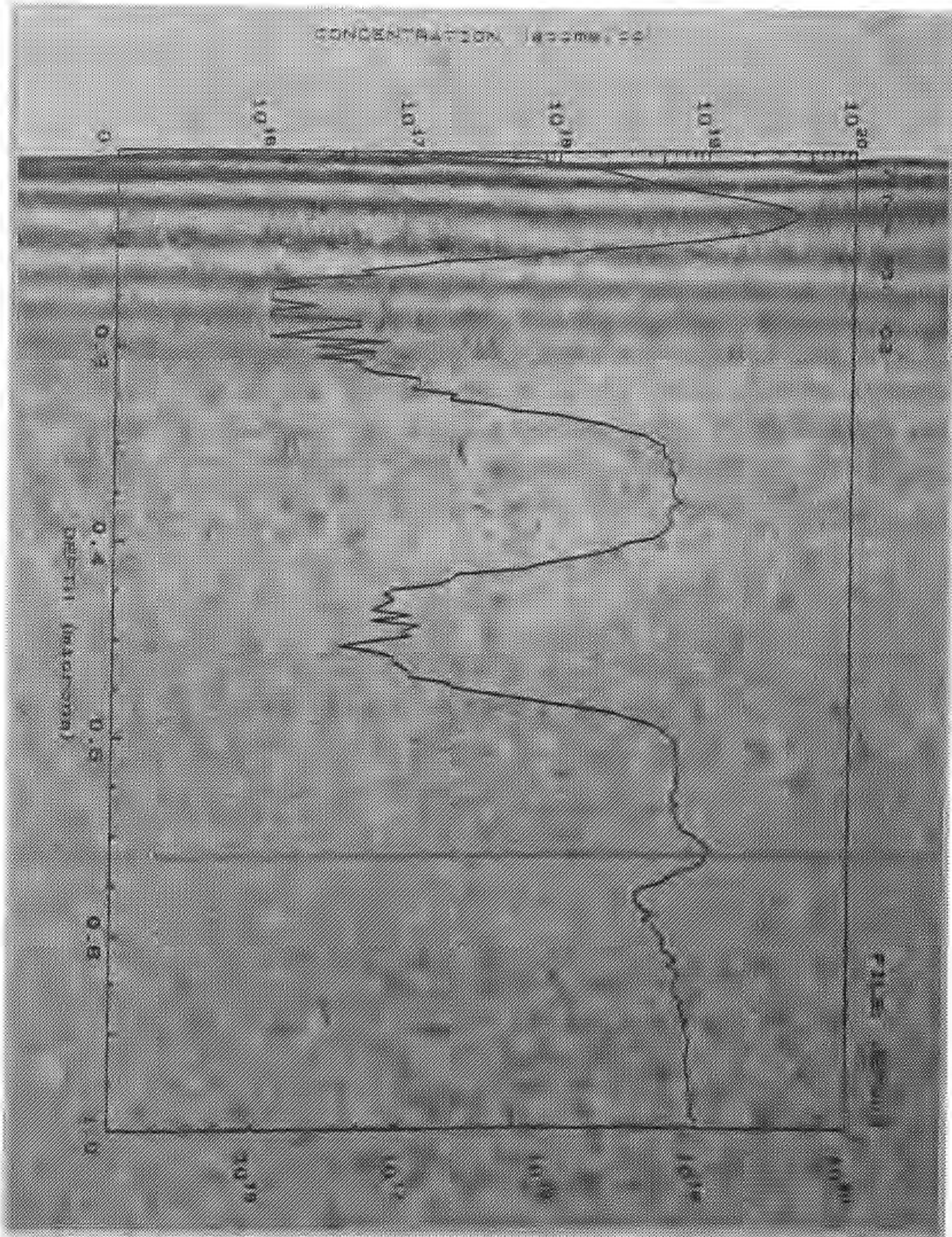


Figure 4.10: Effect of the irradiation (370°C) of a wafer on which a boron doped layer had been grown in an epitaxial reactor. Superimposed the SIMS profile is presented. The generation of defects occur in areas with doping level above $5 \cdot 10^{17}$ cm^{-3}.

