TRANSMISSION ELECTRON MICROSCOPE STUDIES OF EMITTERS IN SILICON BIPOLAR TRANSISTORS

Daniel Patrick Gold
Keble College
Oxford

A thesis submitted for the degree of Doctor of Philosophy
in the University of Oxford

Trinity Term 1989

TRANSMISSION ELECTRON MICROSCOPE STUDIES OF EMITTERS IN SILICON BIPOLAR TRANSISTORS

Daniel Patrick Gold
Keble College

A thesis submitted for the degree of Doctor of Philosophy
in the University of Oxford
Trinity Term 1989

Abstract

Transmission Electron Microscope (TEM) studies have been carried out of emitter regions in polysilicon contacted emitter bipolar transistors. The preparation of suitable TEM thin foils is described. In addition a technique is developed for the observation and quantitative interpretation of the break-up of the interfacial oxide layers observed in these samples.

The effect of annealing the samples prior to emitter dopant implantation (pre-annealing) is investigated for phosphorus and arsenic doped samples, implanted into a polysilicon layer 0.4μm thick, with a dose of 1×10^{16}cm^2. Two wafer pre-cleans have been used prior to polysilicon deposition to produce a thin oxide (0-8Å) and a thicker oxide (14Å). In the presence of the thinner oxide, the phosphorus doped samples enhance epitaxial regrowth of the polysilicon layer compared with the arsenic doped or undoped samples. In the presence of the thicker oxide, no difference is observed between the samples. A mechanism is proposed to explain this.

The mechanisms controlling the gain of a phosphorus doped device are investigated and a model proposed to explain the observed electrical characteristics. It is concluded that there are two mechanisms responsible for the observed supression of hole current. Firstly tunnelling through the interfacial oxide and secondly some blocking effect of the interface. Carrier transport in the polysilicon is not believed to contribute to this supression.

A dopant sensitive etch has been applied to TEM thin foils containing fully processed emitters in state-of-the-art devices. The shape of the emitter dopant distribution is revealed in such devices for the first time, and a 2-D profile is obtained from the emitter. It is shown that reduction in the emitter depth to 800Åor less does not alter the emitter dopant geometry. The technique is demonstrated to be capable of resolving spatial separations of dopant iso-concentration contours of 100Åor less.

This thesis is dedicated to my friends

This thesis is an account of work carried out by the author in the Department of Metallurgy and Science of Materials, Oxford University, between October 1983 and April 1989, under the supervision of Dr G R Booker. No part of this work has been submitted for a degree at any other University. The work of others has been drawn upon and is acknowledged in the text, and lists of references are included. Some sections of the work have been published as follows:

SEM/TEM Studies of Etched Silicon Specimens to Determine 1-D and 2-D Dopant Profiles Associated with p-n Junctions.

D.P. Gold, J.H. Wills, G.R. Booker, M.C. Wilson and D.J. Godfrey. Inst. Phys. Conf. Ser. 91 p.519-524, 1989.

The Effect of Emitter Sidewall Isolation on the Emitter Junction in a Double Layer Polysilicon Bipolar Process.

M.C. Wilson, D.P. Gold, P.C. Hunt and G.R. Booker, 1988 proc. IEEE BCTM Minneapolis USA p.128.

Some sections of the work have been submitted for publication as follows:

Epitaxial Regrowth of Phosphorus-Doped Polysilicon and its Effect on the Base Current of Polysilicon Contacted Bipolar Transistors.

D.P. Gold, G.R. Wolstenholme, P.Ashburn, G.R. Booker and L. Castaner (Submitted to Applied Physics Letters)

Study of Thin Oxide Tunnel Parameters for Polysilicon Emitters, using Computer Simulation and Experimental Results.

D.J. Roulston, D.P. Gold, P. Ashburn and G.R. Booker (Submitted to Solid State Electronics)

The author would like to thank Prof Sir Peter Hirsch FRS and Prof J W Christian FRS for the provision of laboratory facilities, and Dr G R Booker for supervising the work. The author was supported from 1984-1987 by the SERC. This project was financially supported from 1987 to 1989 by the Alvey

Directorate. The author collaborated with Plessey Research (Caswell) Ltd. particularly with Dr.P Hunt and Dr.M C Wilson and with the Department of Electronics, University of Southampton. Particular thanks are due to Dr. Peter Ashburn and Dr. Graham Wolstenholme at Southampton for many valuable discussions.

Thanks are also due to Graham, Ron, Paul and John for fixing the microscopes after I had used them, and to Andrew and Mohammed for printing the pics. The semiconductor group have been a continual source of encouragement and tea, but for particular help advice support and sailing and cakes, David Roulston, Amanda Petford-Long and Rob Glaisher. Neil Long also deserves thanks for his tireless patience playing with Semper6 to get the results out.

I would particularly like to thank my co-writing comrades Bob Mallard and Martin Lyster for the ice cream and beer and the knowledge that I wasn't alone. Also my family for being there.

Finally special thanks to Sasha for caring.

The text of this thesis was typeset using TEX.

The contribution of Maurice Lighthill's shirt to this thesis is gratefully acknowledged

"Perfect sir, that is to say slightly flawed"

Dorothy L. Sayers

Have His Carcass

Contents

Chapter 1

INTRODUCTION

1.1 INTRODUCTION

In this chapter an introduction to bipolar transistors is given. The requirements for VLSI applications are described, and the operation of the bipolar transistor is discussed, as is the operation of its principal rival, the field effect transistor (FET). The factors controlling switching speed in the two types of device are considered. The polysilicon emitter bipolar transistor is then described together with its principal advantages and difficulties. Finally a description is given of the aims of the work carried out in this thesis including a thesis layout.

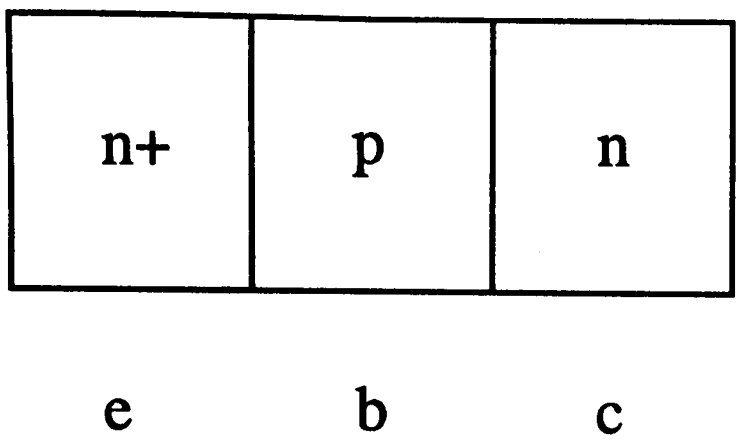
In 1948 a fundamental change took place in the structure of electronics when Bardeen and Brattain created the first transistor [1]. The transistor represented a dramatic move away from the discrete engineering structures which characterised the active elements of the thermionic valve and into a new era where the device active elements were chemically distinct regions within a single piece of material.

The first transistor (from TRANSfer resISTOR) was a point contact device. The diffused contact bipolar transistor followed swiftly however, whilst the field effect transistor (FET) did not appear until the fifties. Semiconductor devices quickly found their way into a number of applications, both as digital switches

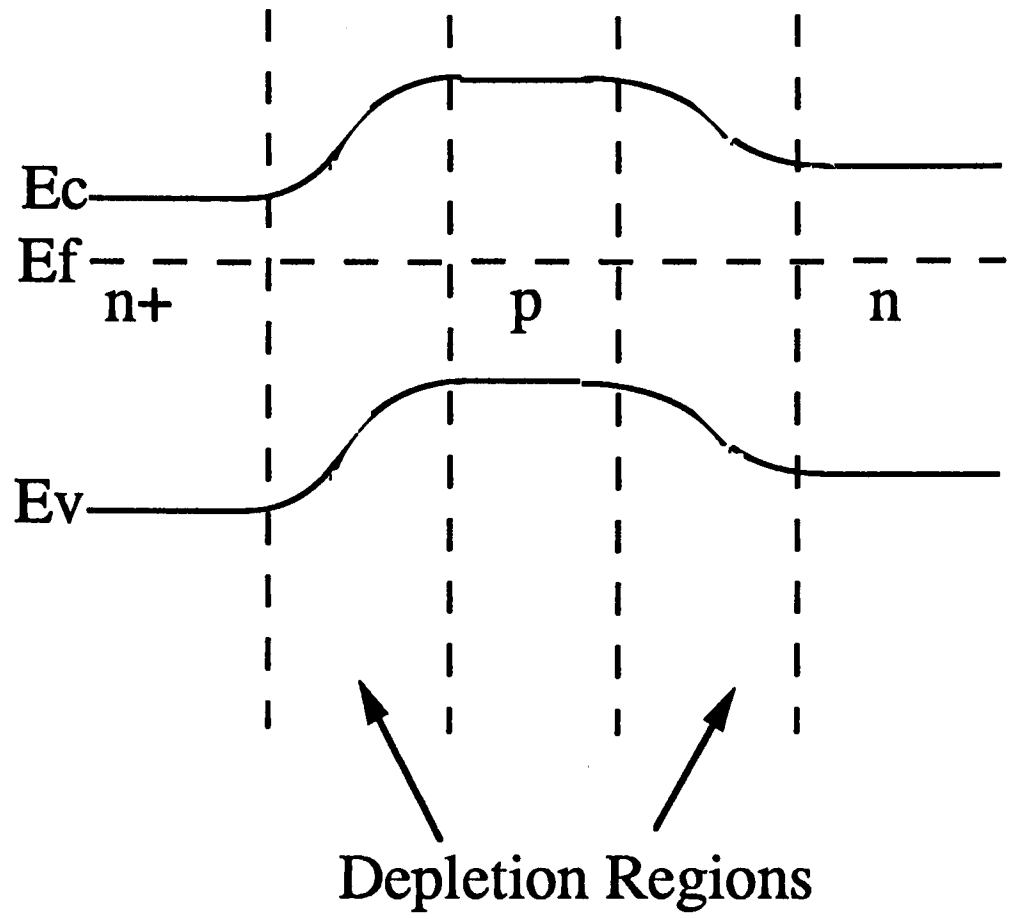
in early computers and as linear amplifiers in a wide variety of analogue applications; in both cases the great reduction in heat generated vastly simplified the design and construction of these systems. The advantages of assembling complete or nearly complete electronic systems on a single slice of semiconducting material were realised with the development of integrated circuits.

Developments in the field of logic integrated circuits have been driven by the industrial requirements for higher electrical performance (particularly speed), and reduced cost. Achieving these design criteria involves a number of different device and chip parameters. Increasing individual device speed is largely achieved by the scaling of device dimensions, this scaling additionally assists in increasing the packing density (active elements per chip) which reduces cost. By this path, levels of chip integration increased initially to large scale integration (LSI), and now to very large scale integration (VLSI). Increasing the level of integration must be accompanied by little or no loss in circuit yield and by sufficient reliability to satisfy the market. This involves not only the development of reduced dimension fabrication technology, but also of new interconnect and contacting techniques, and indeed, packaging systems, especially for ultra high speed applications, as discussed below.

Early semiconducting devices were based on either silicon or germanium material. Germanium has now largely disappeared from the market except in specialist applications, and silicon is the most widely used and well characterised material in the field. This success is largely due to the ease with which silicon can be processed, particularly with regard to the purposeful growth of oxides and nitrides. Other semiconducting materials with useful electronic properties have been developed, for example gallium arsenide (GaAs) which, in contrast to silicon, has a higher electron mobility and a direct band gap. Such materials are limited in application to specialist areas such as radiation hardness, microwave devices and optoelectronics. These materials often have particular processing

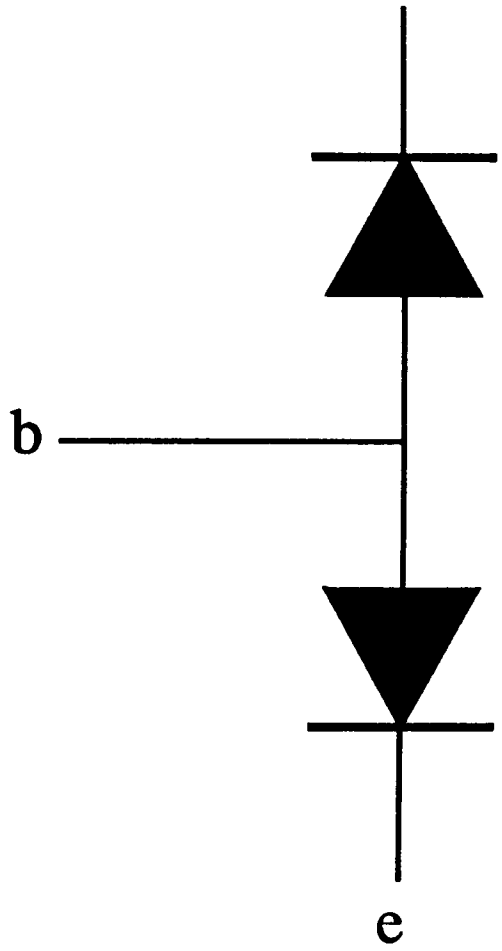


(a)

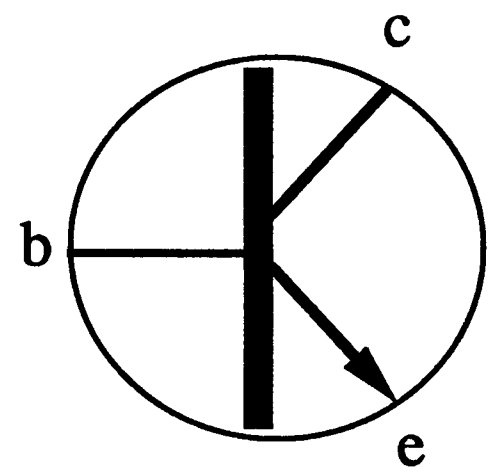


(b)

c



(c)



(d)

Fig.1.1

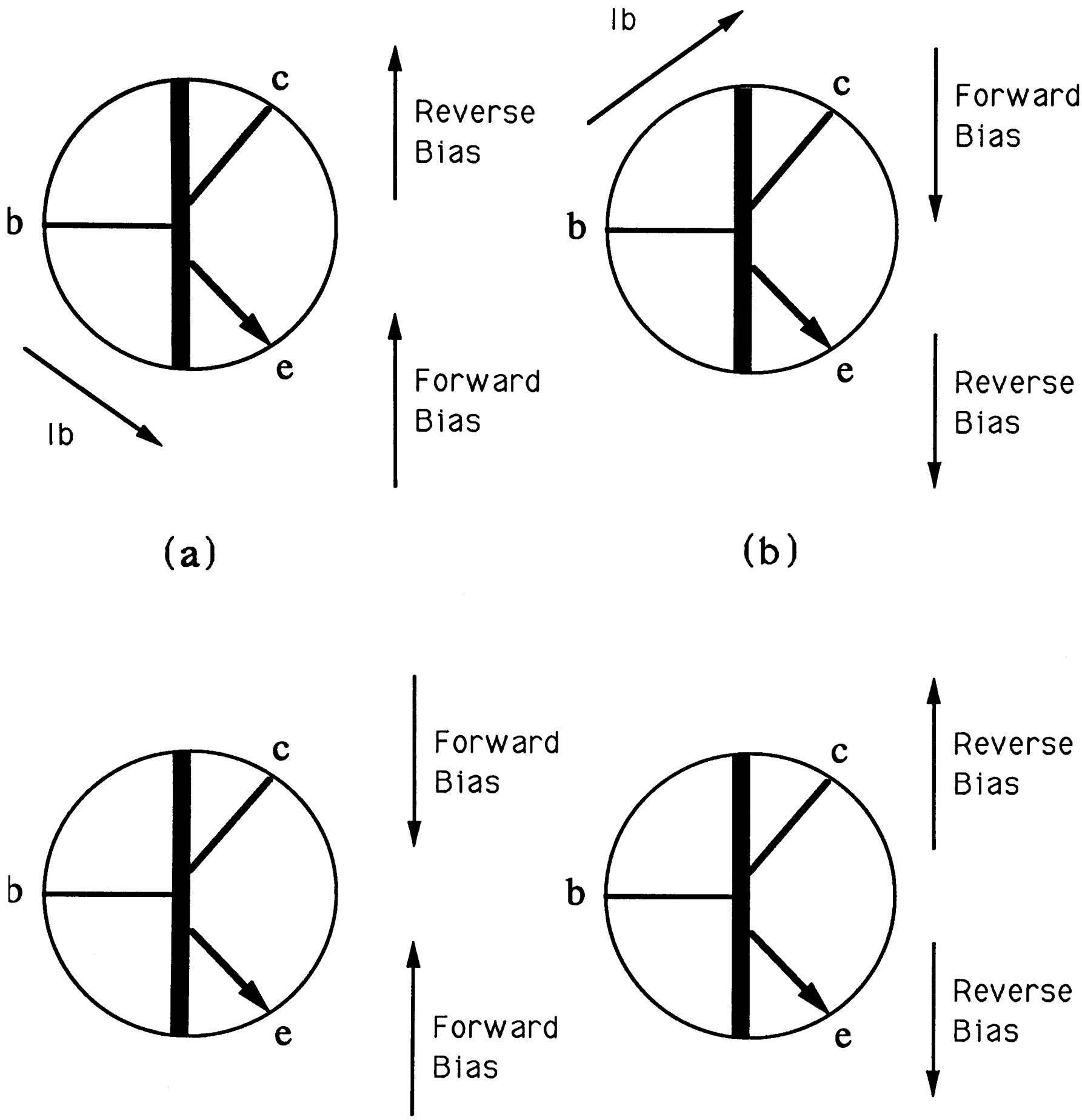

c
b
e
Ib
Reverse
Bias
Forward
Bias
(a)
Ib
c
b
e
Forward
Bias
Reverse
Bias
(b)
c
b
e
Forward
Bias
Forward
Bias
(c)
c
b
e
Reverse
Bias
Reverse
Bias
(d)

Fig.1.2

difficulties such as elemental loss at elevated temperatures or the lack of suitable native oxides and nitrides.

Semiconductor manufacture has evolved a planar technology in which devices are defined photolithographically on the surface of the silicon wafer as discussed in chapter 5. The work in this thesis is based on the development of bipolar silicon planar technology in high speed devices for use in computers. In the following section a brief description will be given of the operation of a bipolar device.

1.2 DEVICE OPERATION

1.2.1 OPERATION OF BIPOLAR DEVICES

A basic knowledge of semiconductor physics is assumed in this section. A variety of text books exist which give good descriptions of both the basic physics and device operation [2] [3] Fig.1.1(a) shows a schematic representation of a npn bipolar device. Fig.1.1(b) shows the equilibrium band structure of the device and fig.1.1(c) illustrates the simple equivalent circuit comprising two diodes back to back, fig.1.1(d) illustrates the schematic symbol used for these devices. Obviously there are four possible combinations of bias which may be applied to this device, as illustrated in fig.1.2.

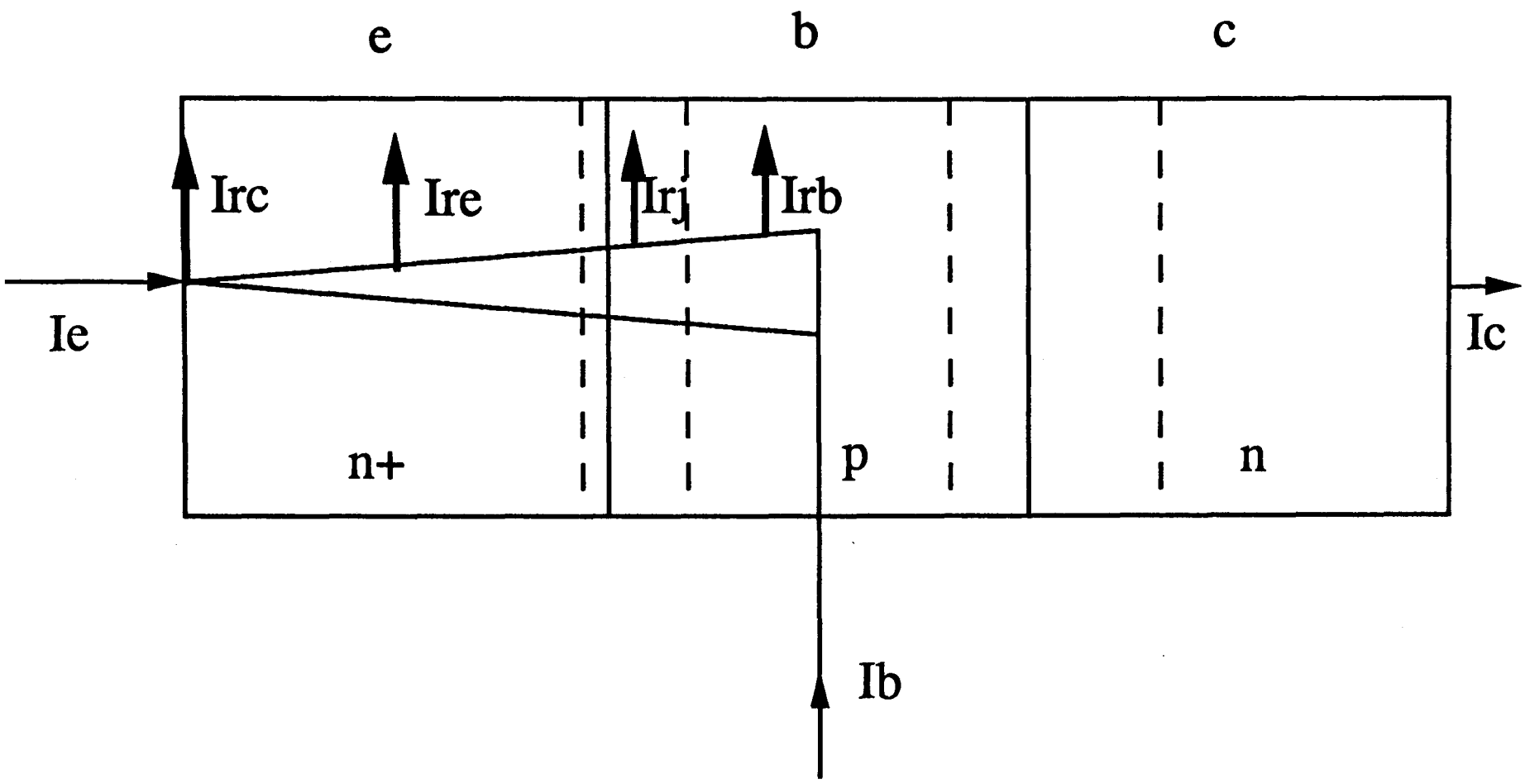
Considering first the forward active region, when the emitter/base junction is forward biased and the base/collector junction is reverse biased. Fig.1.2(a) illustrates this mode, and indicates the current flowing at the three terminals. I_E is the injected electron current, I_C is the ejected electron current, and I_B is the injected hole current, this is equivalent to an electron current of the opposite

sense at the base terminal. In such a situation, electrons are injected across the forward biased junction, from the emitter region into the base region. Once in the base region, they become minority carriers and diffuse across the base to the collector. The electrons which reach the reverse biased base/collector junction are swept across into the collector by the applied field where they once again become majority carriers. The hole current is injected into the base, the field on the base/collector junction blocks the hole current whilst the field on the emitter/base junction sweeps it across. The total hole current can be expressed by considering the various contributions of different device regions to recombination as detailed below; the holes must of course recombine within the device. The common emitter current gain of the device (β) is expressed as the ratio of the collector current (I_C) to the base current (I_B), that is, the ratio of the electron current at the collector terminal to holes injected into the base. Gain occurs primarily because the base is narrow compared with the minority carrier diffusion length and therefore the majority of the electrons can transit the base without recombining. The gain is however dependent on majority carrier effects as well. These are often expressed by the emitter injection efficiency. A useful expression for gain in vertical bipolar devices is given by the expression;

$$\beta = \frac{D_{nb} W_E N_{de}}{D_{pe} W_B N_{ab}}$$

Where D_{nb} is the diffusion coefficient of electrons in the base, W_E is the emitter width, N_{de} is the donor concentration in the emitter (active doping level), D_{pe} is the diffusion coefficient of holes in the emitter, W_B is the base width and N_{ab} is the acceptor concentration in the base (active doping level).

In the reverse active mode (fig.1.2b), the device operates similarly but electrons are injected from the collector into the base. Because of the assymetric construction of real devices, this does not result in similar current gains to the forward active mode. In the saturation mode (fig.1.2c), both junctions are forward biased and a large current can flow from the emitter to the collector. The



Contributions to hole current in an npn transistor

Fig.1.3

device may considered in this condition as a closed switch. When operating in this mode, a large charge is contained within the device. The discharge time associated with this makes the switching speed in this mode of operation relatively slow. For this reason, devices operating in high speed logic applications are constrained by circuit design from entering saturation. Finally, in the reverse inactive mode(fig.1.2d), both junctions are reverse biased and the current flow is very low, this is approximately equivalent to a switch that is off. It is important to realise that the device may be wired into a variety of circuit combinations which can allow different combinations of modes. The most commonly used circuit configuration is the common emitter mode, where the emitter is connected to the supply rail and the load is connected to the collector. Other possible configurations are common collector and common base, these will not be dealt with further here. Where gain is discussed in this thesis, the reference is always to common emitter gain (β).

Fig.1.3 illustrates the various current distributions within the device. Examining firstly the hole current, the total hole current injected into the device is the base current I_b. The holes must all recombine within the device, and so the hole current can be broken down into recombination in the different regions of the device. I_{rb} represents recombination in the neutral base region. I_{rj} represents recombination in the emitter/base space charge region. I_{re} represents recombination in the emitter and I_{rc} represents recombination at the contact. It can be seen that;

$$I_b = I_{rb} + I_{rj} + I_{re} + I_{rc} ;$$

In a useful device the base is very narrow compared with the electron diffusion length, and so I_{rB} is very small and is usually neglected.

Two recombination situations may occur in the emitter. If it is wide, I_{re} consists of recombination in the bulk of the emitter. If it is narrow, some of the holes reach the emitter contact where they must recombine. The rate at which

they recombine is controlled by the nature of the contact interface. The interface effects giving rise to this recombination are often lumped together into a single term called the surface recombination velocity, S_p. This term is important in that it allows any effects occuring more than a set distance from the junction to be considered as the recombination velocity of an interface in the emitter, at that distance from the junction. The surface recombination velocity term is used by many authors when considering the behaviour of devices, as discussed in chapter 4.

The electron requirements for recombination of the holes must be supplied by the emitter current. The collector current is therefore given by the emitter current minus the electron current associated with the recombination processes. Or;

$$I_C = I_E + I_{rj} + I_{rE} + I_{rc}$$

The recombination currents are additive as they are hole currents and therefore of opposite sense to the electron current.

1.2.2 DIFFERENCES BETWEEN BIPOLAR AND FET OPERATION

The operation of the FET relies on the conduction through a bar of semiconducting material, known as the channel. The current flow is injected at one end of the channel (the source) and leaves at the far end (the drain). Conduction is modulated by a third contact (the gate) producing a field in the channel. There are two possible conduction mechanisms in the channel, depending on the type of device, inversion and depletion. In the former, the channel doping level is low, and the conducting layer is formed by using the field to create an inverted layer below the gate containing free carriers. In the second, the channel doping level is high, and conduction occurs when no field is applied to the gate. As a voltage is applied to the gate, a depletion region forms in the conducting channel

below the gate. As the voltage is increased the depletion region spreads through the channel until it reaches the far side at which point no further conduction occurs and the device is said to be 'pinched off'

There are two principle ways of constructing the gate, either by forming a diodic junction (p-n or Schottky) or by creating a capacitive contact using a thin layer of insulator (oxide or nitride). So the important aspects of a FET are that conduction is by majority carriers only and the operation is controlled by the applied voltage (rather than the injected current in the case of a bipolar device).

In the logic market, the principle objectives are speed, both in individual devices and through novel circuit architecture, and economy, by increasing levels of integration whilst maintaining yield. Both technologies suffer similar problems in terms of interconnects, where the reduced dimensions have increased the interconnect resistance and hence, with parasistic capacitance, an RC network is created which results in propagation delays between devices. Similarly, as contact pads are scaled, the difficulty of making ohmic contacts increases, resulting in exotic combinations of contact materials.

The switching speed of an FET is controlled principally by two effects, the input capacitance which must be charged/discharged to turn the device off, and the transit time of the remaining carriers across the bar once it is off. This transit time is related to a number of effects, one of which is the physical separation of the source and drain. These are defined lithographically and so the minimum conventionally defined linear dimension is limited at present to approximately $1\mu m$, the pattern transfer limit of existing technology.

In a bipolar device the same two effects dominate, namely input capacitance and transit time; the subject of capacitance is discussed further below. The transit time in a bipolar device is controlled by the width of the base (W_B). As bipolar devices are constructed vertically, it is possible to construct a base as shallow as desired by controlling the diffusion of the dopants. The minimum

dimensions of the base are in fact limited by the neccessity of preventing the two depletion regions from overlapping in the base during operation; this effect is known as punch through and is tantamount to a short circuit between the emitter and collector.

1.3 POLYSILICON CONTACTED BIPOLAR TRANSISTORS

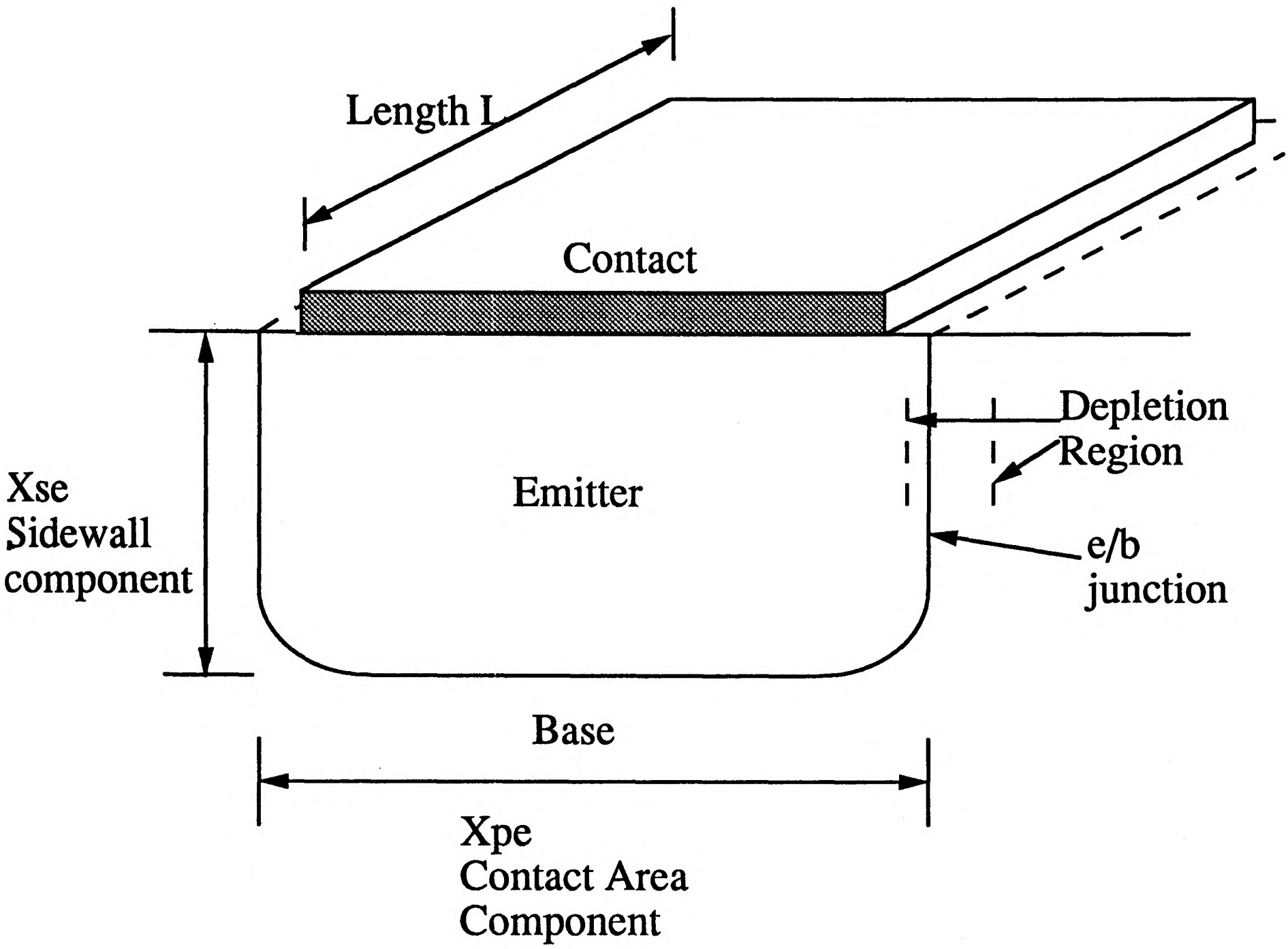
1.3.1 POLYSILICON

Polysilicon, whilst still a semiconducting material has unique properties which separate it's behaviour from that of monocrystalline material. Polycrystalline materials contain grain boundaries, which alter the properties of the material. A good perspective on grain boundaries in semiconductors has been written by Grovenor [4] The theoretical and practical investigation of grain boundary structure is a burgeoning field, and it is not intended to cover it here except to note that there are two generally accepted structural models. The older is the coincident site lattice model (CSL), this model defines grain boundaries by the ratio of actual lattice site separation in the plane of the boundary, to the separation of coincident lattice sites (i.e. coherent bonds) between the two grains, this number is written as Σ=n, where n is the CSL number as defined. The second and more recent model, is the structural unit model (SUM). This model is an intellectual development of the CSL model. In the case of the SUM, a three dimensional atomic structure is visualised at the boundary such that the structure of the boundary can be completely defined by a repetetive array of the chosen SUM.

The effect of the grain boundaries on the electrical properties of silicon is twofold; firstly, it's resistivity is higher than that of monocrystalline material of the same nominal doping level and secondly it can exhibit enhanced recombination, especially at low doping concentrations; however it is not intended to discuss enhanced recombination further here. There are two models, not mutually exclusive, to explain the effect of grain boundaries on carrier transport, the theory of grain boundary segregation and the theory of grain boundary trapping. In the former model, dopant atoms are assumed to segregate to the boundaries where they become inactive. This can result in a reduction of the active dopant concentration within the grains giving rise to an increase in resistivity at the grain interiors, in effect a lower average doping concentration. Segregation of dopant to grain boundaries has been observed by a number of workers, including Baccarani et al [5] and Wong et al [6].

The model of grain boundary trapping assumes that a large number of trapping states exist at the grain boundary. These states are quickly filled by carriers and the Fermi level is pinned, resulting in local band bending at the interface. This may be considered as equivalent to a pair of Schottky barriers back to back, the filled traps having the same effect as the surface states in the Schottky barrier. Carrier transport may then proceed either by thermionic emission of carriers over the potential barrier or by tunnelling through the barrier if it is sufficiently narrow. Seto [7] for example produced a model involving such a thermionic barrier. Seto assumed that the traps were monovalent and approximately centred in the gap, as observed later by Baccarani.

Whilst it now seems clear from the result of numerous experimental results, such as those detailed above, that there are states in the band gap associated with the grain boundary, the origin of these states is less certain. These states have been investigated by a number of workers who have attributed different energies within the band gap to these states, many authors have either assumed or measured states centred around mid-gap. Classically it has been assumed that



Emitter region of a conventional npn transistor

Fig.1.4

these states are an intrinsic property of grain boundaries, i.e. they are associated with the boundary disorder, perhaps in the form of 'dangling' or unsaturated bonds. However, recent work by Paxton and Sutton [8] in modelling the band structure of grain boundaries has suggested that there are no intrinsic defects in silicon grain boundaries which result in the presence of states in the gap in silicon. The implication of this is that the states in the gap are associated with extrinsic effects, for example, the segregation of impurities to the grain boundaries.

The observed increase in resistivity (equivalent to a reduction in mobility) in polysilicon is of critical importance in the discussion of minority carrier behaviour in chapter 4. The behaviour of dopant atoms in polysilicon is also different from their behaviour in monocrystalline material. The diffusion rate of dopant atoms along grain boundaries is known to be significantly greater than the diffusion rate in the bulk [9], and the presence of a heavily doped polysilicon layer on a monocrystalline substrate therefore acts as an effective diffusion source for the substrate.

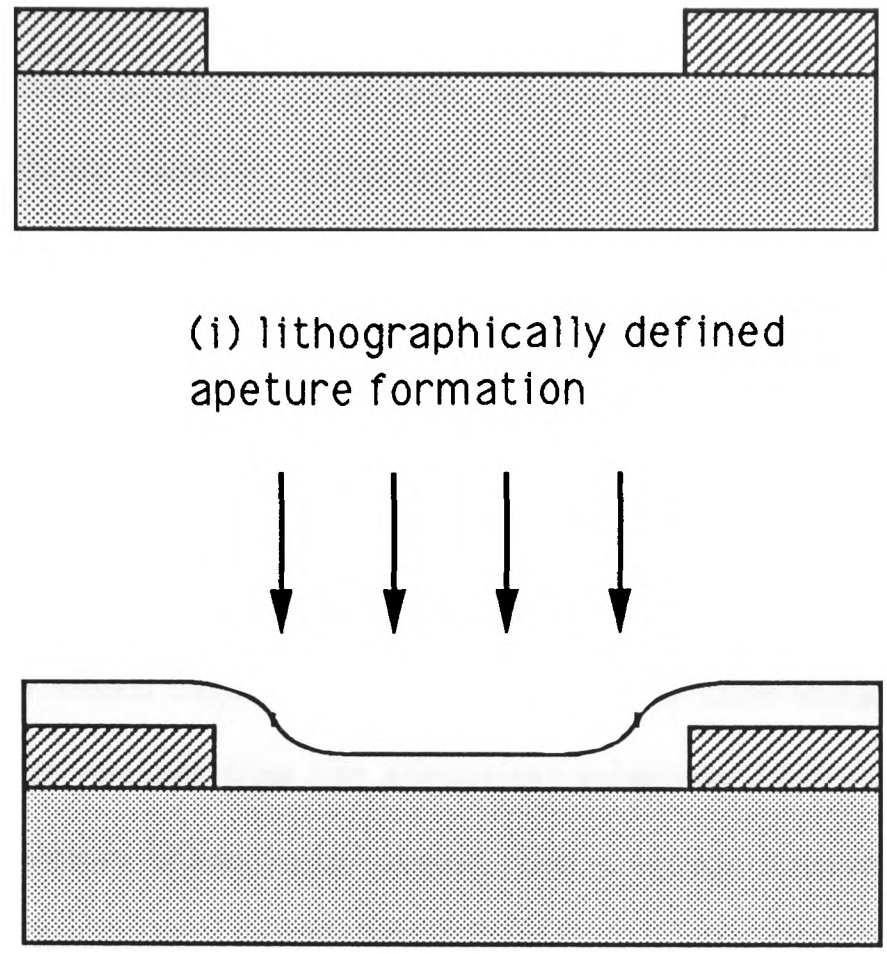
1.3.2 IMPROVING SWITCHING SPEED

The capacitance of the depletion regions and the resistance of the base have a strong influence on the maximum operating frequency of bipolar devices, through the rate of charge and discharge of the resulting RC network. There is therefore a need to reduce the size of these parameters in high speed applications. In the devices investigated here, the diffusional properties of polycrystalline silicon are utilised in the first instance to reduce the area and hence the capacitance of the emitter/base (e/b) depletion region. A typical emitter region from a nominally conventional device is illustrated in fig.1.4. It can be seen from this that the area of the e/b depletion region is controlled by two components, the width

and length of the e/b junction, and a sidewall component related to the depth which the emitter extends into the base. The former component is controlled by the emitter shape and is fixed during manufacture by the mask dimensions (but is not neccessarily of the same dimensions, see below). The latter component may be considered as an undesirable corollary of the junction formation.

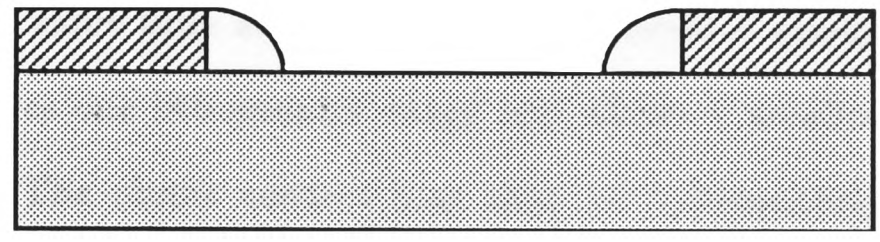
In a polysilicon contacted emitter, the base is implanted/diffused and the wafer surface is then cleaned to remove any native oxide; this is successful to a greater or lesser extent but some oxide always remains which may influence device performance (see chapters 3 and 4). A layer of polysilicon is then grown on the patterned wafer and implanted with the emitter dopant. The wafer then receives an emitter heat treatment drive-in. During this drive-in, the dopant diffuses rapidly down the grain boundaries to the interface with the single crystal substrate. From here it diffuses much more slowly into the bulk. Thus the polysilicon acts as a diffusion source for the formation of the emitter. Additionally it serves as an emitter contact. By careful control of the emitter drive-in the diffusion of the emitter into the substrate can be reduced to provide minimal penetration, and hence minimal sidewall component to the depletion region. Junctions as shallow as 100Å have been reported [10].

It is of course possible to produce a shallow junction by other means; for example, low energy ion implantation, however such a junction must still be contacted. S_p of an ohmic contact is very high, and the presence of such a contact close to the e/b junction acts as sink for holes. This results in a large hole current, and so normal contacting technology is not suitable for the fabrication of ultra-shallow emitter devices due to the unacceptable increase in hole current. The presence of a silicon or polysilicon contact to the emitter however, results in an S_p which is the same if not lower than that of a single crystal emitter[3]. This may be more easily visualised as a conventional deep emitter device with most of the sidewall removed from the e/b capacitance term. It is of course possible to fabricate a polysilicon contact onto an implanted emitter, although



(i) lithographically defined
apeture formation

(ii) layer deposition followed
by anisotropic etch



(iii) reduced width apeture

Reduction of apeture dimensions below lithographically defined dimensions

Fig.1.5

this process is not inherently self-aligning (see below). Such emitters are referred to as polysilicon contacted emitters. The type of emitter discussed above is usually termed a polysilicon emitter, although strictly speaking this too is a polysilicon contacted emitter. The two names merely serve to differentiate the fabrication processes and in general physical terms, they are very similar.