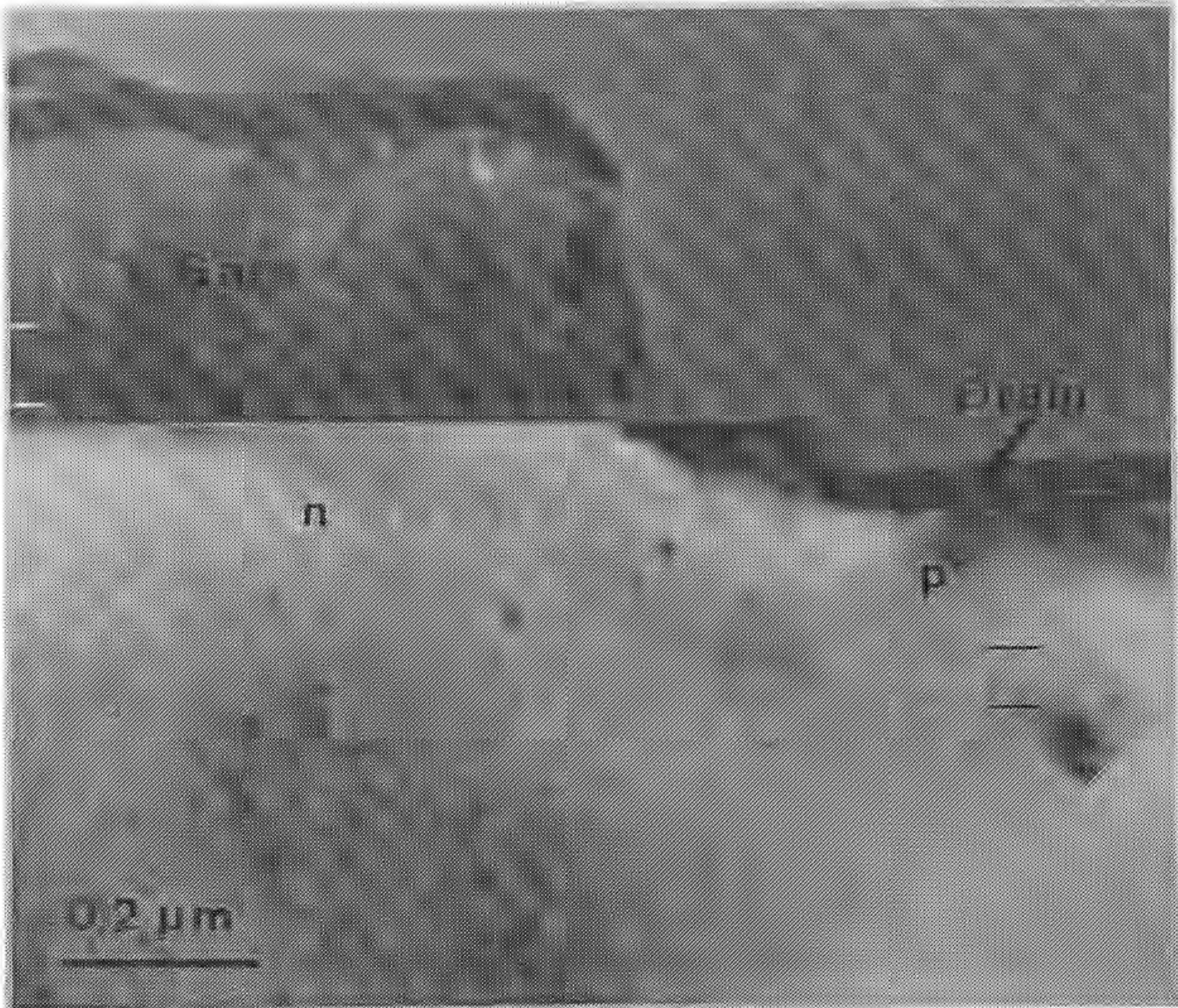


Figure 4.11: PMOS transistor irradiated at room temperature. The area with high density of defects, indicated by two black lines, corresponds to levels between $5 \cdot 10^{17}$ and $5 \cdot 10^{18}$ cm^{-3}. In higher doped regions of the p^{+} area no defects are visible.

files in integrated circuits, figure 4.11 shows the gate area of a 4 μm PMOS transistor, which has been irradiated for 30 minutes at room temperature. This specimen has been prepared using the localized techniques described in section 2.3. It can be seen that in the region close to the source contact no radiation damage occurs. Below this region an area with high density of defects is visible. The bending of the defect area below the spacer oxide can also be observed. Correlating the depth of the areas with the high defect density with SRP measurements shows that the delineated level corresponds to the previously determined ones.

4.3.1.2 Phosphorus

Similar irradiation experiments have been performed using phosphorus ion implanted wafers. Figure 4.12 shows the result of the irradiation at 370°C of a phosphorus implanted wafer. The implanted dose was $2 \cdot 10^{14}$ cm^{-2}. The resulting SIMS profile is presented in the same figure. It is important to point out that the peak concentration in this case is below 10^{19} cm^{-3}, as will be discussed in section 4.3.2. In this specimen three areas are distinguished:

a) As in the case of boron, a low density of large defects are created in the bulk. When comparing with the extrapolation of the SIMS profile, it is clear that the defect area extends to phosphorus doping levels up to 10^{15} cm^{-3}. This doping level is very close to the metallurgical junction.

b) A large denuded zone is present between the just mentioned defects and the end of range defects, i.e., until about $2 \cdot 10^{18}$ cm^{-3}.

c) Above this level also extended defects are created, which are visible by comparing the first and the third pictures.

After reirradiation of this specimen at room temperature, irradiation induced defect annihilation is again observed. In addition, as was also observed for boron, the end of range defects shrank during this second irradiation.

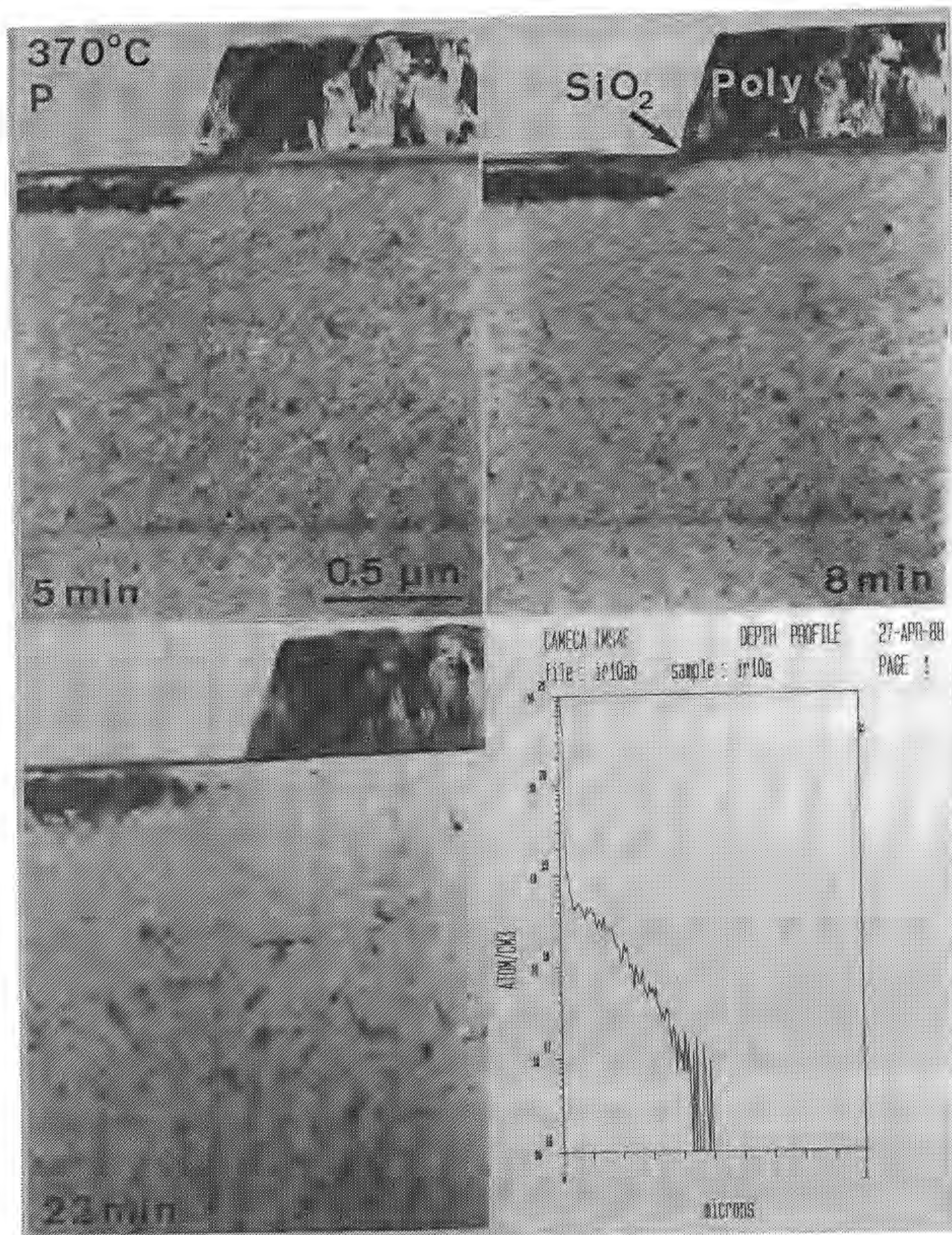


Figure 4.12: Irradiation at 370°C of a phosphorus implanted wafer. The lower right figure shows the SIMS profile. A high density of IID's is generated above dopant concentrations of $2 \cdot 10^{18}$ cm^{-3}.

Another phosphorus doped specimen, with a peak concentration of about 10^{20} cm^{-3}, has been irradiated at 370°C. The obtained results are similar to those presented in figure 4.12. For this sample a denuded zone was observed close to the surface, corresponding to doping levels above $2 \cdot 10^{19}$ cm^{-3}.

4.3.1.3 Arsenic

In figure 4.13 the result of irradiation at 370°C of an As doped specimen is presented. The implantation energy and the dose were, respectively, 100 keV and $2 \cdot 10^{16}$ cm^{-2}. The wafer was annealed at 1000°C for 30 minutes. These anneal conditions are such that all end of range defects have been eliminated. Important in this specimen is that the peak concentration is of the order of 10^{21} cm^{-3}, which is very close to the maximum solid solubility of arsenic in silicon [4.28] and which might give rise to precipitation.

The main characteristics of the irradiation experiments on arsenic doped wafers above room temperature are:

a) In the highest doped areas, above $2 \cdot 10^{20}$ cm^{-3}, creation of small defects can be observed.

b) A denuded zone is present between $3 \cdot 10^{19}$ and $2 \cdot 10^{20}$ cm^{-3}. A strongly marked line is seen in the middle of this denuded zone, situated at a depth of about 460 nm, which corresponds to 10^{20} cm^{-3}.

c) Below this level an area with a high density of large defects is visible. Its lower limit is beyond the sensitivity of SIMS. The size and the density of these defects below the polysilicon mask is smaller.

d) A uniform density of defects can be observed in the bulk, similar to the other samples.

Reirradiation has been performed in the same conditions as for the other two dopants, observing again the shrinkage of the previously introduced defects. This is illustrated in figure

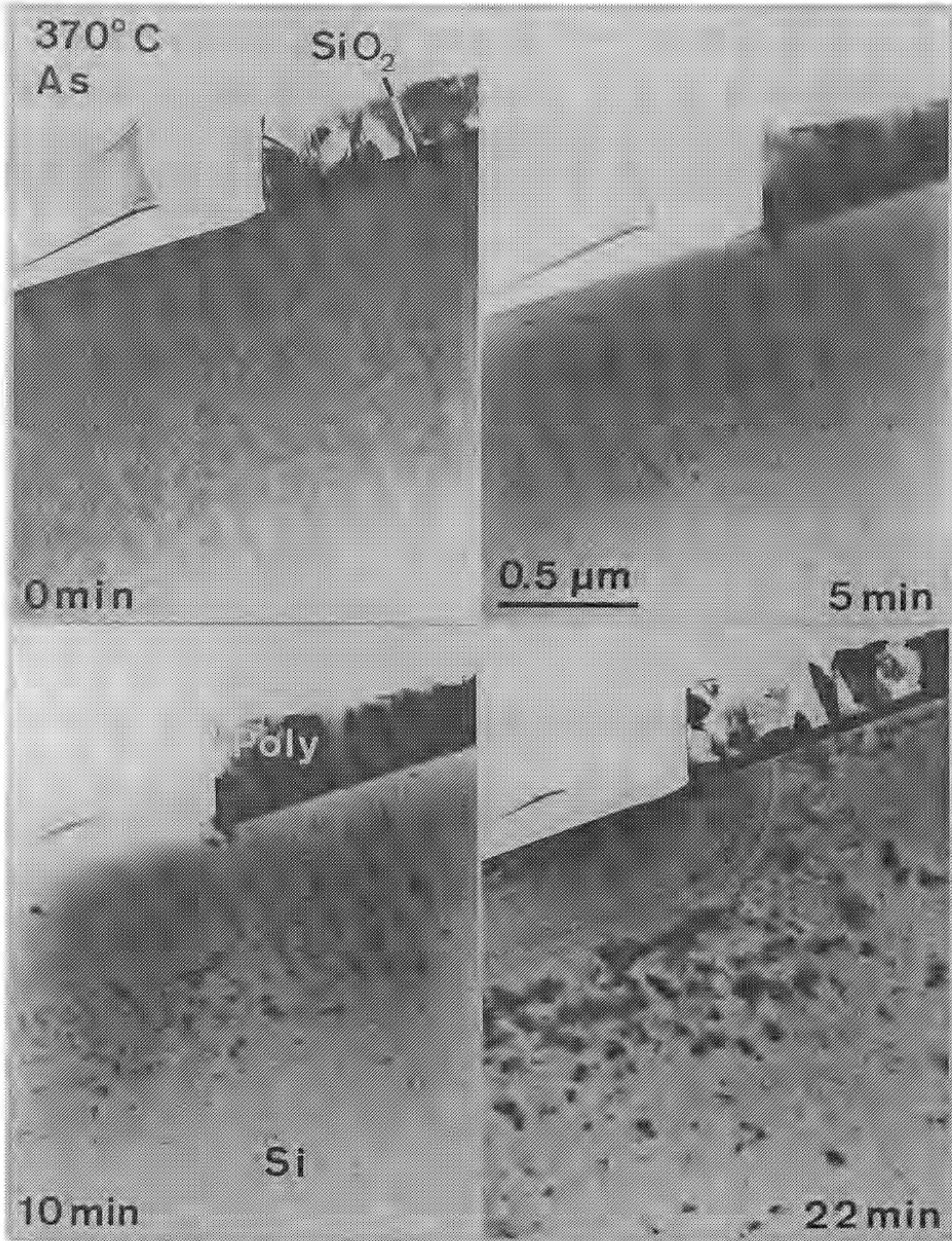


Figure 4.13: Generation and growth of the radiation induced defects of an arsenic ion implanted specimen, irradiated at 370°C. A strong line can be observed, corresponding to a level of 10^{20} cm^{-3}.