In the photolithographic definition of devices, there are two important physical constraints on device dimensions. Firstly the minimum definable feature size and secondly the alignment error or repeatability of the mask overlays. This means that any region to which contact must be made, must have linear dimensions equal to those of the contact plus the alignment tolerance (assuming that the two stages are consecutive lithographic steps). For a square contact having linear dimension L and alignment error δL, the increase in contact area δA, and hence projected area of depletion region is;

$$\delta A = \delta L(2L + \delta L)$$

The use of polysilicon contacts reduces these problems in two ways. Because the contact is formed at the same time as the emitter, the process is inherently self aligning and this removes one lithographic stage and allows a smaller emitter contact to be formed. The ohmic contact may then be subsequently formed onto a larger area of polysilicon. A further advantage arises because the emitter window into which the contact polysilicon grows is not defined lithographically. The initially defined one micron window is reduced in size at the point of contact with the substrate during the formation of the sidewall isolation; the process relies on the anisotropic nature of certain plasma etches, as detailed in fig 1.5. In the present work, emitters with windows as narrow as 0.4μm were formed. Both of these effects result in a reduction in the projected area of the emitter, which along with the reduction in sidewall component detailed above, reduce the total area of the depletion region and allow higher switching speeds [11].

A similar technology may also be used to contact the base [12]. After the base has been implanted into the substrate, a layer of polysilicon sometimes known as the poly 1 layer is grown on the wafer. This layer is implanted with p-type dopant and then lithographically defined. It then receives a base drive-in and diffuses into the single crystal material in the same way as the emitter contact. After sufficient annealing, the dopant from the contact reaches the implanted base dopant and the contact is formed. In such cases, the contact profile is generally referred to as the extrinsic base, and the actual base is referred to as the intrinsic base. It is important however to realise that this is only a contacting technology and has no self aligning properties. Following these stages, the processing proceeds for the emitter as described above. In such devices, the emitter polysilicon is referred to as the poly 2 layer and the whole technology is referred to as 'double polysilicon contact technology'. When utilising this technology, isolation of the emitter and base becomes important and a variety of isolation technologies may be applied to the manufacturing process in addition to thermal oxides. This is detailed in chapter 5.

1.4 THESIS CONTENT

The work in this thesis was carried out as part of a collaborative intiative within the Alvey program, project title Alvey 061. The collaborating bodies were Oxford University, Southampton University and Plessey Research Ltd (Caswell). The aim of the work was to develop a new bipolar device, using polysilicon contacting technology and 1μm optical lithography, to operate in VLSI logic applications, with a maximum switching speed of 20GHz. The final pre-production devices fabricated at Plessey using a novel isolation technology designed by Dr. Peter Hunt were capable of switching at speeds of up to 22GHz.

The author was preceded at Oxford by Dr. Norman Jorgensen, who was involved in the structural investigation of arsenic doped devices, and also of novel contacting technologies, such as platinum silicides.

In this thesis, a variety of techniques have been applied to the study of these high performance bipolar transistor emitters, based around the transmission electron microscope (TEM). In particular, two distinct approaches were undertaken. Working with Plessey on fully fabricated, state of the art, arsenic doped devices, a technique was developed, which for the first time allowed the dopant profile in a real device structure to be examined. Collaboration with Dr. Peter Ashburn and Dr. Graham Wolstenholme at Southampton University involved structural investigation and characterisation of real devices and device-like blanket specimens, doped with phosphorus, fabricated at Southampton. These results were compared and contrasted with D.C. electrical results from the device specimens and with results for arsenic doped samples. The work is detailed in this thesis, and a model is proposed to explain the electrical behaviour of devices with discontinuous interfaces and also with regrown polysilicon layers. It is also demonstrated that in principle, a device may be fabricated using phosphorous as the emitter dopant with significant electrical advantages over a similarly doped arsenic sample.

The analytical techniques used are described in chapter 2. The variation of polysilicon/monosilicon interfacial structure, polysilicon morphology and the oxide behaviour with dopant and annealing conditions is examined in chapter 3 using the TEM, and a device is proposed, as detailed above, using phosphorus as the emitter dopant, having an epitaxially regrown emitter with no increase in the emitter's thermal budget from a conventional, arsenic doped device. In chapter 4 it is shown that a device of this type may be fabricated with useful D.C. characteristics. In conjunction with this, an investigation of the effect of structure on electrical performance in phosphorous doped emitters is performed. The various contributions to hole current during structural change are

considered and a model is proposed to explain the measured electrical performance. Chapter 5 shows the emitter and base structures in a fully processed, high speed bipolar device in the TEM. It then describes the use of a dopant sensitive etching technique to delineate the doping profile in the emitter region and the intrinsic base region, and produce two dimensional profiles of the dopant distribution under the emitter and the base contact for the first time in a real device. This technique is applied to two similar types of device with slight variations in architecture and it is shown that variations in surface architecture may have an effect on emitter shape. The effect of reducing the emitter depth on the diffused emitter geometry is also investigated and it is shown for the first time that the shape of deeper diffused emitters is preserved when the emitter depth is scaled for the devices investigated. Finally chapter 6 lists the conclusions and suggestions for further work.

References

[1] . J. Bardeen and W.H. Brattain, 'The Transistor, a semiconductor triode', Phys Rev., 74, 230(1948)

[2] . S.M. Sze, Physics of Semiconducting Devices. Wiley(1981)

[3] . P. Ashburn, Design and Realisation of Bipolar Transistors. Wiley(1988)

[4] . C.R.M. Grovenor, J. Phys. C, 18 p.4079, 1985.

[5] . G. Baccarini, B. Ricco and G. Spandini, 'Transport properties of polycrystalline silicon films,' J.Appl.Phys., vol 49, p.5565(1978)

[6] . C.Y. Wong, C.R.M. Grovenor, P.E. Batson and D.A. Smith, M J.Appl.Phys. vol.57 p.438(1985)

[7] . J.Y.W. Seto, 'The electrical properties of polycrystalline silicon,' J.Appl.Phys., vol 46,p.2457(1975)

[8] . A.T. Paxton and A.P. Sutton, Acta Met. 37(7), pp.1693-1715,1989.

[9] . W.J.M.J. Josquin, P.R. Boudewijn and Y. Tamminga, Appl.Phys. Lett. vol.43, p.960(1983)

[10] . A. Cuthbertson and P. Ashburn, IEEE Trans Electron Devices. ED32, p.2399(1985)

[11] . S. Konada, Y. Yamamoto and T. Sakai, IEEE Trans. Electron. Devices, vol.ED-33, p.526(1986)

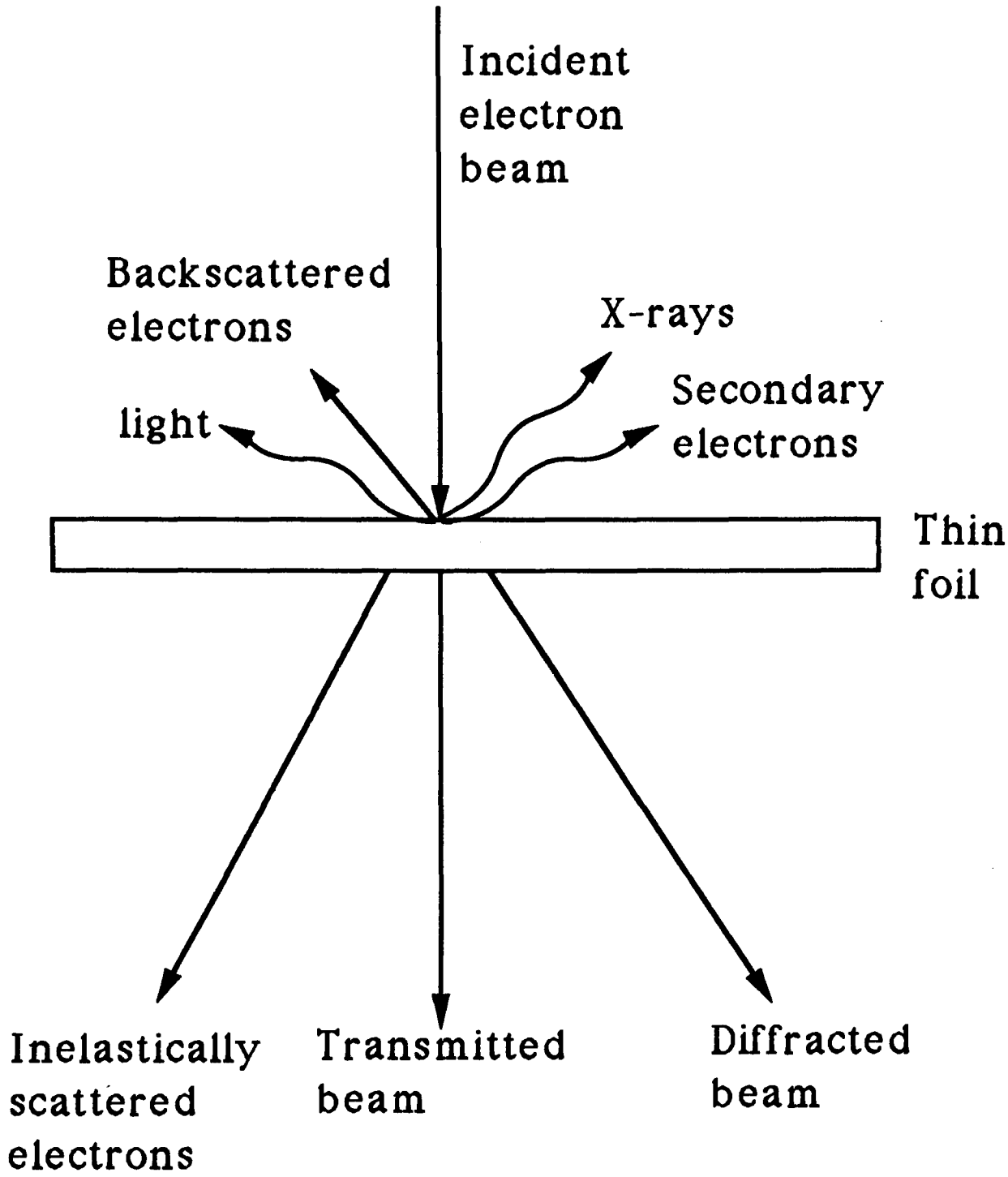
[12] . H.K. Park, K. Boyer, A. Tang, C. Clawson, S. Yu, T. Yamaguchi and J. Sachitano, Proc. BCTM p.39(1986)

Chapter 2

EXPERIMENTAL TECHNIQUES

2.1 INTRODUCTION

The techniques discussed in this chapter may be broken down into two broad areas, sample preparation and analytical techniques. The preparation of both cross-sectional (XTEM) and plan view (PVTEM) transmission electron microscope (TEM) samples is described, as is the dopant sensitive etching technique employed to produce the results discussed in chapter 5. A variety of analytical techniques have been employed in the present work. The bulk of these utilise Transmisson Electron Microscopy (TEM) in some form. These may be divided into three areas, diffraction contrast, high resolution electron microscopy (HREM) and convergent beam electron diffraction (CBED). The other techniques discussed are secondary ion mass spectroscopy (SIMS), spreading resistance (SR) measurement and the technique used to measure the area of coverage of the broken interfacial oxide layer as used in chapter 4.



Electron/specimen interactions

Fig.2.1

2.2 THE TRANSMISSION ELECTRON MICROSCOPE

The TEM column consists of essentially three areas. An electron gun containing a thermionic source, a simple electrostatic lens (the wenhelt) and an accelerating anode with a stable potential of 100-400kV between it and the cathode; all other lenses are electromagnetic. This beam then passes through a series of probe forming lenses (the condenser lenses) which allow the beam size and convergence angle to be controlled. This beam is incident on the specimen and the resulting electron beams produced on the other side of the specimen, pass through the image forming lenses (the objective lens and the projector lenses); the final image is displayed on an electron sensitive phosphor screen or is exposed onto a photographic plate. A detailed description of this equipment can be found in a variety of good text books on the subject [1] [2]. The Phillips CM12 used to produce most of the diffraction contrast results is equipped with an electromagnetic scanning system which allows the incident beam to be scanned in a raster across the specimen. A scanning transmission image (STEM) can then be obtained using the detector fitted to the microscope. This system is computer controlled and a facility is available to stop the beam on any given region marked by a set of movable computer generated cross-wires. This technique was used to define the point from which convergent beam thickness measurements were taken. The high resolution electron microscope images were obtained from a JEOL 4000EX operating at either 300kV or 400kV, or from a JEOL 200CX operating at 200kV. When the electron beam is incident on a thin foil, a variety of events may occur. These are detailed in fig.2.1. All the TEM techniques used in this thesis rely on the information contained in one or more of the diffracted beams.

2.3 TEM SAMPLE PREPARATION

To examine a sample in the TEM it is neccessary to produce a foil sufficiently thin to transmit the incident electron beam. If the sample is insufficiently thin, the transmitted beams will contain a large percentage of inelastically scattered electrons, which contribute noise to the final image and blur fine detail. For diffraction contrast examination of silicon a sample of 1000Å or less is desirable, although it is possible to produce useful images from samples up to approximately 3000Å thick. For HREM imaging, samples should be less than 200Å thick and preferably as little as 50Å. All the samples examined were taken from (100) silicon wafers, and two two types of foil were prepared, (100) plan view foils and (110) cross-sectional foils. All the samples were prepared using a modified version of the technique of Pettit and Booker [3] and Fletcher [4]. The techniques employed are discussed in some detail as small variations in these techniques can result in failure to produce a useful sample and this has been observed by the present author to be one of the most difficult aspects of the field to master.

It is important to realise that the analytical area in TEM is very small and is therefore susceptible to statistical scatter, indeed it has been estimated that the total volume of material ever observed in the TEM is of the order of one cubic millimeter. Thus it is neccessary to test results for consistency with other samples from the same or similar groups wherevever possible. This generally applied technique is not intended to mask real, macroscopic changes in sample behaviour. Howevér in cases where unusual or unexpected differences are observed between samples, it is sometimes neccessary to re-examine a series of samples, thus it is hoped that 'real' changes may be either ratified or refuted. This statistical 'risk' is an inherent part of all microscopy.

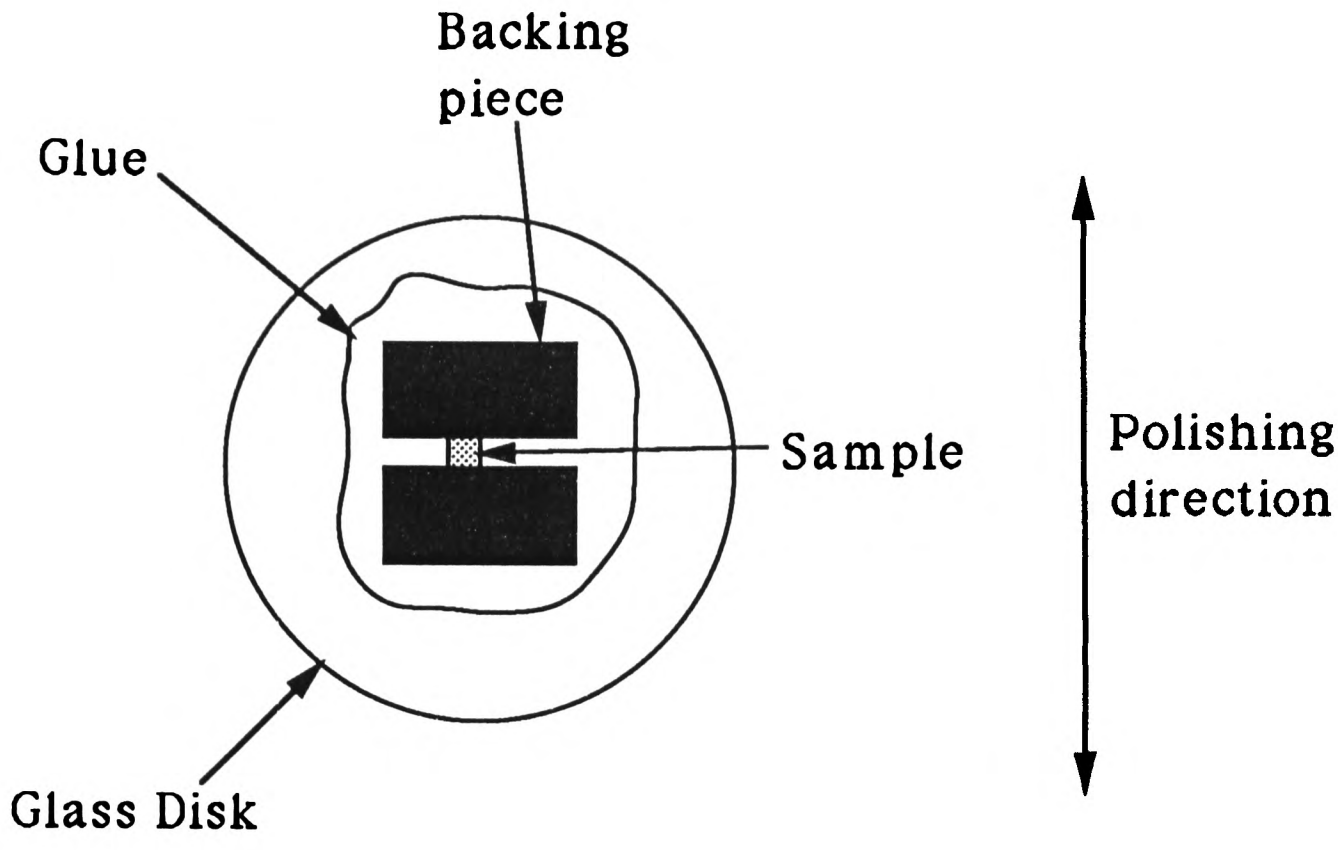
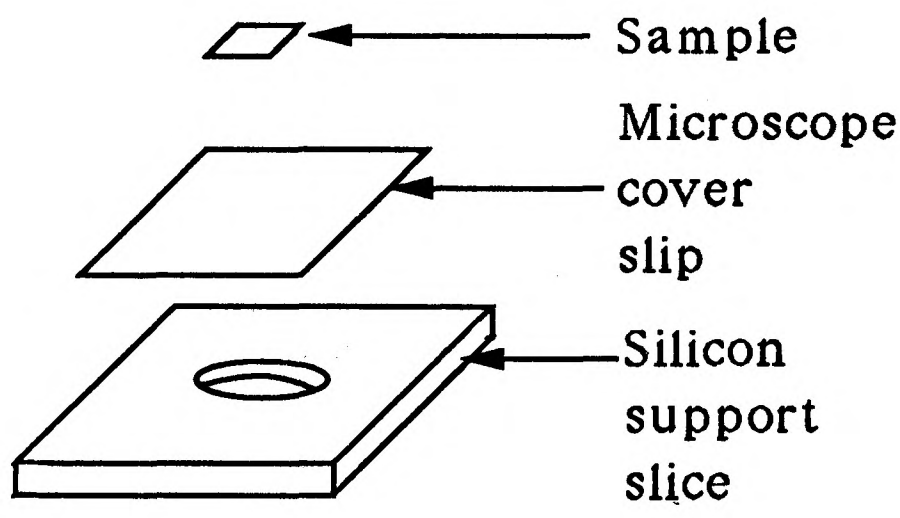


Fig.2.2



Exploded view of plan-view sample support

Fig.2.3

2.3.1 PLAN VIEW SAMPLE PREPARATION

Plan view samples were prepared by cutting a small piece of the sample material approximately 3mm×3mm or less using a diamond scribe to score a guide on the wafer surface prior to fracture. The resulting sample was mounted on an optically flat glass disc approximately 5 cm in diameter with the surface of interest against the glass, between two backing pieces of scrap silicon (fig.2.2). The adhesive used was Crystalbond 509; this is a thermoplastic wax which melts at approximately 120°C and is soluble in acetone. The sample was then ground down to a thickness of approximately 40μm using a commercial rotating platten covered in 600 grit wet and dry paper constantly lubricated with water. The sample was hand held during the grinding operation. The sample was checked regularly to ensure that the ground face was parallel to the surface of interest. This was initially measured using a micrometer, but touch was found to be satisfactory to maintain an approximately flat surface. The final thickness of the specimen was estimated by polishing the sample until the outside edges of the backing pieces could be seen to begin todissappear when viewed through the glass disc during grinding. This gave a sample thickness close to 40μm. If one side was observed to be removed more rapidly, then the pressure on the disc was adjusted to balance this.