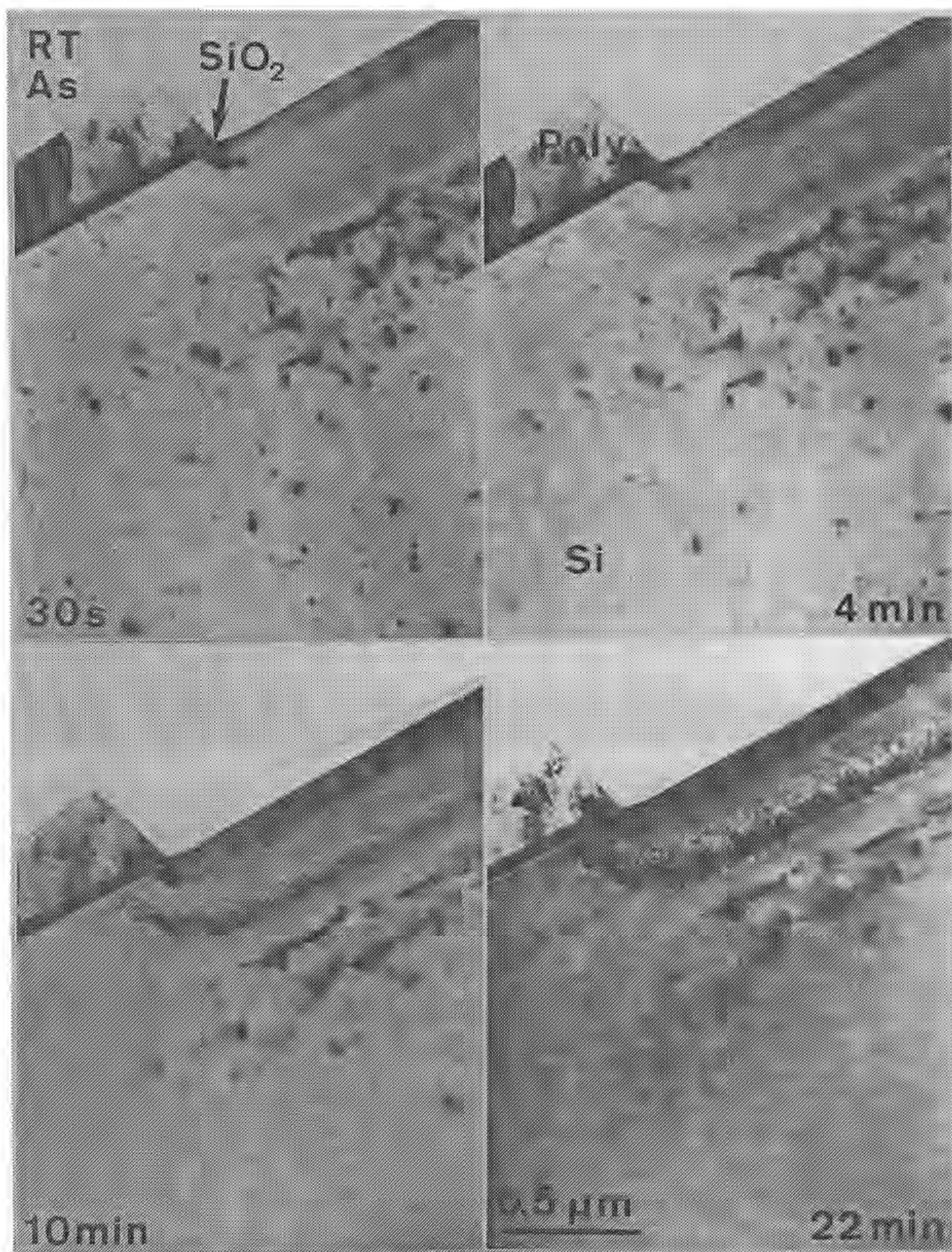


Figure 4.14: Effect of the irradiation at room temperature of the previously irradiated specimen of figure 4.14. The defects shrink and even disappear during this procedure. A region with a high density of small dot-like defects appears where previously a denuded zone was.

4.14, corresponding to the same specimen of figure 4.13. A high density of small defects is now observed, corresponding to doping levels between 2 and $5 \cdot 10^{20}$ cm^{-3}. Furthermore the marked line disappears during this reirradiation.

The specimens have been further thinned in order to study the created defects using HREM. Figure 4.15 shows one of these defects, which has been found in the bulk of the wafer, i.e., far away from the implanted area. It can be observed that it lies in a {113} plane. This indicates that the defects are formed by clustering of interstitials, according to the model by Salisbury and Loretto [4.26].

4.3.2 Discussion

Some general features have been observed for all irradiation experiments:

a) Radiation induced defects can only be created in the relatively thick areas (thicker than about 100 nm).

b) There exists always a certain "incubation" time before the irradiation induced extended defect formation takes place. This incubation time decreases with increasing temperature.

Brown and Fathy [4.29] also irradiated p-n junctions in the HVEM. They observed a different defect generation for n- and p-type silicon. For their irradiation conditions (680 keV, $1.2 \cdot 10^{19}$ cm^{-2} s^{-1} and temperatures between 300 and 600K), defect generation was reported to be suppressed in p-doped areas ($6 \cdot 10^{16}$ cm^{-3}), while no generation in highly doped areas ($5 \cdot 10^{19}$ cm^{-3}) of n-doped material was observed. In the p-n junction depletion area the electric field ionization was explained by assuming that charging of the created intrinsic point defects suppresses clustering, while uncharged defects can cluster more easily. The assumption of the presence of an interstitial level, E_i, near the middle of the band gap, about 0.4 eV above the top of the valence band, implies that the number of charged (n_i^+) and uncharged (n_i^o) interstitials depend on the position of the Fermi level according to:

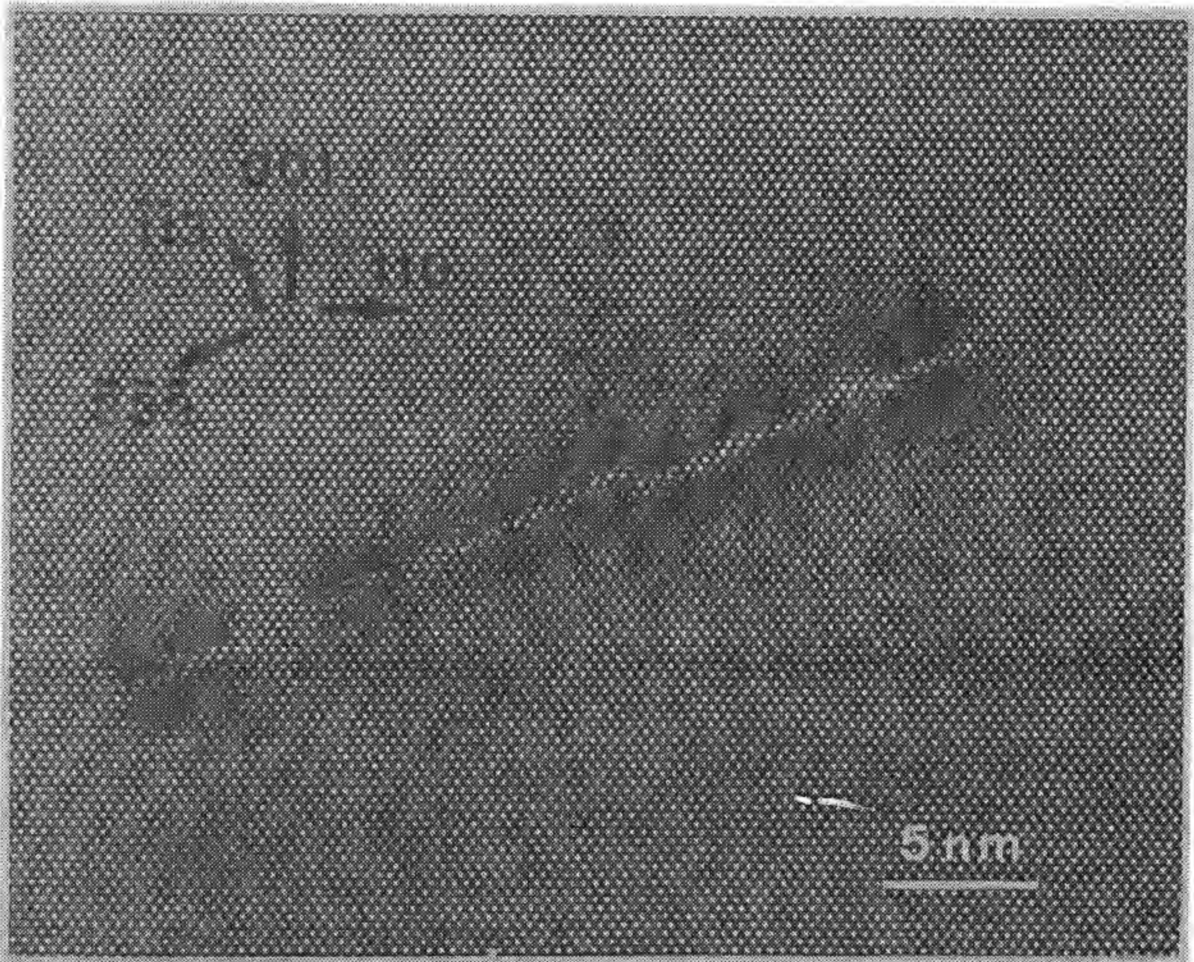


Figure 4.15: HREM image of one of the radiation induced defects generated in the bulk of the previous specimen. The defect lies in a {113} plane.

$$n_i^o/n_i^+ \approx \exp -\frac{Ei-Ef}{kT} \qquad (4.2)$$

For highly doped p-type silicon the Fermi level is below the self-interstitial level. The generated self-interstitials will then be positively charged, so that Coulomb repulsion exists, making their clustering into extended defects more difficult. When the Fermi level is above the self-interstitial level, the interstitials are neutral and can cluster more easily.

An alternative explanation for the lack of extended defects in the highly doped areas is the creation of B_sB_i pairs formed by the reaction [4.30]:

$$B_s + Si_i \rightarrow B_i$$
$$B_s + B_i \rightarrow (B_sB_i) \qquad (4.3)$$

where the subscripts s and i indicate, respectively, substitutional or interstitial position. The probability of formation of (B_sB_i) pairs is proportional to $\Phi\ [B_s]^2$ (Φ is the total electron flux on the specimen). These (B_sB_i) pairs are stable up to relatively high temperatures, up to 400°C. At very high B_s concentrations, i.e. above 10^{19} cm^{-3}, B_iB_i pairs are also formed. The formation of large numbers of B_sB_i pairs represents a strong sink for the generated self-interstitials, explaining why in the highly doped areas no extended defects are formed for not too long irradiation times and too large fluxes.

The density of radiation damage will also be affected by the presence of electric fields in the junction region and also by the elastic stresses at the silicon-silicon oxide interface. The presence of large electric fields will enhance the diffusion of charged point defects. The influence of stresses of an oxide layer on the defect density can be observed in figure 4.8 below the field and the gate oxides.

As mentioned in reference [4.13], the strain field introduced by the dopant atoms gives rise to a contrast change in the electron microscope when using the correct diffraction condi-

tions. We believe that the strain field could also influence the generation mechanism of Frenkel pairs or the clustering.