This technique was found to give suitably flat specimens of approximately 40μm thickness. These samples were then polished on a commercial rotating platten running at full speed with 6μm diamond polish using a water based lubricating fluid (oil based lubricants were believed in some cases to be reducing the adhesion of the waxes employed) until the characteristic coarse grinding damage could no longer be observed in the optical microscope. From this point, the samples were polished for approximately 3-5 minutes more. The samples were finally polished using 1μm diamond paste in a similar fashion until the

6μm polishing marks were eradicated (again, as observed in the optical microscope). This grinding and polishing technique was satisfactorily used in all cases, no artefacts were observed in any foils which it was believed could be attributed to this process. When finished, the sample was immersed in a beaker of Dimethyl Formamide (DMF) on a hotplate at approximately 150°C. After a period of not less than one hour the sample could be removed from the disc and cleaned by gently brushing with a fine sable hair brush (size 000), wetted with hot (150°C) DMF. Four different sable hair brushes were used for sample preparation. Cleaning as described was carried out using an as supplied 000 size brush, the samples thinned by polishing were maneouvered and manipulated using another 000 brush lightly wetted in methanol or DMF, but in this case modified with a scalpel to produce a thinner, more tapered tip. A third brush was cropped to a length of approximately 2mm, this was used for the application of Lacomit as described below, its smaller fluid capacity allowing thin layers to be applied. Finally a brush was cropped to leave two single full length hairs, this sample was used to remove the films of polymerised Lacomit which sometimes adhered to the ion beam thinned samples even after cleaning.

A holder was then prepared using a piece of scrap silicon with an ultrasonically drilled hole. This hole was drilled using the standard tool designed for cutting 2.3mm sample discs. A piece of glass cover slip was stuck to this as shown using Lacomit (an air drying acetone soluble resin, not unlike nail varnish). The sample was, in turn, glued to this cover slip with the surface of interest facing the cover slip, so that it was aligned over the hole in the scrap silicon, again using lacomit. Fig.2.3 illustrates the complete holder. This construction was allowed to air dry for approximately 30 minutes. It was then mounted on a single sided holder as described by Roberts [5], for use in a Gatan 600 Duomill ion beam thinner (IBT).

Ion beam thinning was carried out using a single gun at an angle of approximately 17° to the surface with the sample rotating. The sample was then

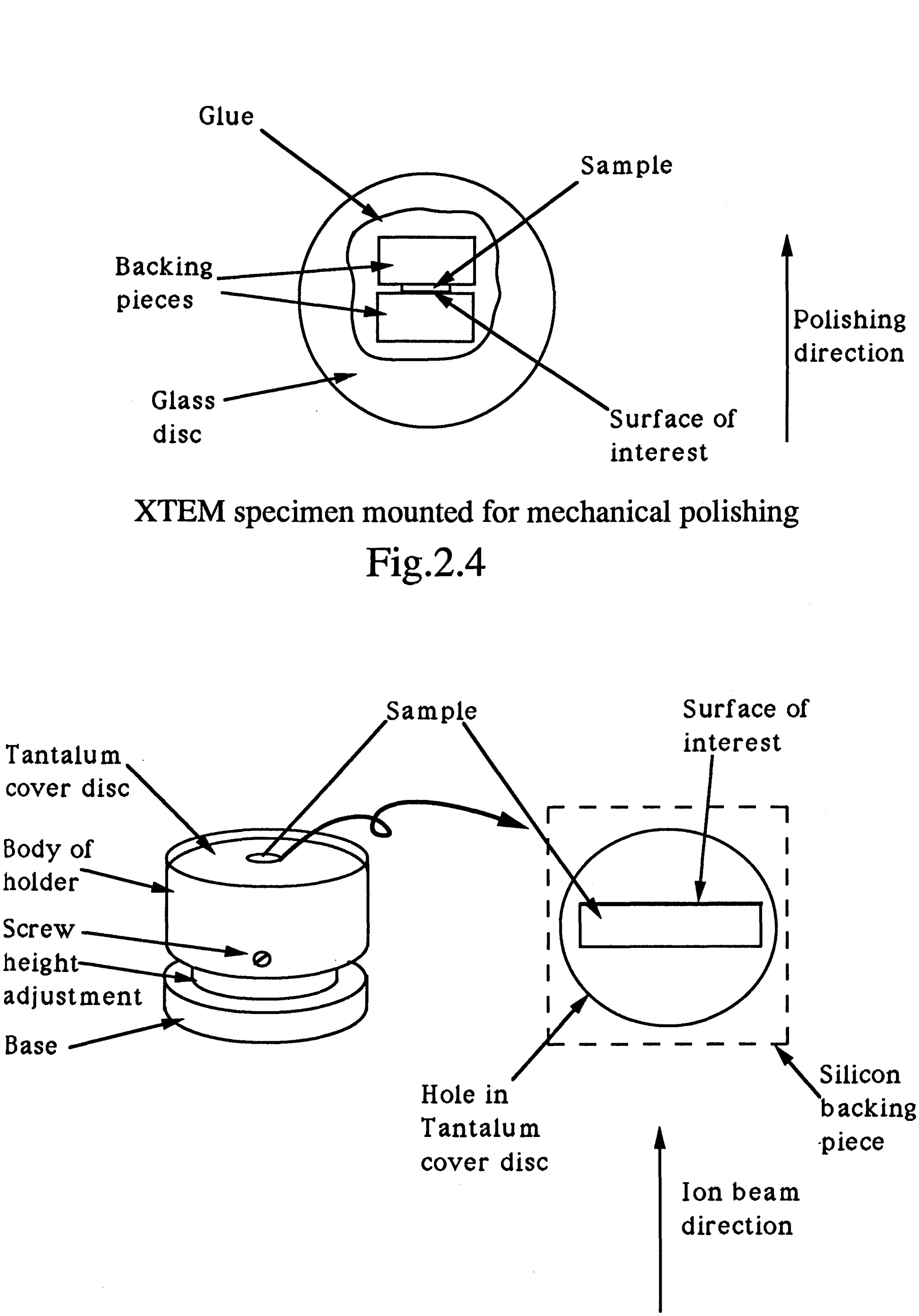
thinned until perforation. For plan view specimens, perforation was detected using the Gatan laser terminator attachment. This shines laser light through the airlock window and onto the sample. When the sample has perforated the laser light is detected by a light sensor on the far side of the sample and the ion beam is turned off. The sensitivity of the terminator is abitrarily adjustable from 0-10. With a clean airlock window, a sensitivity of approximately 3.0 was found to terminate the specimen close to perforation. However, if the sample is milled for more than approximately eight hours, the window must be re-cleaned or the light transmission is reduced by sputtered material on the viewing port, and the perforation can become larger than desired. The milled surface of PVTEM foils often has a mottled appearance; this was not found to cause significant problems when imaging the foils in the TEM.

Exposure of the lacomit to the ion beam causes it to polymerise. It is therefore neccessary to immerse the sample in hot DMF for at least 24 hours after milling. After this time the sample can usually be removed from the holder by gentle agitation with a fine sable brush, taking care to avoid contact with the thin area of the sample. Samples thinned to this point can only be moved by floating off filter paper on a pool of liquid (in this case Analar methanol). A 3mm copper TEM mounting grid (2x1mm slot) is then degreased in methanol and has a thin layer of a cyanocrylate adhesive (Tixo K) applied. The grid is then touched onto the milled side of the specimen, so that the perforation is over the slot. Care must be taken when mounting specimens that only a thin layer of glue is used on the grid, or the surface tension of the glue can move the specimen on the grid and result in the perforation sticking to the copper.

In the case of the oxide balling measurements, the oxide layer to be examined is covered by 0.4μm of polysilicon, and this must be removed to allow accurate examination of the interfacial oxide. This is carried out by mounting the specimen on a grid in a double sided Gatan holder and milling only the surface of interest using identical conditions to those described above. The time

XTEM specimen mounted for mechanical polishing

Fig.2.4



Polished XTEM sample mounted for ion milling

Fig 2.5

required is estimated by assuming a milling rate of approximately 5μm an hour. So to remove a layer 0.4μ thick requires approximately 8 minutes. Following this, the sample was given a two sided mill at 3kV and 0.25mA per gun for 3-4 minutes. This removes the material sputtered through the perforation during the previous milling stage. Samples prepared in this way provided a simple process that produced suitable samples with the minimum of supervision.

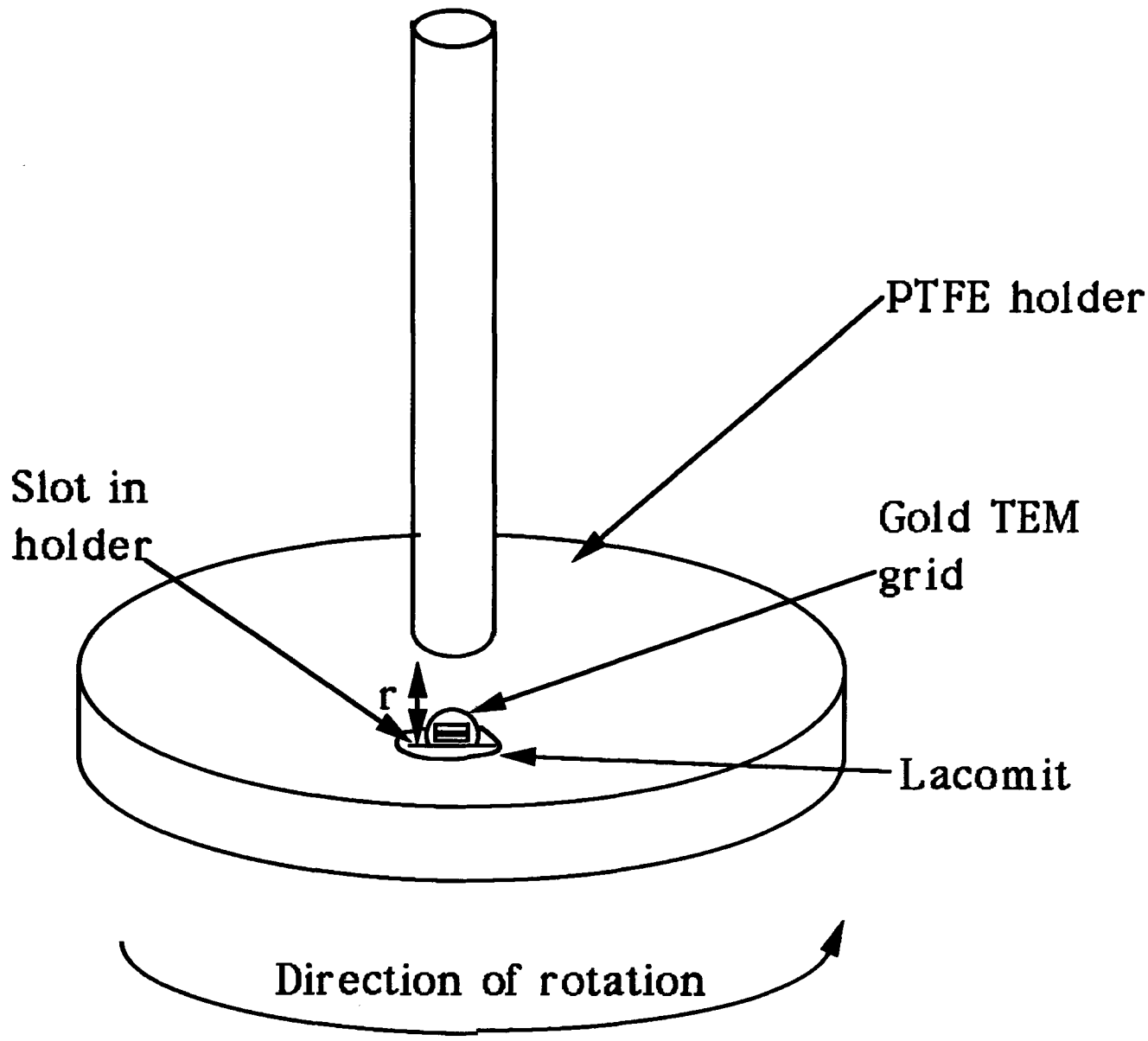
2.3.2 CROSS-SECTIONAL SAMPLE PREPARATION

Cross-sectional samples were fabricated by a similar technique to plan-view samples. A small piece of the sample was produced from the wafer by cleaving, and mounted on its edge on an optically flat disc (similar to that used for PVTEM sample preparation), between two backing pieces of scrap silicon as shown in fig.2.4. Generally these backing pieces were of different size and/or shape to each other and the surface of interest was pushed against the larger piece to facilitate easy identification during grinding. It is important that the sample remains approximately perpendicular to the disc while the adhesive cools. The sample was then ground down until the backing pieces were in contact with the grinding paper. The grinding direction was always towards the surface of interest as illustrated, to protect it from excessive grinding damage. The sample was polished as described above. The glass disc was re-heated on the hot plate ($\approx$150°C) until the wax softened, and then the sample was carefully turned over using wooden cocktail sticks, so that the polished face was now against the glass and the surface of interest was still against the larger backing piece. Care must be taken that the polished surface is in close contact with the glass disc. This was then ground down as before to approximately 40μm and again polished. The sample was removed from the disc as before.

A sample 3mm or less in length is required and the ground and polished sample was cut by placing it on filter paper wetted with methanol and touching

it at the appropriate place with a diamond scribe. The methanol soaked filter paper prevents the sample moving when it is broken. The sample was carefully cleaned on both sides using a sable haired brush dipped in hot DMF and placed on a clean slab of scrap silicon. This was then mounted in a single sided Gatan holder as detailed above and as shown in fig.2.5. The holder was orientated, using the specimen rotate system until the surface of interest was facing away from the beam as shown. Single gun milling was then carried out at 5kV with a beam current of 0.5mA for approximately two hours. After this, the sample was removed from the slab. The slab was covered with a thin layer of Lacomit and the milled surface of the sample applied to this. After drying, the sample was reinserted in the mill on the holder, as detailed above. Milling was then recommenced until perforation occured at the interface. The existence of a suitably thin region could be confirmed by the presence of white light fringes at the perforation edges. The sample was mounted on a TEM grid and then re-inserted into the ion mill and given a final low energy mill exactly as detailed for the plan-view samples. This was found to result in less amorphous material on the foil surfaces when examined in the TEM. Occasionally XTEM foils would perforate initially from the back surface. This results from bevelling of the sample during mechanical preparation. Continuing to ion mill the specimen results in the perforation moving forward through the sample until it meets the front surface. This generally produces two half foils containing thin area. Such samples were usually found suitable for examination in the TEM, although the amount of thin area is significantly reduced from the optimum possible.

Sample prepared for etch delineation followed essentially the same preparation routine as above, with a few important changes. The copper grids were entirely unsuitable as they were eaten away by the etch and so gold grids were used. Cyanaocrylate adhesives such as Tixo K were found to be sensitive to the etch used and therefore foils were bonded to the grids using a two part epoxy glue designed for strain guage mounting; called M-bond 600. This glue must be



Teflon holder for the etching of TEM foils

Fig.2.6

cured at 150°C. The active components of this glue are carried in a solvent base and the evaporation of the solvent can leave a thin coating of epoxy on the foil surface. This film is impervious to the etch of course, and therefore must be removed before etching. The only effective method ascertained for this is a low energy mill. In this case, using both guns, set at 3kV and 0.25mA per gun and rotating the sample on a double sided holder, for approximately 5 minutes.

2.4 ETCHING OF TEM SPECIMENS

Etching of TEM specimens was carried out to delineate dopant profiles as detailed in chapter 5. The technique used is a development of the process described by Roberts for the etching of TEM foils [6]. A holder was fabricated similar in dimensions to that used by Roberts, but instead of a composite holder, one was machined from a single bar of PTFE, as the original holder was found to degrade in the etch. Wills [7] has shown that the radial position of the sample has a strong effect on the shape of the etching vs. dopant curve, therefore a circle was scribed on the surface of the holder centred on the axis of the holder with a radius of 17mm. This provided a marker for repeatable positioning of the specimen. Mounting the sample whilst still on its backing piece, as described by Roberts was used briefly but was felt to leave doubt as to whether the protected face was in fact unetched. To avoid this problem, the samples were first mounted on gold grids as described above and then positioned vertically in slots cut in the holder at the marker radius (fig.2.6) and this additionally allowed simple re-etching of specimens previously examined in the TEM. The samples were secured using Lacomit. Care must be taken to avoid the Lacomit creeping onto the thin area of the foil under surface tension.

Etching was carried out using the TEM etch as described by Roberts consisting of 0.3%HF and 99.7%HNO_3 cooled to 5°C. These measurements are by

volume from stock solutions of Analar reagents (40% and 69.5% respectively). $100cm^3$ of etchant was prepared freshly for each etching experiment, the HNO_3 was measured using a $100cm^3$ measuring cylinder and the HF was measured using a micropipette previously calibrated according to the manufacturers instructions. The etch was cooled in a thermostatically controlled water bath with a maximum heating capability of 2kW, containing a 1kW continuously operating immersion cooler. The fluid in the bath was a 50/50 mix of tap water and commercial ethylene glycol anti-freeze. Hollow plastic spheres floated on this fluid to reduce heat loss. Temperature control throughout the bath was $\pm 0.3°C$. The sample and holder were pre-chilled in a refridgerator to approximately 3-5°C immediately before use.

A dummy run of the experiment using distilled water instead of etchant showed that the water temperature in the beaker was 0.7°C higher than in the bath and varied by approximately $\pm 0.6°C$. Insertion of the chilled holder resulted in a temperature rise of 0.5°C and the temperature rose by a further 0.3°C during the course of a two minute experiment. Therefore the bath temperature was set at 3.5°C and the etch was allowed a period of 10 minutes for the temperature in the beaker to stabilise before etching commenced; additionally the nitric acid used was pre-cooled in the water bath for at least two hours. The holder was rotated exactly as described by Roberts, using a "Mecchano" electric motor, geared at 60:1, driven from a 6V supply to give a rotational speed of approximately 1.8Hz.

The termination of the etch by removing the holder from the etch and immersing it in distilled water, as described by Roberts, was felt to be a potential source of error due to the relatively long period (up to ten seconds) between removal from the etch and immersion in the water. Such a situation leads to non-linearity in the etch as the highly doped regions are depleted of etchant more rapidly, whilst the lower doped regions may continue to etch. To avoid this, termination was achieved by spraying distilled water from a wash bottle

onto the centre of the rotating holder immediately after removal from the etch, whilst still suspended over the beaker for approximately twenty seconds. After twenty seconds, the holder was given a final immersion in a one litre beaker of distilled water. The sample was then removed using acetone, and finally rinsed carefully in Analar methanol; acetone was found to be too dirty to use as a final clean.

A number of problems were encountered in the development of this etch, and these are detailed in this section. Initially it was found that the sample would not remain attached to the grid, and a variety of adhesives were tested for resistance. The acid test involved a five minute immersion in the etchant leaving the foil still securely attached. The adhesives tested were Tixo K, normal and rapid epoxy (from RS Components), crystalbond 509, Lacomit and M bond 600. Only the latter was succesful, the Lacomit being unsuitable in thin films. However, the use of M-bond 600 entailed a further low energy mill for the reasons detailed in the preparation section. The original intention of this work was to carry out an in-situ calibration of the etch by direct foil thickness measurement using CBED, however, it was found that the sample would no longer etch after these measurements. This was attributed to the polymerisation of hydrocarbons by the electron beam. Strenuous efforts were made to ensure that the sample was clean when placed in the microscope by cleaning it in a variety of organic solvents, such as trichloroethylene at 100°C and diethyl ether. Additionally, the cold trap on the microscope was always filled at least one hour before insertion of the specimen, and the specimen was left in the column for at least half an hour before the beam was turned on. The problem however persisted, and has therefore been attribued to contamination from the vacuum system of the CM12. As this was the only machine available with a facility for repeatable positioning of the probe, the technique was abandoned for direct measurement of the etch rate. When observing the samples in diffraction contrast, a diffuse

beam was used wherever possible; samples have been successfully re-etched after such imaging.

The surface preparation of the samples to be etched, turned out to be critical in producing useful results. For this reason, a standarised pre-etch low energy ion mill on both sides of the specimen was used. Madl [8] has suggested that ion milling the surface has a significant effect on the etch and therefore renderes the technique unsuitable for ion milled specimens. However in that work Madl has used an incident beam angle of 45°, and it is felt that such high incident ion angles may indeed result in unnacceptable levels of damage. In the present work, and in the work of Roberts, no effects were observed which it was felt could be attributed to damage caused by the 17° or less, incident ion beam used in the present work.