If only the elastic effects are taken into account, the planar strain introduced by the dopant at a concentration C into the silicon lattice is given by [4.31]:

$$\varepsilon = \beta C \tag{4.4}$$

where β is the lattice-contraction coefficient. β can be calculated using a volumetric model [4.32]:

$$\beta = \frac{1}{3 N}\left[1 - \left(\frac{r_{dop}}{r_{Si}}\right)^3\right] \tag{4.5}$$

where N is the density of lattice sites in silicon ($5 \cdot 10^{22}$ cm^{-3}), and r_{dop} and r_{Si} are the covalent radii of the dopant and of the silicon atom, respectively.

Figure 4.16 shows the absolute strain introduced in the silicon lattice calculated using these formulae. As pointed out in section 4.3.1.1 the region of the boron doped specimens where an enhanced clustering of defects takes place is between $5 \cdot 10^{17}$ and $5 \cdot 10^{18}$ cm^{-3}. The corresponding strain can be deduced from the figure and lies between $2 \cdot 10^{-6}$ and $2 \cdot 10^{-5}$. The doping values corresponding to this strain levels for phosphorus doping are between $1.8 \cdot 10^{18}$ and $1.8 \cdot 10^{19}$ cm^{-3}. This values correspond perfectly to the area with high density of defects in the phosphorus doped wafers.

Arsenic, because it has a larger covalent radius, gives a tensile stress in the silicon lattice. The doping level corresponding to the previously indicated strain values are $1.2 \cdot 10^{19}$ and $1.2 \cdot 10^{20}$ cm^{-3}. These doping levels are the lower and upper limits of the denuded zone shown in figure 4.13.

4.4 CONCLUSIONS

In this chapter two TEM techniques for dopant profiling of silicon doped material have been presented. One of them is based

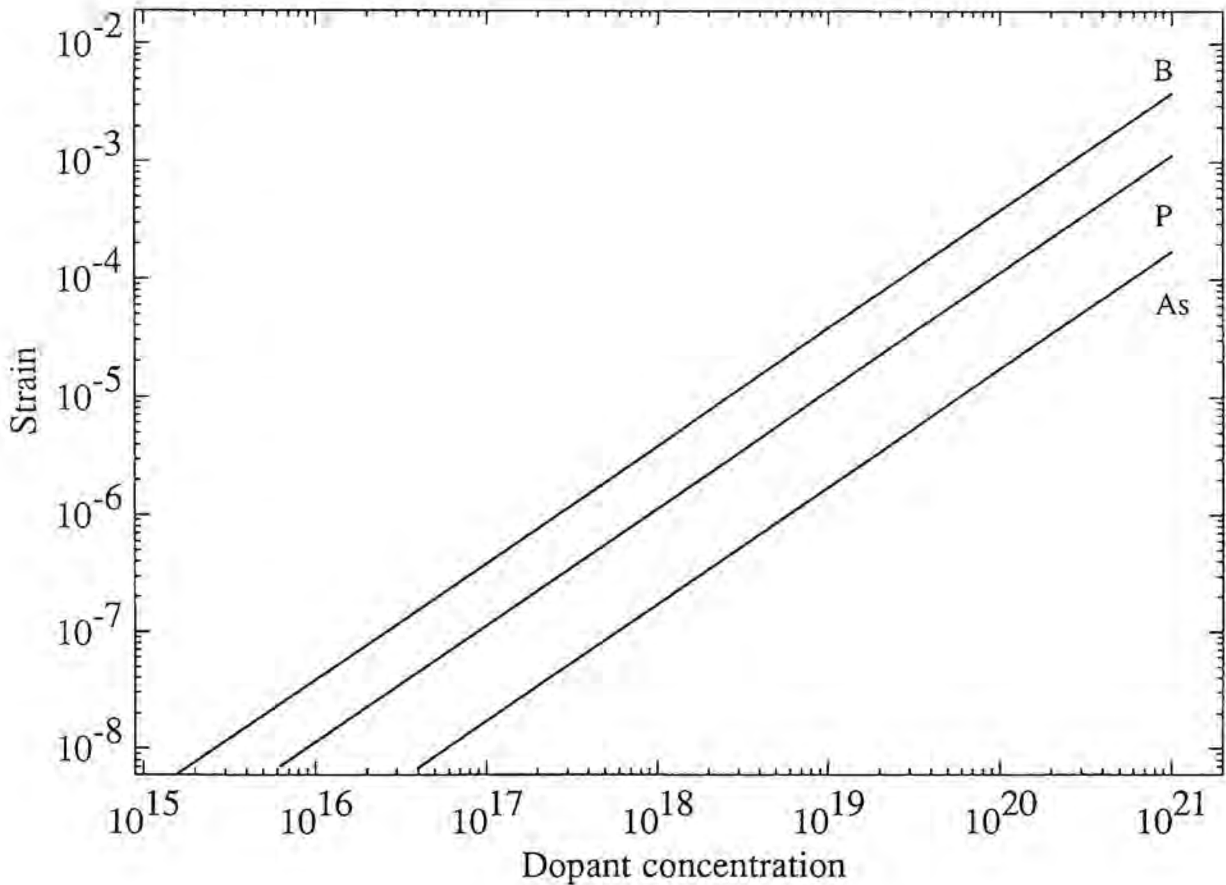


Figure 4.16: Calculation of the elastic strain induced by dopant atoms in substitutional position as a function of the dopant concentration.. The regions where high density of irradiation induced defects were generated in the case of boron and phosphorus are indicated and correspond to the same values of strain.

in the preferential etching of cross-section specimens using solutions with a dopant dependent etching rate. Two solutions have been proposed:

- A solution HF:HNO3=1:300 at 5°C has been proved to delineate doping levels down to $6\cdot10^{17}$ cm^{-3}, both for boron and phosphorus doped wafers.
- A solution $K_2Cr_2O_7$ 0.15M and HF 3M has been found to delineate two different levels in phosphorus doped material: a first level at $3\cdot10^{18}$ cm^{-3}, and a second level, deeper in the bulk and that might correspond to the metallurgical junction. For boron doped silicon only one level has been delineated, corresponding to $3\cdot10^{17}$ cm^{-3}. For this method the cross-section specimens have to be prepared along [010] directions instead of the more common [110].

The use of the chemical solutions has the drawbacks of the poor reproducibility rate, the strong variation with the acid lot and the specimen aging and the danger of destroying the specimen during etching.

A second delineation method has been developed, consisting in in-situ extended defect generation during electron irradiation in the HVEM. These defects are generated in well defined areas of the specimen, which have been correlated with fixed doping levels in a reproducible way. The influence of the temperature and of the electron flux on the defect behaviour has been investigated. Boron, phosphorus and arsenic implanted silicon wafers have been studied, showing the following behaviour during irradiation:

a) For boron an area with a high density of defects is observed at doping levels between $5\cdot10^{17}$ and $5\cdot10^{18}$ cm^{-3}. For higher doped areas small defects appear only after irradiation at high temperature (370°C). At lower boron levels a defect poor zone is also observed, which might correspond to the metallurgical junction. In the bulk (n-doped) a uniform density of {113} defects is observed.

b) Phosphorus doped wafers show a similar behaviour. The dopant level at which the high density of defects are generated is however higher and lies between $2\cdot10^{18}$ and $2\cdot10^{19}$ cm^{-3}.

c) The irradiation of arsenic doped cross-section specimens reveals a different defect behaviour. A defect poor zone is observed at levels between $2\cdot10^{19}$ and $2\cdot10^{20}$ cm^{-3}. Below this level an area with a high density of defects is obtained, which extends into the bulk, probably until the junction. At temperatures above 220°C in this defect poor area a marked line is visible, corresponding to 10^{20} cm^{-3}.

Possible explanations for the observed behaviour are:

- The position of the Fermi level can affect the clustering mechanism. For an interstitial level in the lower half of the gap, the generated interstitials will be charged or uncharged depending on the position of the Fermi level. Charged interstitial will repel each other, so that clustering in extended defects is suppressed.

- for boron doped material, in the highly doped areas boron interstitial-boron substitutional pairs can be formed, acting as a dominant sink for the generated self-interstitials so tat clustering will be retarded.

- The elastic strain introduced by the dopant can enhance the defect growth. The highly defective areas in the boron and phosphorus doped wafers correspond to the same strain values.

Further work is required to understand the basic mechanisms that give rise to this dopant dependent radiation induced defect generation.

REFERENCES CHAPTER 4

[4.1] R. Subrahmanyan, H.Z. Massoud and R.B. Fair, Appl. Phys. Lett. 52 (25), 2145 (1988).

[4.2] G.J.L. Ouwerling and M. Kleefstra, Proc. ESSDERC 1990, 193 (1990)

[4.3] C. Hill, P.J. Pearson, B. Lewis, A.J. Holden and R.W. Allen, Proc. ESSDERC 1987, 926 (1987).

[4.4] C. Curling, R. Hokke and A.H. Reader, Inst. Phys. Conf. Ser. 100, 531 (1989).

[4.5] T.T. Sheng and R.B. Marcus, J. Electrochem. Soc. 128, 881 (1987).

[4.6] M.C. Roberts, K.J. Yallup and G.R. Booker, Inst. Phys. Conf. Ser. 76, 483 (1985).

[4.7] D.P. Gold, J.H. Wills, G.R. Booker, M.C. Wilson and D.J. Godfrey, Inst. Phys. Conf. Ser. 100, 537 (1989).

[4.8] A. Romano, J. Vanhellemont, J.R. Morante and W. Vandervorst, Micron and Microscopica Acta 20 (2), 151 (1989)

[4.9] P. Renteln, D.G. Ast, T.C. Mele and J.P. Krusius, J. Electrochem. Soc. 136, 3828 (1989).

[4.10] S. Kordic, E.J. van Loenen, D. Dijkkamp, A.J. Hoeven and H.K. Moraal, J. Vac. Sci. Technol. A 8, 549 (1990).

[4.11] W. Vandervorst, T. Clarysse, J. Vanhellemont and A. Romano-Rodríguez, accepted for presentation at the "International Workshop on the Measurement and Characterization of Ultra-Shallow Doping Profiles in Semiconductors", Triangle Research Park, USA, March 1991; to be published in J. Vac. Sci. Technol.

[4.12] M.G. Dowsett, G. Cooke, C. Hill, E.A. Clark, P. Pearson, I. Snowden and B. Lewis, Proc. of SIMS VII, (1989).

[**4.13**] D.D. Perovic, G.C. Weatherly, R.F. Egerton, D.C. Houghton and T.E. Jackman, to be published in Phil. Mag. A (1991).

[**4.14**] B. Schwartz and H. Robbins, J. Electrochem. Soc.123, 1903 (1976).

[**4.15**] E. Sirtl and A. Adler, Z. Metallkde. 52, 529 (1961).

[**4.16**] J. Secco d'Aragona, J. Electrochem. Soc. 119, 948 (1972).

[**4.17**] D.G. Schimmel, J. Electrochem. Soc. 126, 479 (1979).

[**4.18**] K.H. Yang, Semiconductor Processing, ASTM STP 850, American Society for Testing and Materials, 309 (1984).

[**4.19**] T.L. Chu and J.R. Gavaler, Electrochimica Acta 10, 1141 (1965).

[**4.20**] W. Maszara, C. Carter and G.A. Rozgonyi, in VLSI Science and Technology 1984, The Electrochemical Society, Pennington, 36 (1984).

[**4.21**] J.E.A.M. van den Meerakker and J.H.C. van Vegchel, J. Electrochem. Soc. 136, 1949 (1989); ibid. 136, 1954 (1989).

[**4.22**] A. Romano, J. Vanhellemont, A. De Keersgieter, W. Vandervorst, J.R. Morante and J. Van Landuyt, Proc. XIIth International Congress for Electron Microscopy, Vol. 4, 748 (1990).

[**4.23**] L.W. Hobbs, in Quantitative Electron Microscopy, eds. J.N. Chapman and A.J. Craven, Scottish Univ. Summer School of Physics, Edinburgh, 1984.

[**4.24**] L.I. Fedina, A.L. Aseev, S.G. Denisenko and L.S. Smirnov, in Defects in Semiconductors, ed. H.J. von Bardeleben, Trans, Tech. Publ. Ltd., Switzerland, 1123 (1986).

[**4.25**] M.D. Matthews and S.J. Ashby, Phil. Mag. A 27, 1313 (1973).