It was observed that etching did not occur in some specimens, even after a low energy mill and before insertion in the microsope. To try and detect the source of the problem, the technique was returned to that used by Roberts, including the original specimen holder, mounting technique, and cooling system. All containers were carefully cleaned, rinsed in distilled water and dried using disposable paper towels. The technique was then updated step by step to the present system, but no discernable pattern emerged as to the cause of the failure to etch. As a further stage, a variety of chemical surface cleans were used on the ion milled samples in the event that a surface film was preventing the etch. These included organic solvents as detailed above and HF, both Analar (40% by volume) and buffered in the ratio of 1:6 by volume with a 40% (by volume) solution of NH_4F. Again, no pattern emerged; however the Analar HF had a marked effect on samples which had already been etched, as detailed in chapter 5. It was possible to tell if the buffered HF had acted on the foil by observing the device oxides in the TEM. If the sample had been in contact with the solution, the capping and isolation oxides of the device would be preferentially removed from the original wafer surface, and this was generally observed in such cases,

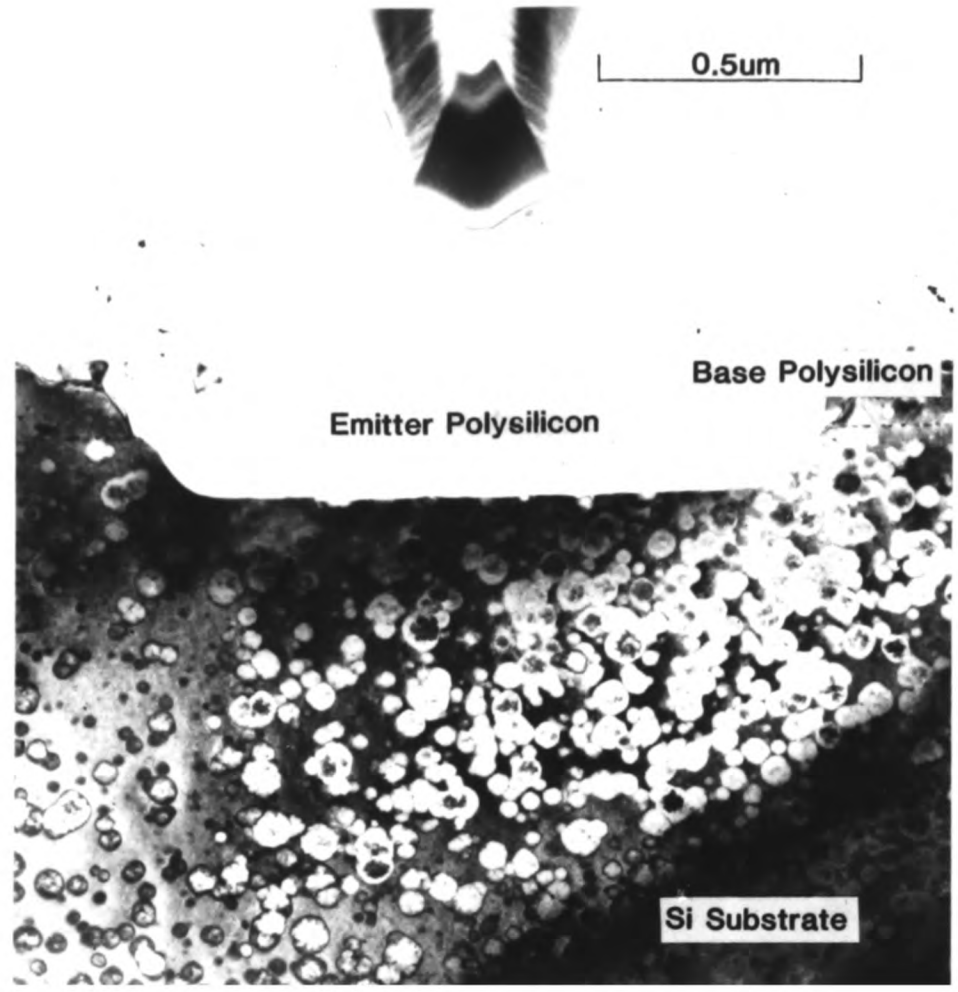


Fig.2.7
Random Etching in a TEM Foil

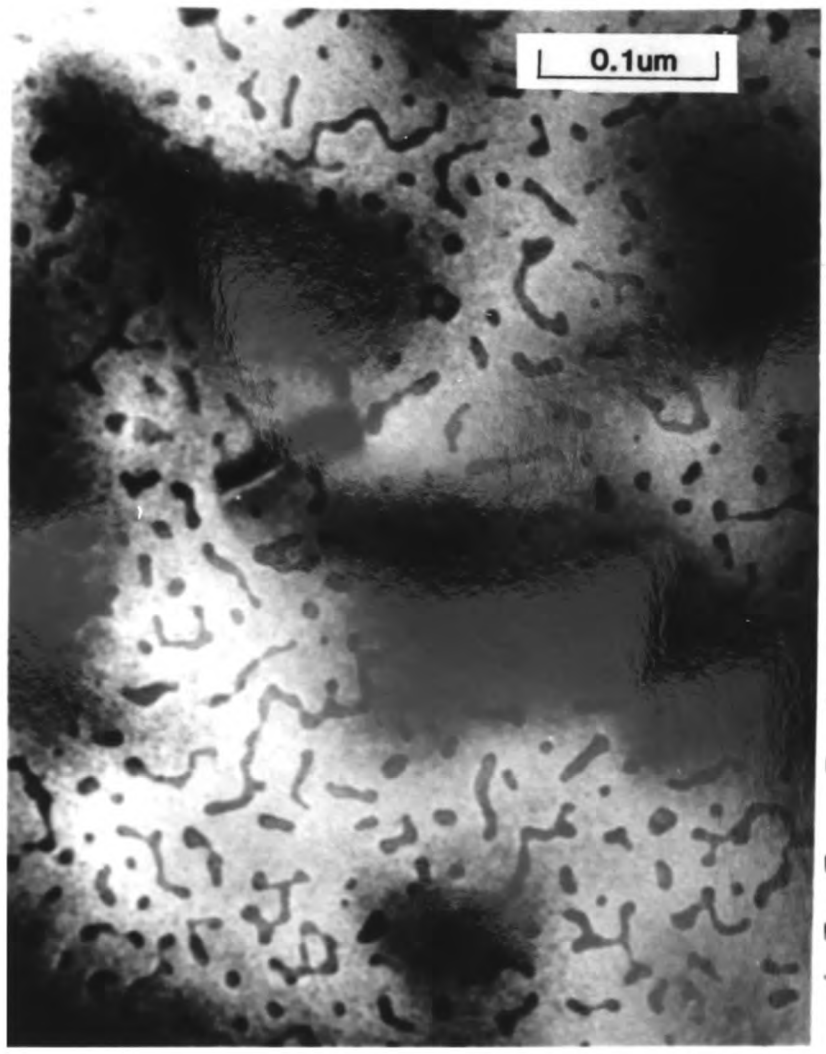


Fig.2.8a
PVTEM of Broken Oxide Layer

Fig.2.8b
High Contrast Tracing of 2.8a

however the delineating etch often did not work after such treatments. This strongly suggests that surface films are not responsible. It is now believed that the failure to etch may have been caused by electrical contact (or the lack of it), between the foil and the gold grid. This may be tested in the future by completely coating a gold grid in M-bond 600, allowing it to cure, and then attaching the foil, thus ensuring no electrical contact between grid and foil.

It was sometimes observed that specimens which had received the delineating etch produced a random etch pattern of pits and blobs superimposed over the desired etch patterns. These could distort the neccessary fringes and prevent analysis, however the existence of this non-linearity was easily observed; as shown in fig.2.7. The repetition of this behaviour from sample to sample declined as the surface clean was developed, but occurred from time to time, even in the most carefully prepared samples.

2.5 MEASUREMENT OF INTERFACIAL OXIDE BEHAVIOUR

In this work, the structural behaviour of polysilicon contacted bipolar devices has been investigated. One of the most important components of these devices is the interfacial oxide which exists between the single crystal silicon substrate and the polysilicon layer. During thermal treatment, the oxide breaks up, and in this work, the break up of this interfacial layer is directly measured to provide quantatative information about oxide particle size and the percentage of interface not covered by oxide.

Plan view TEM samples were prepared as described above, including the additional ion milling stages to remove the polysilicon capping layer. The sample was then examined in the TEM using bright field (001) pole conditions. The area of interest was found by moving the area under examination away from the

perforation until polycrystalline material was imaged; this could be confirmed by inspection of the diffraction pattern as discussed below. The area of single crystal material immediately adjoining this region was found to contain the oxide particles, as expected. Images of this type have been obtained previously by both Jorgensen [9], and Albu-Yaron [10]. In this work, these micrographs were then image processed using a Semper6 image processing package on a Dell 386 microcomputer. The Semper 'analyse' routine was used to measure projected particle area. From this, knowing the total analysed area, it was possible to obtain the percentage of the interface covered by the oxide. Particle size was obtained by normalising the results to a common, nominal magnification. A ruler was placed under the Semper camera, and linear magnification measured from this in pixels. It is important of course to remember that a pixel is a two dimensional object, and when measuring areas the linear measurement of the pixel must be squared. When carrying out particle size distribution analysis, the statistical limits introduced by the relatively small analytical area was addressed by increasing the histogram channel size until the histogram assumes some sort of regular shape. As in all TEM, the small analytical volume introduces the neccessity of some reliance on recognition of patterns in the samples investigated.

The presence of additional contrast in these images from stacking faults and some remaining polysilicon complicated the process of data aquisition. To extract data from the 'raw' micrographs involved a great deal of careful image processing, despite the shape and distribution of the oxide particles being immediately discernable to the eye (fig.2.8a). To avoid this problem, the images were traced by hand onto acetate overlays, to produce high contrast images (fig.2.8b) more suitable for the image processing package.

To test the reliability of the hand tracing, a micrograph was selected at random and copied on four separate occasions. On each occasion, the micrograph was inserted at random in a group being traced. In all cases, the micrographs being examined were randomly mixed and assigned a code number not directly

related to the micrograph number. The four tracings from the same micrograph yielded a spread in measured area within ±3% of the mean. When measuring particle size it is neccessary to set a mean background level for the Semper system; this involves a level of interpretation of the image by the user. However, tracings re-measured over a period of several months were found to produce a spread in measured area of only ±2%. In one case, an attempt was made to measure the coverage of the oxide layer in a sample with a relatively low level of break-up. Repeatable contrast was found in the immediate vicinity of the polysilicon layers, which was attributed to the interfacial region. However, it was not possible to define which part of this region constituted regrowth and which oxide. The sample was treated as described, and the result found to be approximately 50% (see chapters 4). Whilst this result is fortuitously useful, particle size analysis, if indeed the layer had broken up, could not be meaningfully applied.

2.5.1 DIFFRACTION CONTRAST

The subject of TEM imaging and interpretation of normal diffraction contrast imaging has been covered in great detail by a number of excellent texts. Of particularly interest to the practical microscopist are Loretto[1] and Eddington [11]. It is not intended to quote verbatim from these here. In the present work, four types of imaging condition have been applied, many beam imaging using only the central transmitted beam from a sample aligned on the pole, two beam imaging in which only the central spot and one other are excited. This means that the information in the image results almost entirely from diffraction due to the planes associated with the excited spot, or inelastic scattering. The second component is controlled by selecting a suitably small aperture. The third technique is weak beam imaging, in which a weakly excited diffraction spot is

used to image the specimen. Weak beam imaging results in a smaller extinction distance than two beam or many beam imaging conditions. The fourth imaging technique, high resolution electron microscopy (HREM) is discussed later.

Contrast in the image for a single crystal sample arises principally from local distortion of the lattice around crystalline defects or inclusions. The bent planes are then under different diffraction conditions to the bulk and thus contrast is generated. The image observed does not have the dimensions or neccessarily the geometry of the defect, as the imaging process is indirect i.e. it is an observation of strain in the lattice. Contrast may also be generated by regions of different composition as the energies scattered into the different beam directions are dependent both on the crystal structure and the atomic number. The third source of contrast is inelastic scattering. As the energies of the electrons are changed during this process, the diffraction conditions for these electrons are different as is the effect of the lenses on them. This results in a loss of contrast and resolution. To avoid this, thin samples and small objective apertures are used. Contrast in polycrystalline materials arises from both defect structure, and from the variation in crystal orientation between grains; this places each grain in a different diffracting condition, grains which are in a strongly diffacting condition appear dark in bright field imaging. Diffraction patterns of polycrystalline material show a radial distribution of small spots, sometimes referred to as a specular pattern, at the principle d spacings of the single crystal material.

2.5.2 CONVERGENT BEAM ELECTRON DIFFRACTION

Convergent beam electron diffraction (CBED), was used in the present work to measure TEM foil thickness, as the technique features only briefly; a simple description is given. The technique used is as described by Tanaka and Terauchi [12] after the work of Jotsons and Kelly [13]. A suitable two beam condition is obtained, and a highly convergent beam is applied at this condition to the specimen. In the present work, the beam stop facility in the CM12 was used to position the convergent probe at a known position relative to a known device on the foil, and at least four measurements taken from the region of each device. The diffraction pattern produced by a convergent beam in the two beam condition consists of an incident disc and a diffracted disc. The diffracted disc contains detail in the form of a number of light and dark bands. These bands result from the large number of beam incident angles, thus the Bragg condition is satisfied for the chosen planes for several incident beam angles. By measuring the separation of the discs and of the bands, the thickness of the foil and the extinction distance of the chosen condition (a useful check) may be extracted by a simple graphical method as described in the references. Jotson and Kelly suggested an accuracy of approximately 2% for thickness measurement, but Ecob [14] has suggested a more realistic limit of approximately 5% under ideal conditions. For silicon, using 120kV and 200 or 220 type reflections, it was found possible to measure thicknesses from approximately 850Å to 2500Å, to within approxmately 7%; a simple computer program was used to process the data. The graphical method as described by Tanaka and Terauchi produced values for both thickness and extinction distance. The extinction distance was found to be a useful check of the accuracy of the measurements.

2.5.3 HREM IMAGING

The image formation mechanism under HREM conditions involves the construction of an image from the phase information contained in the diffracted beams, that is, an interference pattern. The wavelength of all the beams being essentially the same, this interference pattern is related to the periodicity of the diffracting medium, i.e. the crystal planes in the lattice. The phase information in each beam is defined by three conditions, the relative phases of the diffracted beam Bloch waves which exit the bottom of the specimen, the relative path lengths of these beams in the image formation process in the microscope and inelastic scattering events. The first condition is controlled by multiple diffraction events in the specimen which are strongly dependent on thickness. The second is controlled by the imaging conditions in the microscope, for example by the focus and the spherical abberation (C_s) of the objective lens. The effect of the objective focus on the relative phase shifts of the diffracted beams is defined for each machine by a series of empirical plots known as the contrast transfer function (CTF), in which relative phase is plotted against the reciprocal of spatial frequency (plane spacing) for a variety of defocus values. In accurate studies of crystal structure, it is neccessary to carry out careful image matching with computer image simulations. This means that even under conditions of abrupt change in structure it is not possible to measure heterogeneous structures by simply placing a ruler on the micrograph. However, certain aspects of the image are invariant, for example plane spacings and symmetry. The symmetry of the image must match the projected symmetry of the crystal. A good text on the subject of HREM has been written by Spence [15].

2.5.4 SECONDARY ION MASS SPECTROSCOPY

Secondary ion mass spectroscopy (SIMS) is a technique for measuring depth profiles of chemical distribution in a bulk material. The technique has excellent chemical resolution depending on the element being measured (10^{16}-10^{17}cm^{-3} for most dopants in silicon), and can have good depth resolution if correctly operated (approximately100Å). It lacks lateral resolution, typically having an analytical probe width of approximately 50μm.

A stream of positively charged ions, oxygen in this case, is produced in a beam from a suitable source. This beam is fired at the sample of interest and sputters a hole in the surface. The material removed from the sample is ejected in the form of ions which may be subsequently collected to determine the ion species, using a mass spectrometer. Thus a plot of the depth of the sputtered hole against the ion species produced gives a depth resolved chemical profile. Care must be taken that the secondary ions measured are only taken from the centre of the crater, to avoid measurement of sputtered and then redeposited material. A low energy ion beam is used (typically 2kV) which ensures reduced penetration of the primary ions into the sample (typically 20Å in silicon); this does however make the beam more difficult to focus which can reduce the lateral resolution. The accuracy with which the sputtering rate is known is the most critical component in depth resolution. For this reason, the depth of the hole is often measured after the experiment. In this work, SIMS is used exclusively to measure dopant profiles for boron, arsenic and phosphorous in silicon. Obviously, the measured dopant concentrations are total chemical concentrations and not electrically active concentrations.

2.6 CONCLUSIONS

In this chapter, a detailed description is given of sample preparation techniques for the TEM for both plan-view and cross-sectional foils. Modifications to the technique are described to allow the observation of buried interfacial layers by PVTEM. A process for etching TEM foils to perform dopant delineation in two dimensions is given, as is a technique for measuring the break up of thin interfacial oxide layers in device-like polysilicon contacted blanket wafers. A brief description is given of imaging modes in the TEM, and finally a description is given of the operation of the SIMS chemical analysis technique. In the next three chapters, these techniques are applied to both blanket samples, and real, fully fabricated device structures.

References

[1] . M.H. Loretto. Electron Beam Analysis of Materials. Chapman and Hall

[2] . Hirsch, Howie, Paschley, Nicholson and Whelan. Electron Microscopy of Thin Crystals. Krieger

[3] . H.R. Pettit and G.R. Booker. Proc. 25^{th} Anniv. Meeting, EMAG(IOP), p.290 1971.

[4] . J. Fletcher. D.Phil Thesis. Oxford 1979.

[5] . M.C. Roberts. D.Phil Thesis. Oxford. 1985.

[6] . M.C. Roberts, K.J. Yallup and G.R. Booker. Inst. Phys. Conf. Ser. 76(11), pp.483-488, 1985.

[7] . J.H. Wills. MSc. Thesis. Oxford. 1987.

[8] . K. Madl, A.L. Toth, A. Barna. Inst. Phys. Conf. Series, 93(2), pp.65-66, 1988.

[9] . N. Jorgensen. Private Communication.

[10] . A. Albu-Yaron, J.C. Barry and G.R. Booker, 8^{th} Euro. Conf. on Electron Microscopy, 1984.

[11] . J.W. Eddington, Monographs in Practical Electron Microscopy in Material Science, Phillips Technical Library, 1975.

[12] . M. Tanaka and M. Terauchi, Convergent Beam Electron Diffraction, publ. JEOL Ltd. 1985.

[13] . P.M. Kelly, A. Jotsons, R.G. Blake and J.G. Napier. Phys. Stat. Sol. 31 p.771, 1975.

[14] . R.C. Ecob. Scripta Met., 20, pp.1001-1006, 1986.

[15] . J.C.H. Spence, Experimental High Resolution Electron Microscopy, Oxford University Press, 1988.

Chapter 3

POLYSILICON/MONOSILICON INTERFACE BEHAVIOUR

3.1 INTRODUCTION

Polysilicon is now widely used in the construction of high speed bipolar transistors operating with f_{mosc}s of up to 22GHz (see chapter 1). The principle advantages of the polysilicon emitter process centre around the reduction in the length of the emitter base depletion region sidewalls [1] and the inherently self-aligning nature of the process which allows fewer mask steps and therefore a reduction in emitter window area [2] (see also chapter 1).

When the polysilicon is deposited on the single crystal silicon substrate, a thin layer of oxide is present on the wafer surface, this may be either a native oxide formed after cleaning, or a purposefully grown oxide [3]. This oxide forms a boundary between the polysilicon and the substrate which may have dramatic effects on the device characteristics [4](see chapter 1).

In commercial usage, the effects of the interfacial oxide on batch-to-batch reproducibility and more importantly on emitter series resistance in VLSI applications, have made it generally desirable to deliberately break up the oxide as a routine step in the fabrication of such VLSI high speed bipolar devices. This approach has resulted principally from a lack of detailed knowledge of the epitaxial regrowth behaviour of the polysilicon and the associated break-up and balling

up behaviour of the interfacial oxide and in particular how this behaviour may be modified by the presence of dopant species. In this study, an investigation has been carried out to examine the changes in interfacial morphology of P and As doped emitters during thermal processing. To place this work in the context of real device structures, all the processing steps associated with the fabrication of real devices were applied to these samples.

3.2 INTERFACIAL MORPHOLOGY

As processing proceeds following the deposition of the polysilicon layer, the structure of the interfacial region is observed to alter in two distinct ways, firstly oxide layer break-up and secondly epitaxial regrowth of the polysilicon layer, see for example Jorgensen et al [5]. Initially, the oxide begins to 'corrugate', i.e. the oxide interface becomes rough, thickening in certain regions at the expense of other regions . Eventually, the oxide perforates leaving gaps in the interfacial plane where the single crystal silicon (100) surface is now in contact with the polycrystalline material. As processing continues, the oxide continues to thicken in local regions, and the 'windows' which are free of oxide continue to grow in size. The oxide is finally transformed from a thin and approximately continous planar sheet into an array of spherical, dumb-bell and worm-like particles which lie in a plane close to the position of the original oxide layer.

As the oxide perforates, the single crystal substrate comes into direct contact with the polycrystalline material. This in effect represents a grain boundary in the interfacial plane. This boundary is often referred to as a 'pseudo grain boundary', although a more accurate intuitive model is obtained by referring to it as a single multi- or multiple grain boundary. It is well known that during thermal processing subsequent to the deposition of a layer of polycrystalline

material, that the larger grains grow at the expense of the smaller grains, in a process known as 'grain coarsening'. An analogous behaviour might be expected in the case of the single crystal/polycrystalline interface described above, and indeed exactly such a behavour is observed following perforation of the interfacial oxide. As subsequent heat treatment progresses, the single crystal regions grow upwards towards the surface by consuming the polysilicon[3] [6] [7]. The regrowth is driven by a reduction in the surface area resulting from the loss of grain boundaries. Behind this regrowth front the material is now crystallographically aligned to the substrate. This is of course a form of solid phase epitaxy, and the process is referred to as 'epitaxial regrowth'.

In polysilicon emitters, the nature of the interface has been shown to have a critical effect on device performance [8] [9] [10] [11]. It is thus vitally important to understand the effect of the processing variables on the 'balling up' behaviour of the oxide and on the epitaxial regrowth of the polysilicon layer. In this investigation two specific variables are considered, firstly the relative morphological variations associated with the presence of arsenic or phosphorus compared with undoped material and secondly how the effect of pre-anneals can modify the behaviour of the interfacial regions. Such pre-anneals are carried out to promote break up of the interfacial oxide without the need to increase the emitter drive-in time or temperature as this increases the depth of the emitter in the single crystal material to the detriment of device performance.