[**4.26**] I.G. Salisbury and M.H. Loretto, Phil. Mag. A 39 (3), 317 (1979).

[**4.27**] H. Asahi, R. Oshima and F.E. Fujita, Inst. Phys. Conf. Ser. 59, 455 (1980).

[**4.28**] F. Shimura and H.R. Huff, in VLSI Handbook, ed. N.G. Einspruch, Academic Press, New York, 191 (1985).

[**4.29**] L.M. Brown and D. Fathy, Phil. Mag. B 43 (4), 715 (1981).

[**4.30**] L.C. Kimerling, M.T. Asom, J.L. Benton, P.J. Drevinsky and C.E. Caefer, Mater. Sci. Forum Vols. 38-41, 141 (1989).

[**4.31**] B.G. Cohen, Solid State Electron. 10, 33 (1967).

[**4.32**] J.E. Lawrence, J. Electrochem. Soc. 113, 819 (1966).

UNIVERSITAT DE BARCELONA

Biblioteca de Física i Química

5. GENERAL CONCLUSIONS AND FURTHER RESEARCH

In this work some typical applications of transmission electron microscopy for the analysis and development of silicon materials and devices for ULSI technology have been presented.

For the preparation of thin silicon foils transparent for 200 kV or 1000 kV, special techniques had to be developed. In chapter 2 these specimen preparation techniques were discussed in detail. The described procedures to prepare high quality cross-section and plan view specimens can be performed in a short period of time, 4 to 6 hours, and with a high quality and success rate. The advantage of these techniques is that only conventional preparation tools, available in most of the laboratories working on TEM of solid state materials, are required.

In chapter 3 some of the approaches to obtain isolation of devices in integrated circuits have been studied. Conventional LOCOS can only be used for super micron technology. For ULSI circuits, where submicron technologies have to be used, alternative isolation methods are needed. To avoid the introduction of new complicated processing steps some modifications of LOCOS were studied in the third chapter. First the SILO (Sealed Interface Local Oxidation) process is explored. SILO consists in the nitridation of the pad oxide or in the deposition of a nitride layer in direct contact with the silicon substrate. Secondly LOPOS or poly-buffer LOCOS is studied in detail.

Another drawback of LOCOS, not yet reported in the literature, is the observed asymmetry of the bird's beaks in a closely packed structure. This observation has important consequences for memories and other cells with a high number of transistors, because the electrical characteristics of the transistors at the edges of the structures will differ from the characteristics of the others.

The LOPOS technique has been studied in detail. Structures with two different buffer layer types have been investigated.

The use of polysilicon buffer reduces the bird's beak length by about 20% without substrate defect generation.

A modified LOPOS approach, based on the insertion of a microcrystalline layer between the pad oxide and the nitride, was also explored. A reduction of the bird's beak by 50% is observed. This leads, however, to a strong increase of the stresses at the edge of the bird's beak. These stresses can be high enough to give rise to different stress relaxation mechanisms:

- Breaking of the pad oxide.

- Breaking of the nitride.

- Generation of defects in the substrate.

For temperatures below 950°C breaking of the nitride occurs in combination with breaking of the pad oxide. For temperatures above 1000°C the generation of defects is the dominant relaxation mechanism.

Some preliminary results on the implementation of LOCOS on SIMOX substrates have been presented. The TEM observations show that:

- No defect have been generated in the thin silicon film for the oxidation mask parameters employed.

- The bird's beak is of the same length as for LOCOS on bulk material, even for thin field oxide thicknesses.

- The upper surface of the active silicon is strongly bent, which can be harmful during the following processing steps.

- Once the oxygen species reach the buried layer, enhanced diffusion through the buried layer happens, oxidizing the underlying bulk silicon and the lower interface of the active silicon. This causes an increase of the buried oxide thickness and, thus, a lifting of the active silicon.

In order to implement this process in integrated circuit processing further research is required. Possible approaches to solve the problems mentioned above are:

- Use of a LOCOS stack with a thicker silicon nitride mask or use of any modification of LOCOS in order to achieve a short bird's beak and an interface as flat as possible.
- The field oxidation should be stopped just when the buried layer is reached to avoid oxidation of the bottom of the active silicon layer.

In the fourth chapter two TEM techniques for the delineation of shallow junctions were presented. These techniques are based on the enhancement of the contrast between regions with different dopant concentration. Two approaches have been proposed: TEM in combination with preferential chemical etching and in-situ radiation induced defect generation in the HVEM.

Etchants have been tested with etching rates that depend on the doping level. The best choice is the "Roberts" etch with composition $HF:HNO_3=1:300$ at 5°C. It has been applied both to boron and phosphorus implanted specimens, delineating dopant concentration levels of about $6\cdot10^{17}$ cm^{-3}.

The main drawbacks of the use of chemical etching is its poor reproducibility, due to the dependence on the morphology and aging of the specimen and of the solution.

The second approach for junction delineation developed in chapter 4 is based on the preferential creation of radiation induced extended defects. The extended defect nucleation depends on the dopant concentration, so that they are created in well defined areas of the doped specimens.

Possible explanations for the observed dopant concentration dependent defect behaviour are:

- The generation of defects could be affected by the position of the Fermi level. In the case of boron, in the highly doped areas, the Fermi level lies above an interstitial level. The

generated interstitials would be charged and they would prevent of clustering, while for the lower doped the interstitials are neutral and cluster more easily.

- The formation of boron interstitial and boron substitutional pairs is proportional to the amount of boron present in the sample. For highly doped areas also pairs of boron interstitials are formed and could constitute a strong sink for the generated self-interstitials. This would explain why in the highly doped areas no clustering is observed.

- The area with a high density of defects for boron and phosphorus implantation and with low density for arsenic, corresponds to regions with the same strain, calculated using an elastic model. This indicates that the strain induced by the dopant plays an important role in the defect generation mechanism.

In order to understand the fundamental mechanisms that give rise to this enhanced defect generation in the HVEM, further research is required. This should involve the study of:

- Specimens with a very gradual dopant profile.

- The influence of the covalent radii of the dopants. For this purpose the irradiation of antimony and gallium doped specimens should be investigated.

- The effect of the built-in electrical field of the specimen and the use of an externally applied field.