3.2.1 THE INTERFACIAL OXIDE

The 'native' oxides formed by HF cleans can vary signifcantly in thickness depending on the precise details of the cleaning process. Variations in thickness for HF oxide layers from 4Å to 8Å have been reported for example by Wolstenholme[3]. It is useful to consider how these in-situ thickness measurements have been obtained. Three different methods have been applied to in-situ measurement by various workers; all however involve electron microscopy. Firstly, cross-sectional HREM images can be obtained of the interface region, and these have been used to establish the onset of epitaxial regrowth. They also provide a tempting opportunity to measure oxide thickness directly. Amorphous materials show a fine 'worm-like' contrast under HREM imaging conditions, and structure of this type can be observed at the monocrystalline/polycrystalline boundary. This is clearly the 'interfacial layer' of 'native' oxide.

However, interpretation of HREM images can be complex, due to the nature of the image forming mechanism (i.e. by phase contrast) as discussed in chapter 2. Even under carefully optimised conditions, the difficulty of modelling the amorphous and polycrystalline regions is great.

The measurement of oxide thickness may be further complicated if the interface is not planar through the thickness of the foil (at least 50-200Å). This means that superposition of the oxide with both the monocrystalline and polycrystalline regions may occur to an extent that is not easily ascertained. Odhomari et al [12] have attempted to model the monosilicon/oxide interface. They suggest that it would be energetically favourable for the (100) surface to produce pyramid-like (111) facets. They were able to produce a simulated image quite close to their observations. They suggest alternatively that a sub-oxide (SiO_x) may exist between the (100) surface and the oxide.

An alternative approach to HREM as used by Jorgensen [13], involves balling up the oxide in a high temperature annealing stage until the oxide particles have

become relatively thick ($\geq$ 30Å). PVTEM diffraction contrast images of the layer are then obtained similar to those shown in chapter 3. By measuring the size of the balled up oxide particles and assuming that the particles have a rectangular cross-section a figure for the volume of oxide may be calculated. Assuming that significant amounts of oxygen are not lost from the oxide during the anneal and that the balled up oxide has the same density as the original oxide interfacial layer, it is then possible to deduce a figure for the thickness of a uniform film which would contain the same total volume of oxide. This is taken to be the original oxide layer thickness at the interface. In the case of Czochralski silicon where the concentration of oxygen in the wafer may be as high as $10^{18}cm^{-3}$, Ahu et al [14] have calculated that the total amount of oxygen contained in the film will not alter when the sample is annealed. However, there is no evidence to suggest that the density of the oxide remains constant. Additionally, observation has shown that the initial oxide layer is not of uniform thickness and the calculation of an ideal, uniform layer must be considered of limited usefulness, when modelling real devices. The results obtained by this technique do however give some idea of the thicknesses involved and fit reasonably with the thicknesses measured by HREM imaging. The third technique for measuring the interfacial layer thickness is the Fresnel fringe technique of Ross and Stobbs, this is discussed later in this section.

The term 'native oxide' is commonly used to describe the thin layer of oxide which forms on the surface of a 'clean' silicon wafer that is exposed to air. It is now neccessary to examine in more detail what is meant by the term 'native oxide'. In a thick oxide ($\geq$100Å) the chemical composition is generally taken to be approximately SiO_2, but as thinner layers are considered, a number of factors become important. A decision has to be made as to where the oxide begins and this becomes more critical for thinner layers. Observations of layers as thin as 4Å have been reported by Wolstenholme[3]. Clearly in such a case the interfacial layer consists substantially of approximately one monolayer or

less of oxygen and silicon atoms with some bonding to the silicon on either side of the interface. In a thin oxide of the type discussed here, it would not be surprising to observe a periodic arrangement of these oxygen atoms with the substrate. Ourmazd [15] has observed such an effect and has identified this as the tridymite phase of silica. Clearly, as this is less than one unit cell it cannot be adequately described by direct comparison with a bulk, equilibrium phase. The effect if real, is more accurately considered as being similar to sub-critical layer thickness epitaxy which can occur in thin layers [16].

Furthermore, experimental evidence has identified a potential contamination effect in these thin layers. Weinberger et al [17] have observed the presence of up to one monolayer of SiF or SiF_2 on the surface of silicon wafers which have recieved an HF clean, using X-ray photoelectron spectroscopy. They similarly detected carbon contamination and considered that it resulted from air-borne hydrocarbons which have a high sticking coefficient to the silicon surface. Similar carbon contamination has also been reported by Taubenblatt [18]. It is important to consider therefore that in these thin 'oxides', the percentage of oxygen in the layer may be reduced from its notional value by 50% or more.

Grovenor and Cerezo [19] have investigated the stoichiometry of thin native oxide layers using a Pulsed Laser Atom Probe (PLAP). This allows the surface of a sample to be investigated chemically, atom layer by atom layer. They observed that for oxides similar to an RCA clean (see below), the stoichiometry was close to SiO for layers up to 5Å thick and then changed abruptly to a composition close to SiO_2 whilst for native (HF clean see below) oxides they observed a continuous compositional change. They did not however observe significant contamination of the oxide with either carbon or fluorine.

Ross and Stobbs [20] have investigated the thickness and composition of thin oxide layers formed between monosilicon and polysilicon using TEM images of Fresnel fringes at the interface. This technique is extremely difficult to perform accurately. They observed that in all the oxides examined, a layer of sub-oxide

(SiO_x) less than approximately 5Å thick may exist. They also noted a difference between the observed amorphous layer thickness in HREM and their measured compositional oxide layer thickness in thicker oxides (approximately 40Å). They attributed this to the presence of a crystalline oxide phase below the amorphous layer. However they did not observe the presence of such a crystalline phase in thin oxides as investigated by Ourmazd.

Jorgensen[13] observed that modifications in the detailed processing conditions used in the HF clean process resulted in a change in the oxide thickness (as determined both by HREM and by oxide balling measurements). This type of variation illustrates the potential for the production of different interfacial structures by different workers and may well be one of the reasons for the confusion in the literature over the exact nature of the interfacial region and the differing electrical properties obtained by different workers on nominally similar samples.

3.3 SAMPLE PROCESSING

The present work was carried out on 3 inch diameter (001) n-type Czochralski silicon wafers. The wafers were cleaned by one of two processes. The first is the standard process used commercially to produce a 'clean' wafer surface. This consists of a dip in buffered HF followed by a rinse in distilled water. It has been shown however that this does not in fact leave a free silicon surface and that a layer of 'native' oxide ≈4-8Å thick results at the end of the clean.

The oxide which results from the HF clean is non-uniform as has been previously reported (by Wolstenholme[3] and also by Albu-Yaron [21]). To enable the effects of the dopant/heat treatments to be effectively separated, a second series of samples was generated. These replaced the HF oxide with a

	HF	RCA
HF dip etch	1-6,13	
HF dip etch+RCA clean		7-12,14
Polysilicon Deposition	1-6.13	7-12,14
Pre-anneal (10min 1000C Dry N2)	2,4,6	8,10,12
As implant (1E16 cm-2, 70keV)	3,4	9.10
P implant (1E16 cm-2, 50keV)	5,6	11,12
Drive-in (30` 900C, wet 02)	1-6	7-12
pre-anneal (10 min 1000C dry N2)	13A	14A
pre-anneal (10 min 1100C dry N2)	13B	14B
pre-anneal (10 min 1150C dry N2)	13C	14C

Table 3.1

Specimen No.	Oxide	Regrowth	Polysilicon layer
2,HF, no dopant pre-anneal+drive-in	broken	yes, 60x110A	grain size 0.1um
8,RCA, no dopant pre-anneal+drive-in	yes(17A) interfacial contrast	yes, 40x170A	grain size 0.1um
3,HF, As, drive-in	continuous (10A), rough	no	grain size 0.1um
9,RCA, As,drive-in	continuous (10-14A)	no	grain size 0.1um
4,HF, pre-anneal,As + drive-in	broken	yes,130-300A	grain size 0.1um
10,RCA, pre-anneal, As+drive-in	broken(23A)	yes(70-200A)	grain size 0.1um
5,HF,P+drive-in	broken	yes(40-190A)	grain size 0.15um
11,RCA,P+drive-in	slightly broken	yes(70-120A) 1000A apart	grain size 0.15um
6,HF,pre-anneal,P, +drive-in	balled-up	yes (1700-5000A)	
12,RCA,pre-anneal, P+drive-in	broken	yes(60-130A)	grain size 0.2um
13A,HF, 1000C pre-anneal	broken	yes(70-175A) 100A apart	grain size 0.1um
14A,RCA,1000C pre-anneal	broken(12A)	yes(50-150A)	
13B,HF,1100C pre-anneal	balled-up(57A)	regrown	regrown,twinned
14B,RCA,1100C, pre-anneal	balled-up(67A)	extensive	large areas of regrowth,twins
13C,HF,1150C pre-anneal	balled-up(78A)	regrown	few twins and dislocations
14C,RCA,1150C, pre-anneal	balled-up(90A)	regrown	twinned near surface, defect free near interface

Table of TEM results

Table 3.2

thicker, deliberately grown oxide. In this work an oxide $\approx$14Å thick was used, and this was obtained by a standard chemical surface treatment known as an RCA clean [22]. This consists of a buffered HF clean followed by the RCA clean itself (NH_4OH:H_2O_2:H_2O) in the ratio of 1:1:5 for 10 minutes, followed by HCl:H_2O_2:H_2O at 1:1:5 for 10 minutes).

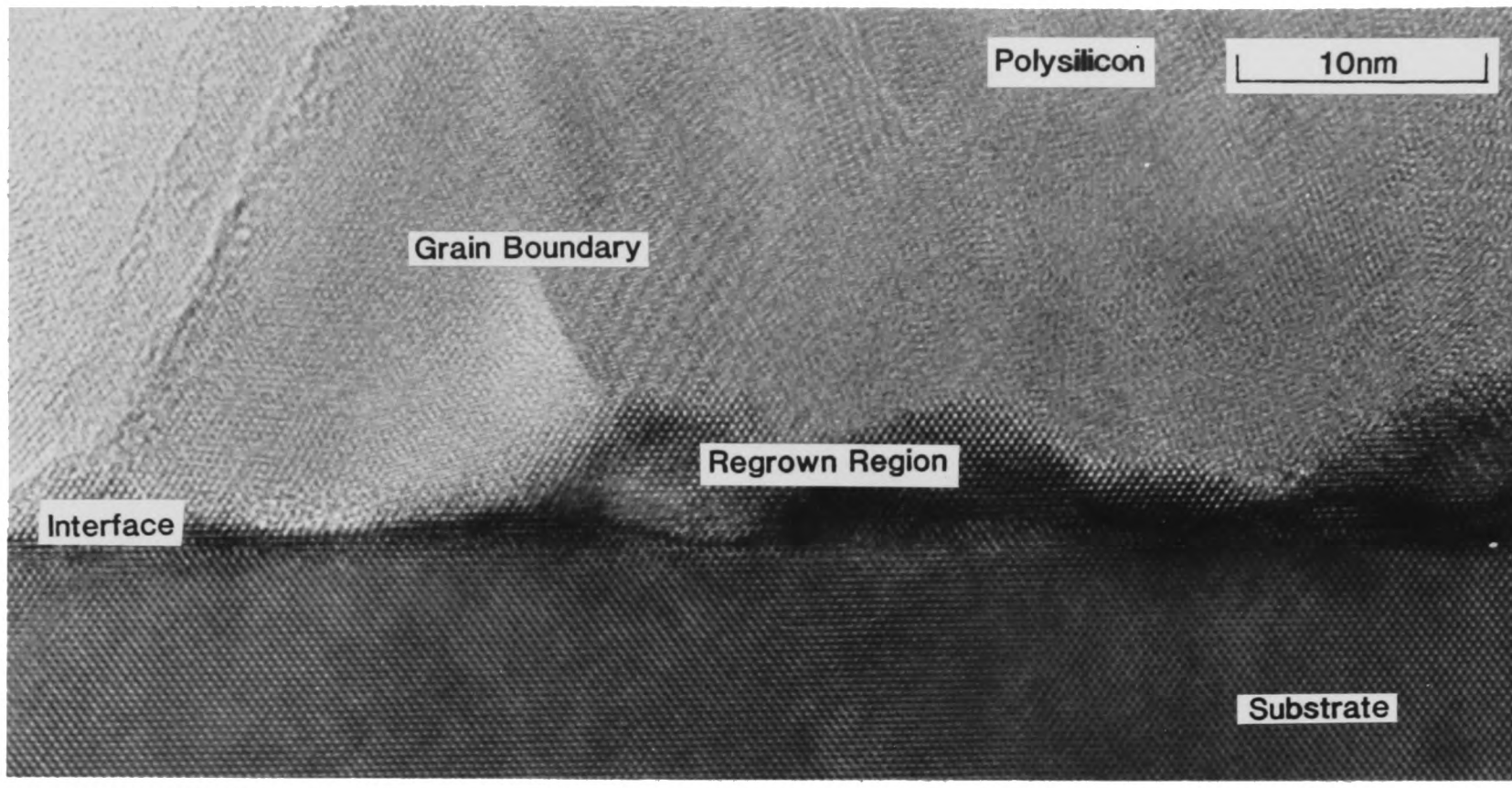
A layer of undoped LPCVD polycrystalline silicon $\approx$0.4μm was then deposited on all the samples in a commercial reactor at 610 °C. The samples then received a combination of different processing steps, as detailed in table 3.1. These allowed an investigation of the effects of different combinations of wafer cleans, dopants and thermal treatments.

3.4 TEM OBSERVATIONS

3.4.1 CROSS-SECTIONAL TEM

Unless otherwise stated, the results listed below all correspond to (110) type cross-sections. The results are discussed in pairs, firstly the HF cleaned sample, and then the RCA cleaned sample which has had identical processing. The samples are referred to by their assigned number in table 3.1, and the observed structures are characterised in table 3.2. The HREM images in this chapter were obtained from a JEOL 4000EX TEM operating at either 300kV or 400kV.

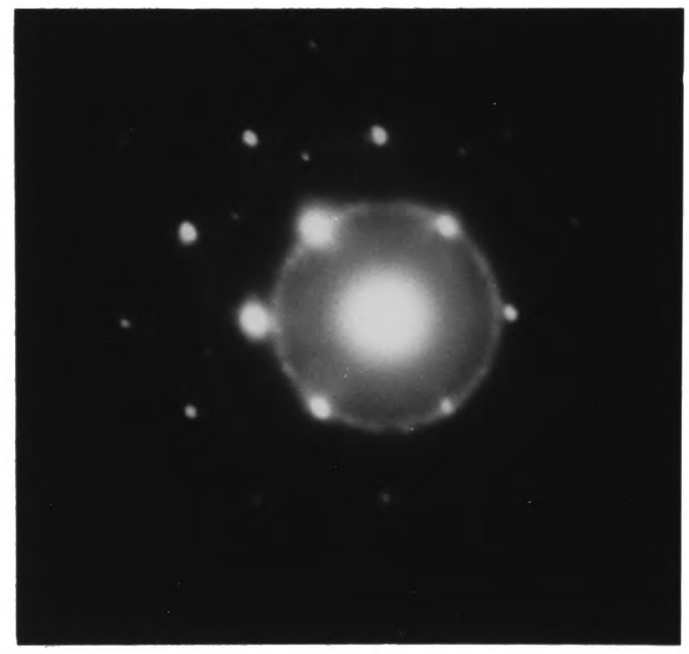
Sample 2 illustrates the effect of the thermal treatments (pre-anneal and emitter drive-in) only on the HF clean sample, i.e. no dopant implant. The sample consists of a layer of polysilicon approximately 0.3μm thick, having a random, equi-axed grain structure with an average grain size of approximately 0.1μm. The interface in this sample is broken up in many places by small areas

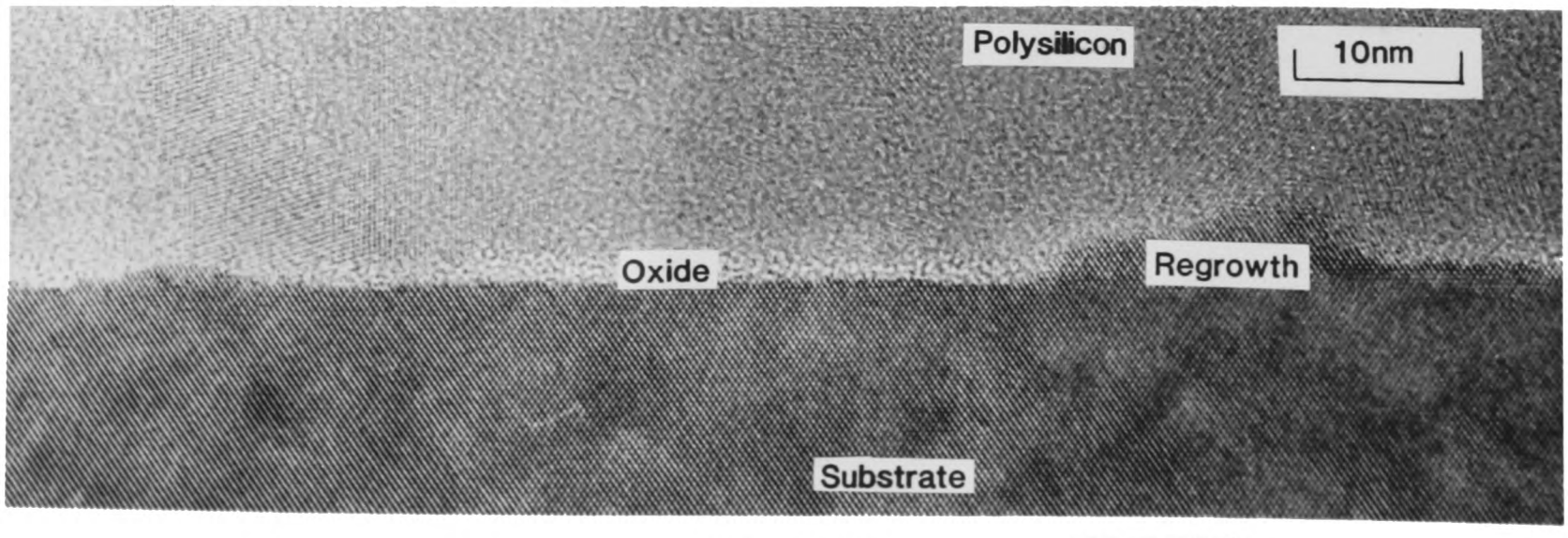


Sample 2 Fig.3.1a XTEM

Fig.3.1b

Diffraction Pattern of Polysilicon





Sample 8 Fig.3.2 XTEM

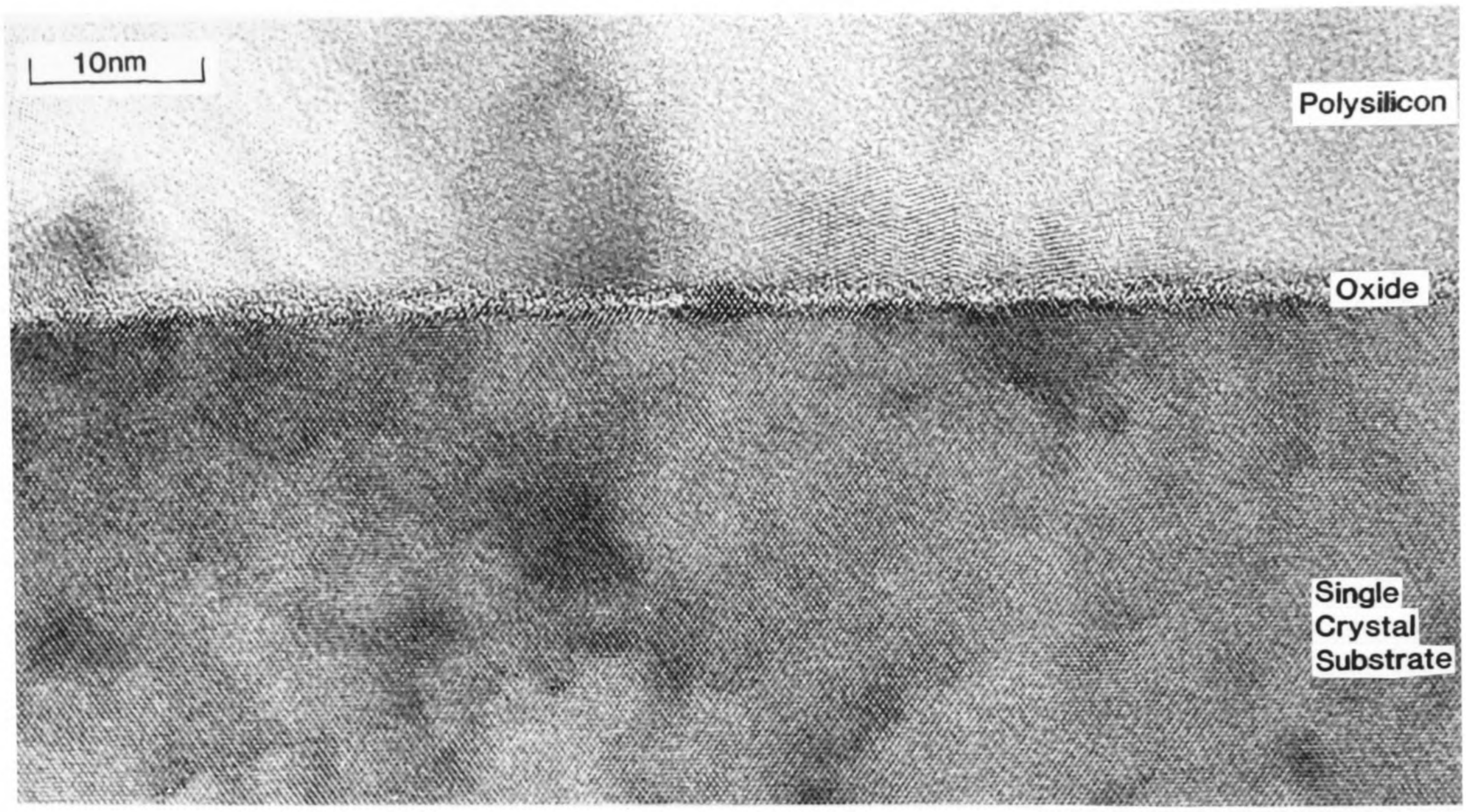


Fig.3.3a
Sample 3
XTEM

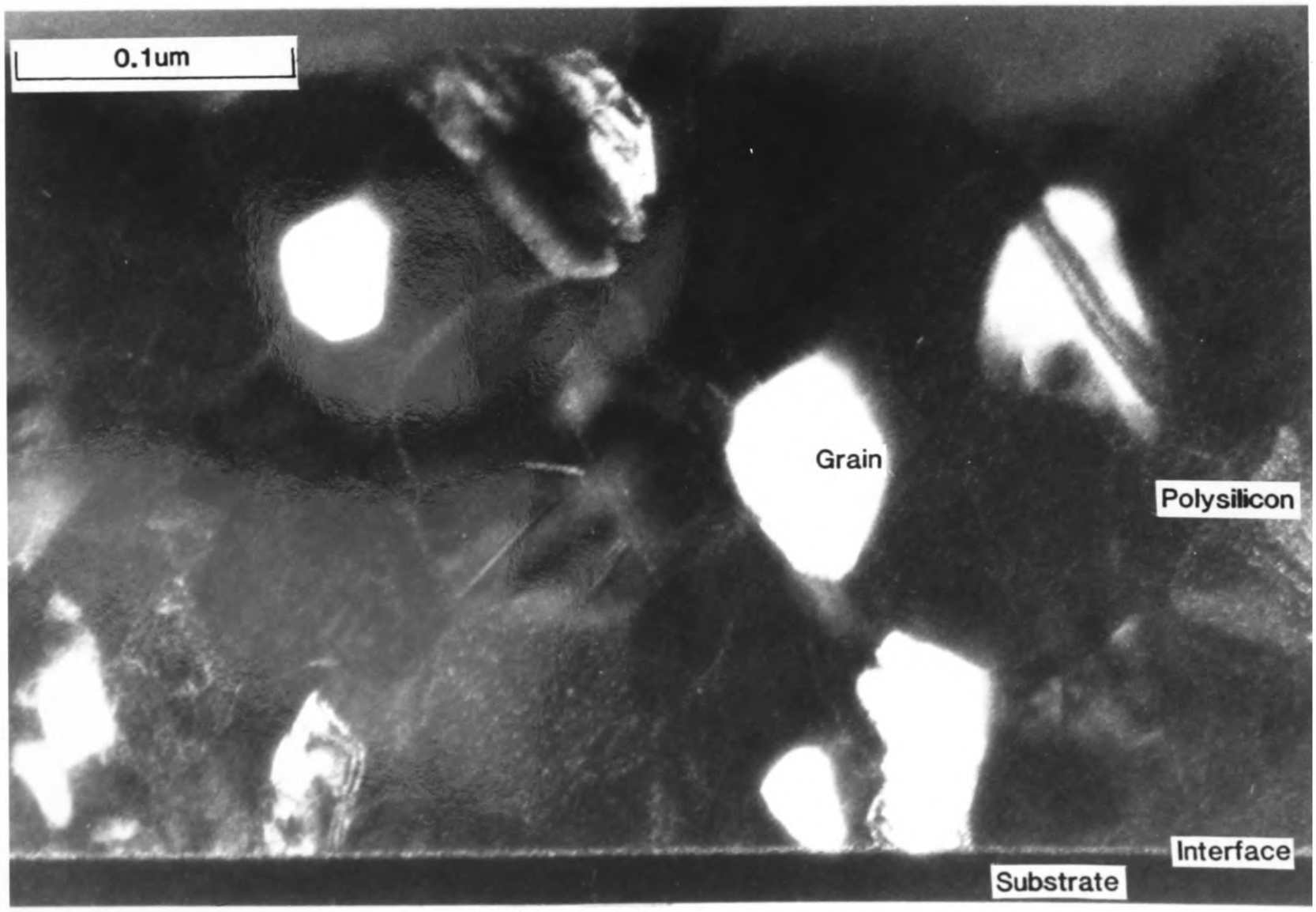


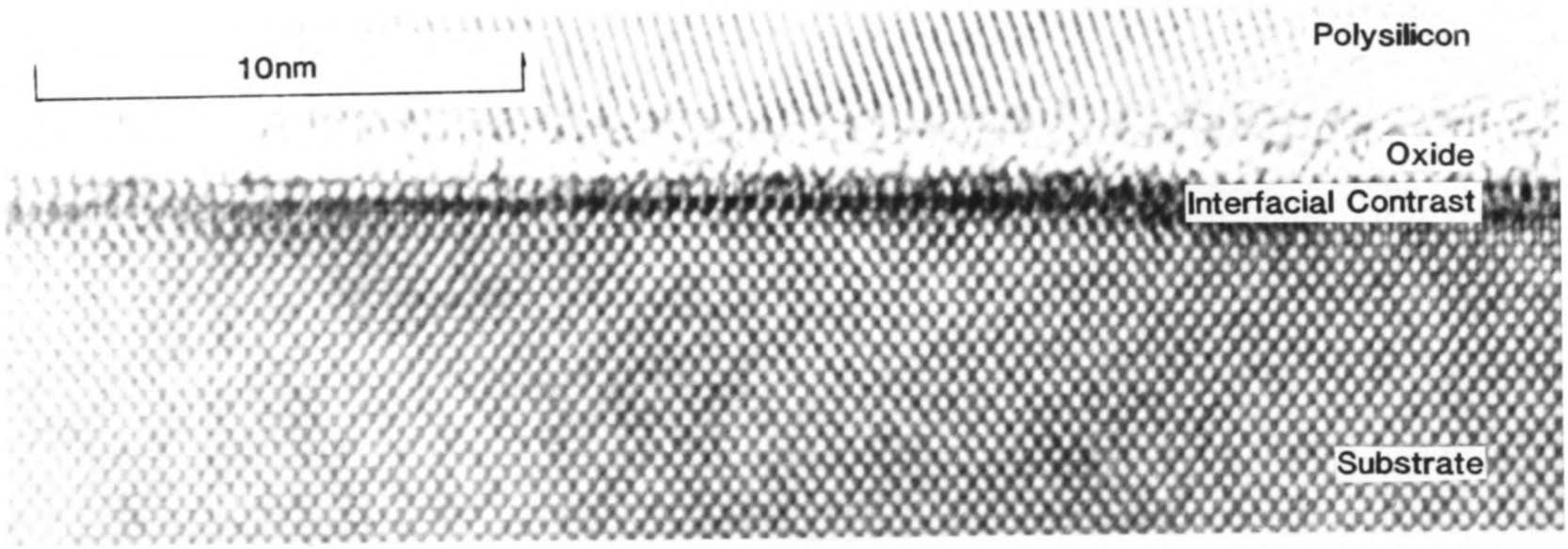
Fig.3.3b
Sample 3
XTEM

of regrowth (fig.3.1a). These areas are typically 60Å×110Å(perpendicular and paralell to the interface respectively). The regions of regrowth are separated by less than 100Å. The regrowth appears to proceed up grain boundaries. The presence of defects in the substrate was observed to be caused by beam damage in the 400kV microscope used for this study. To avoid this problem most of the subsequent samples were examined using an accelerating voltage of 300kV, which reduced the severity of the damage but did not completely eliminate it in all cases. Fig.3.1b shows a typical diffraction pattern from the interface region of a single crystal substrate with a surface layer of polysilicon. The strong spots arise from diffraction in the single crystal (the (110) pole in this case), whilst the annular specular pattern arises from the random orientation of a large number of individual grains in the surface layer.

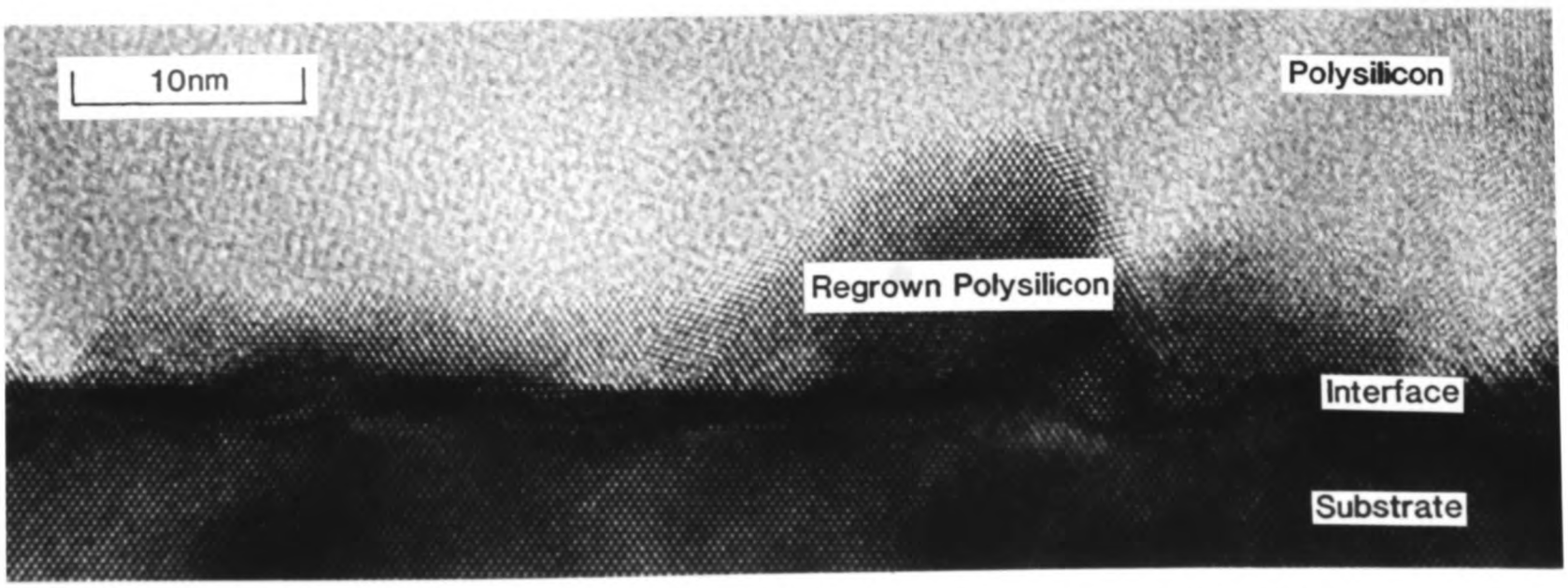
Sample 8 is the equivalent RCA cleaned sample. Regrown regions can be observed in this sample of approximately 40Å×170Å (fig.3.2). The regions of regrowth although of similar size to those in sample 2, are far more widely separated. Separations of up to 1200Å were observed. In these regions the interface has become rough, with the regrowth having pierced the oxide film but not having regrown into the polysilicon. In regions with no local surface roughening, the oxide film is approximately 10-12Å thick. Near to regions of regrowth, the film has thickened to approximately 17Å. In some regions of the sample, a discontinuous change in contrast was observed in the substrate, immediately below the oxide layers, but was never observed at a regrowth front. This discontinuous change in contrast was measured to be approximately two monolayers or about 10Å.

Sample 3 has had an HF clean, an arsenic implant and a drive-in, but no preanneal. This sample has a polysilicon layer approximately 0.3μm thick. Grain size is approximately 0.1μm (fig.3.3b). The oxide layer is very nearly continuous and approximately 10Å thick (fig3.3a). Roughening of the substrate has occurred

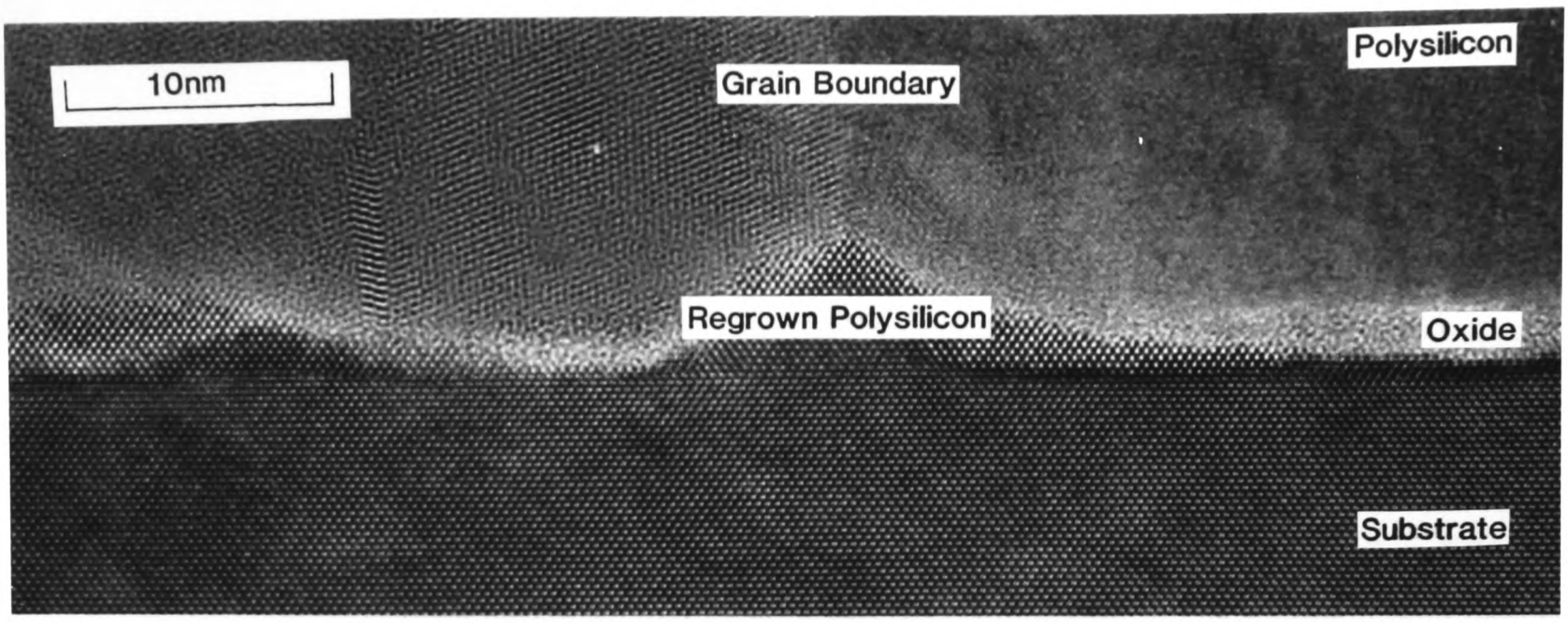
Sample 9 Fig.3.4a XTEM



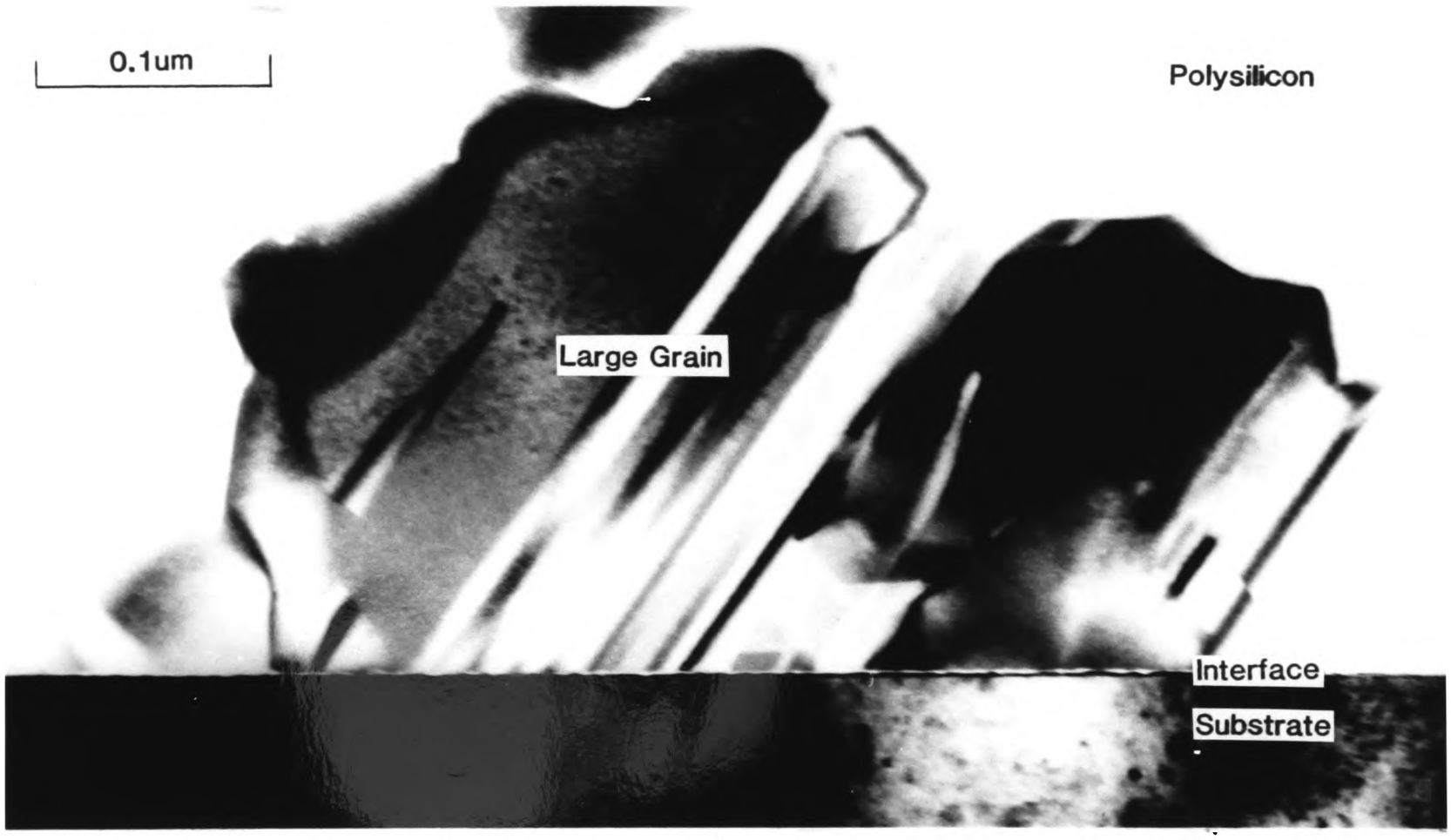
Sample9 Fig.3.4b XTEM



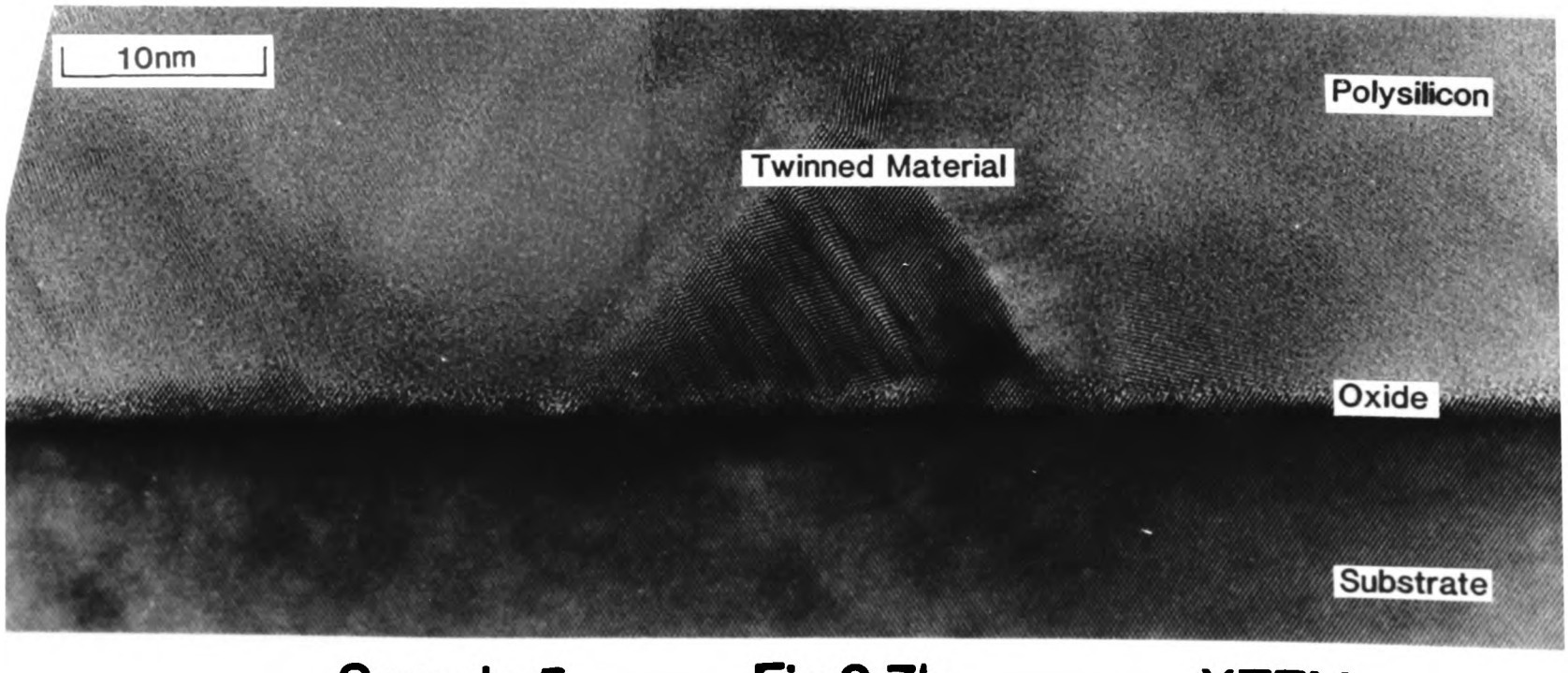
Sample 4 Fig.3.5 XTEM



Sample 10 Fig.3.6 XTEM



Sample 5 Fig.3.7a XTEM



Sample 5 Fig.3.7b XTEM

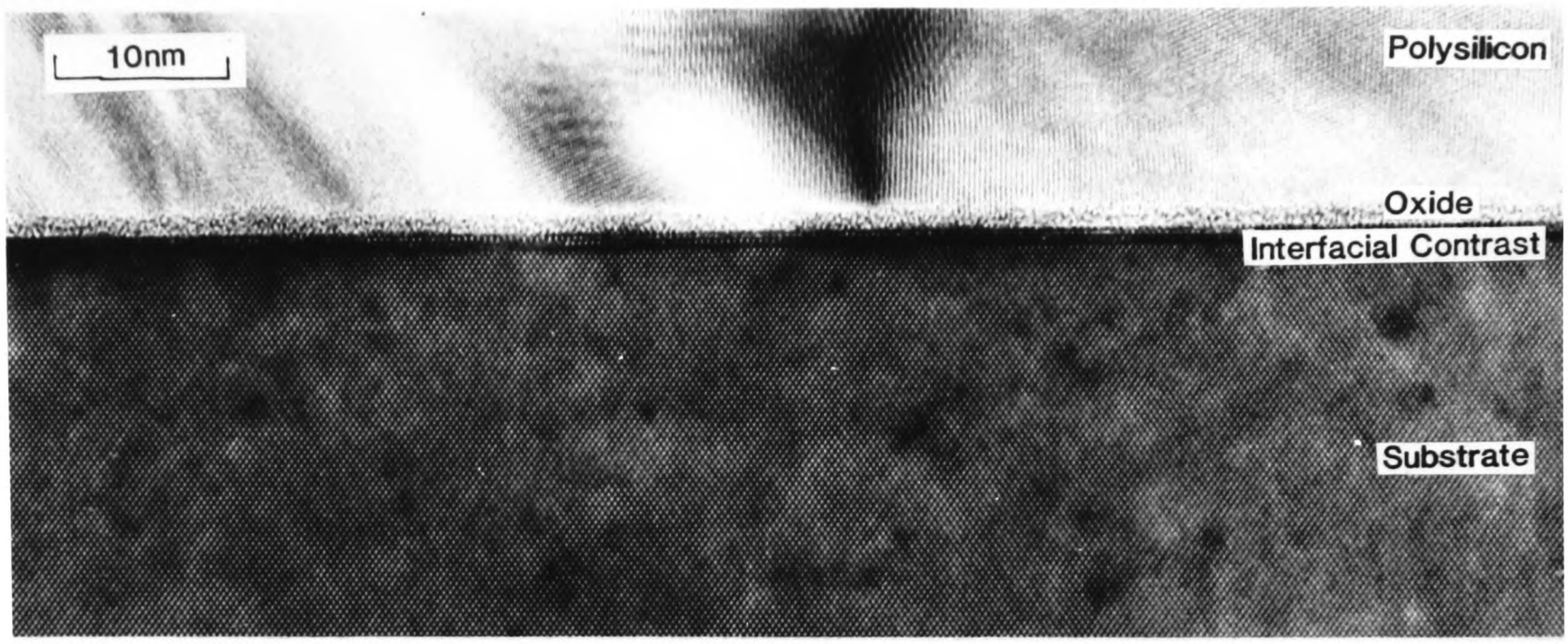
in areas separated by 100-175Å, but no regrowth into the polysilicon layer has occurred.

Sample 9 is the equivalent RCA cleaned device. The polysilicon layer is approximately 0.3μm thick with a mean grain size of approximately 0.1μm (fig.3.4a). The oxide layer is continuous across the observed sample area and has a thickness of approximately 10-14Å. Once again, a two monolayer region of discontinuous contrast was observed in the substrate (fig3.4b). In this case, this contrast was observed over most of the sample examined. No regrowth or significant thinning of the oxide can be observed.

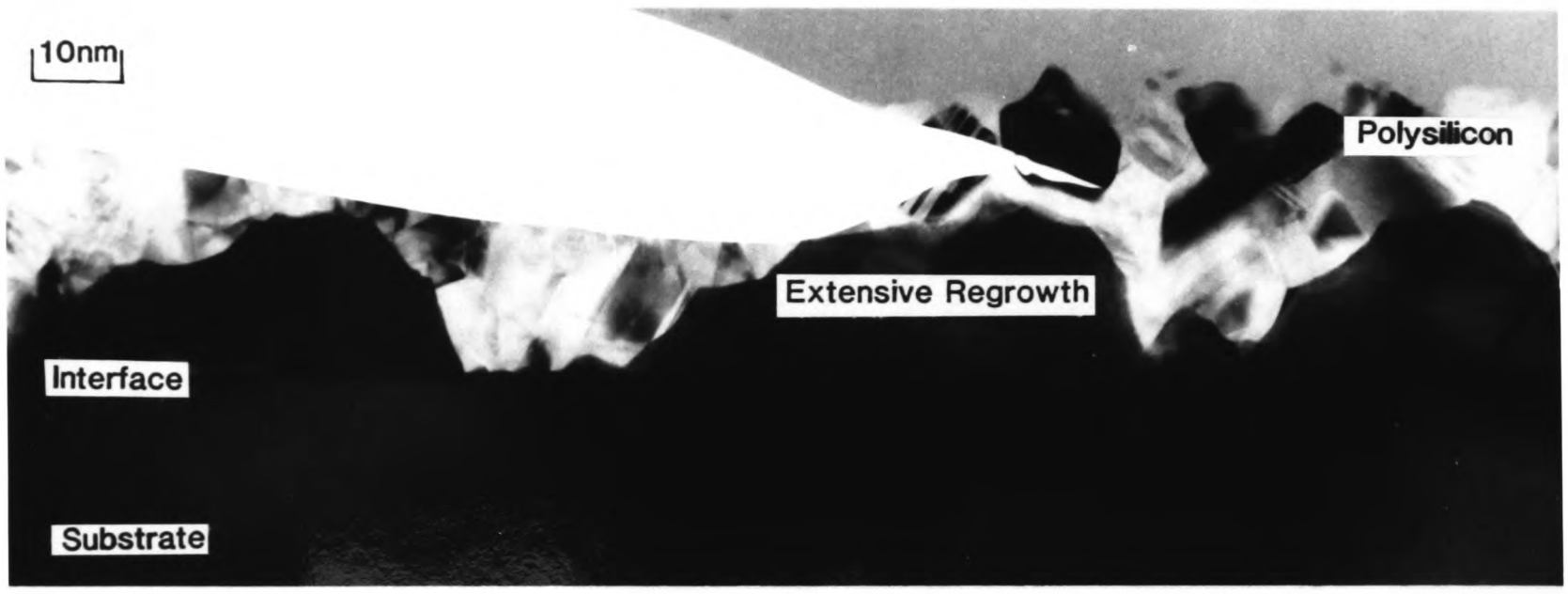
Sample 4 has had an HF clean, pre-anneal, As implant and drive-in. The entire interface is now covered by regrowth, having maximum dimensions of 130Å×300Å (fig.3.5). The two monolayer region of discontinuous contrast is not observed in this case.

Sample 10 is the equivalent RCA cleaned sample. The polysilicon layer is approximately 0.3μm thick and contains equiaxed grains with an average size of approximately 0.1μm. HREM examination shows regions of regrowth of approximately 70Å×200Å. These regions can be clearly observed to regrow up grain boundaries (fig.3.6). The amorphous layer has formed 'pools' between the regrowth regions having a maximum thickness of approximately 23Å. The regrowth regions are separated by between 100Å and 250Å. The two monolayer region of discontinuous contrast can be observed in some regions of this sample, it is not present below regrown areas.

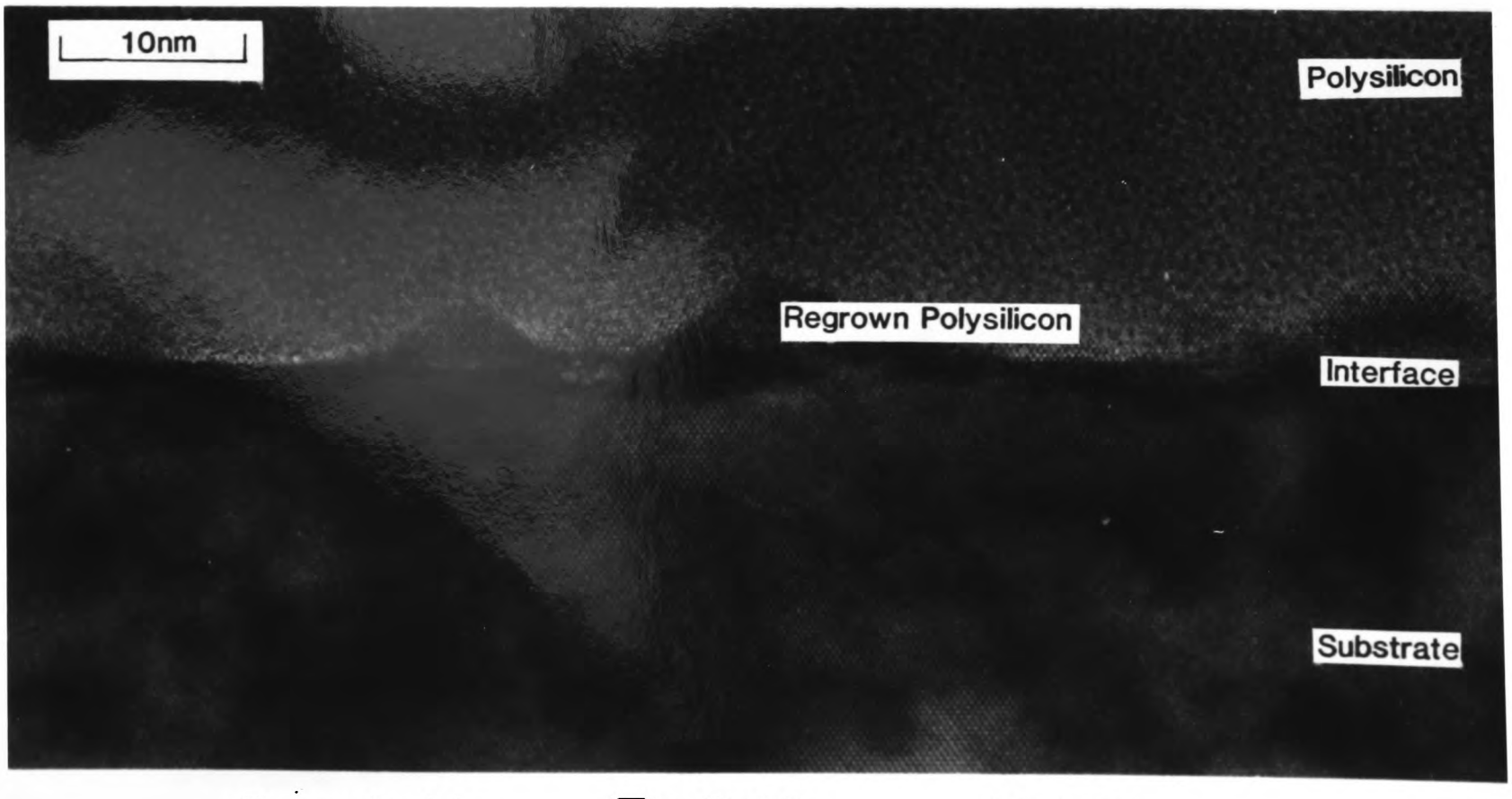
Sample 5 has received an HF clean, no pre-anneal, a P implant and a drive-in. Average grain size is approximately 0.15μm (fig.3.7a), with one observed grain having a lateral dimension of approximately 0.3μm. This particular grain contained largely defect free material with a few twins Such a low defect structure is often typical of grain growth. Small areas of regrowth at the interface can be observed approximately 40Å×190Å, with one region of 140Å×340Å (fig.3.7b). This region is illustrated as it clearly shows the initiation of microtwins in the



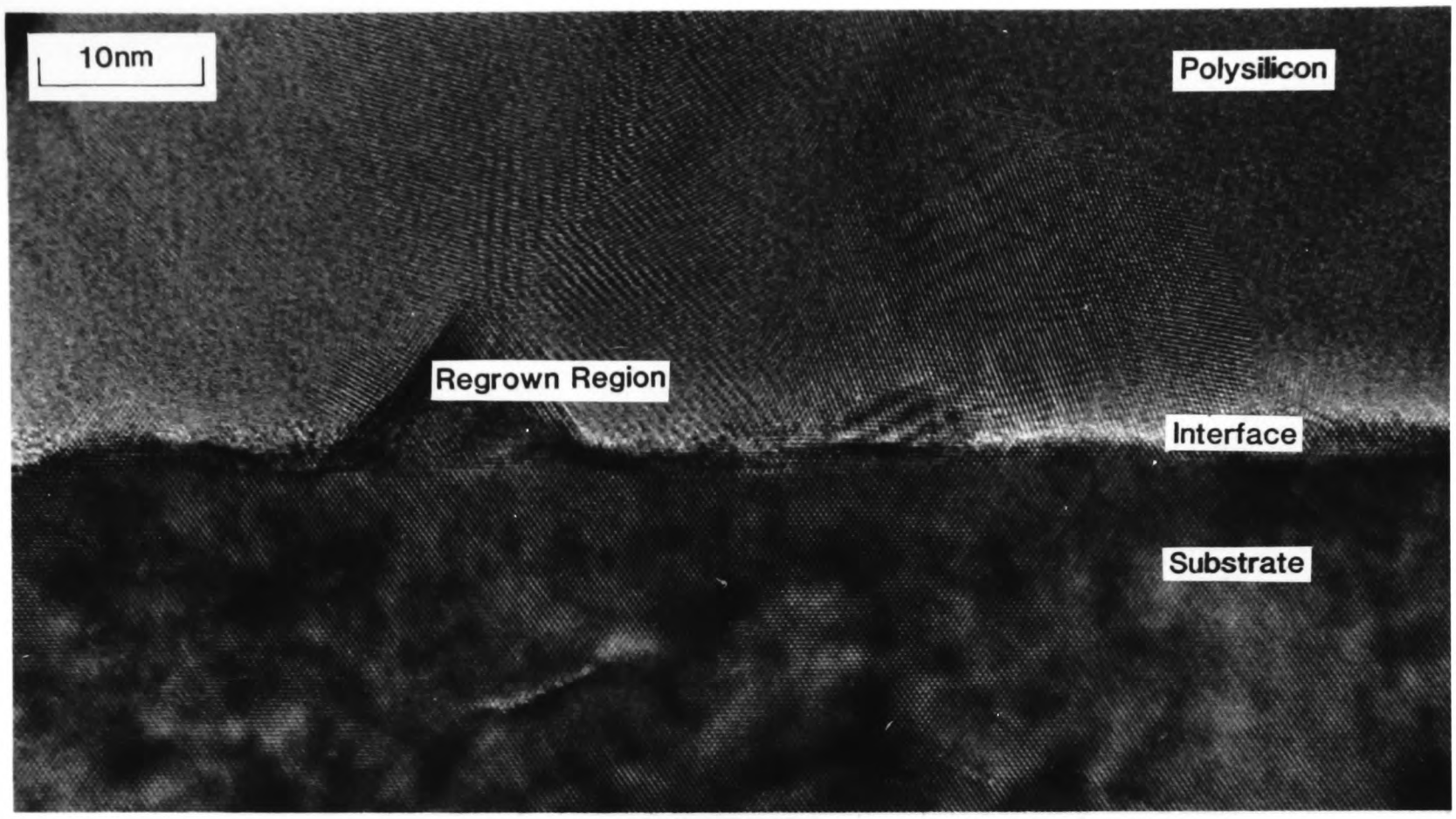
Sample 11 Fig.3.8 XTEM



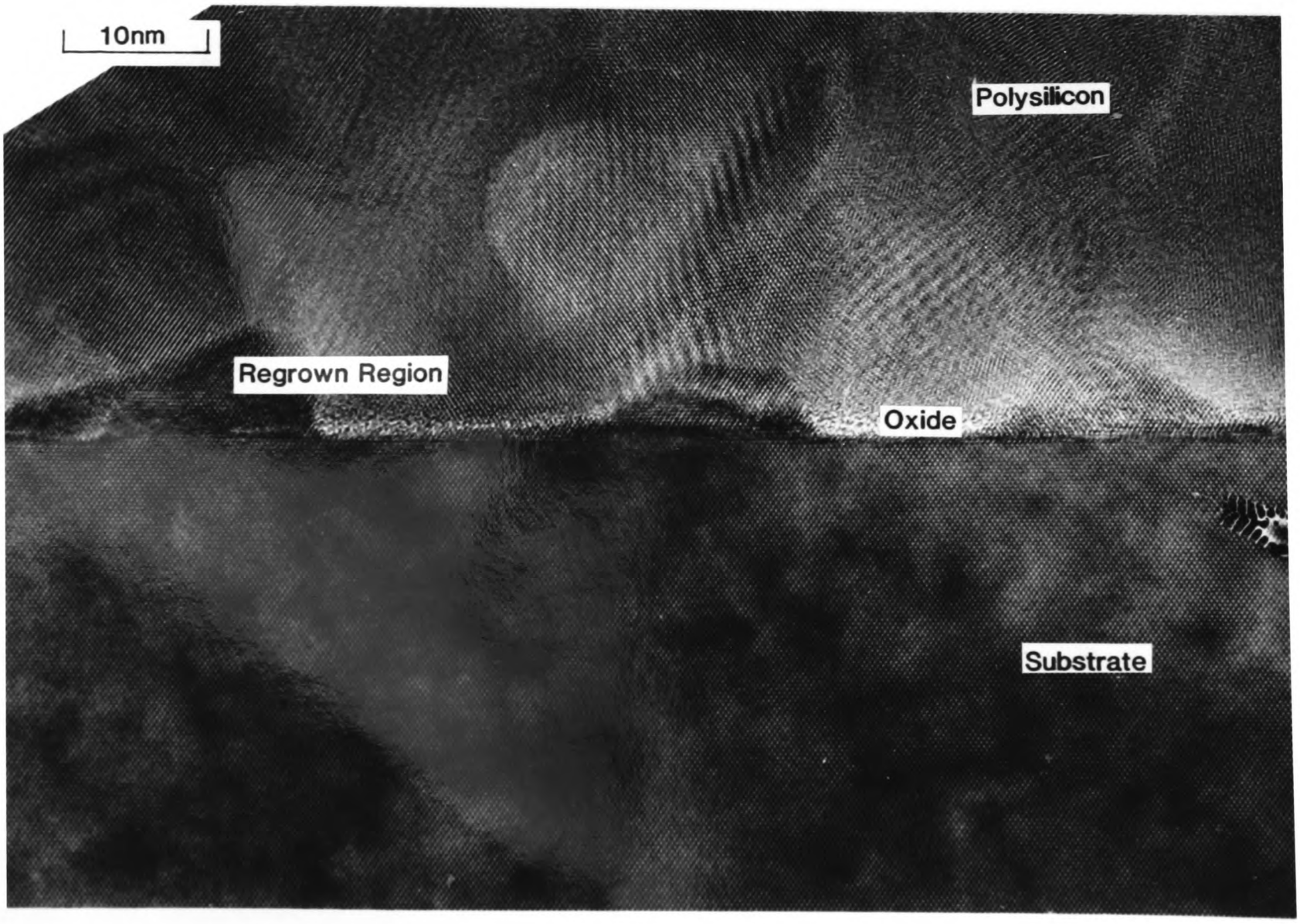
Sample 6 Fig.3.9 XTEM



Sample12 Fig.3.10 XTEM



Sample 13A Fig.3.11 XTEM



Sample 14A Fig.3.12 XTEM

regrowth. Once again regrowth can be observed to proceed up grain boundaries. Although the interface has begun to corrugate in many places, the regions of actual regrowth into the polysilicon are separated by up to 1000Å. The abrupt, two monolayer interface contrast is observed in some cases.

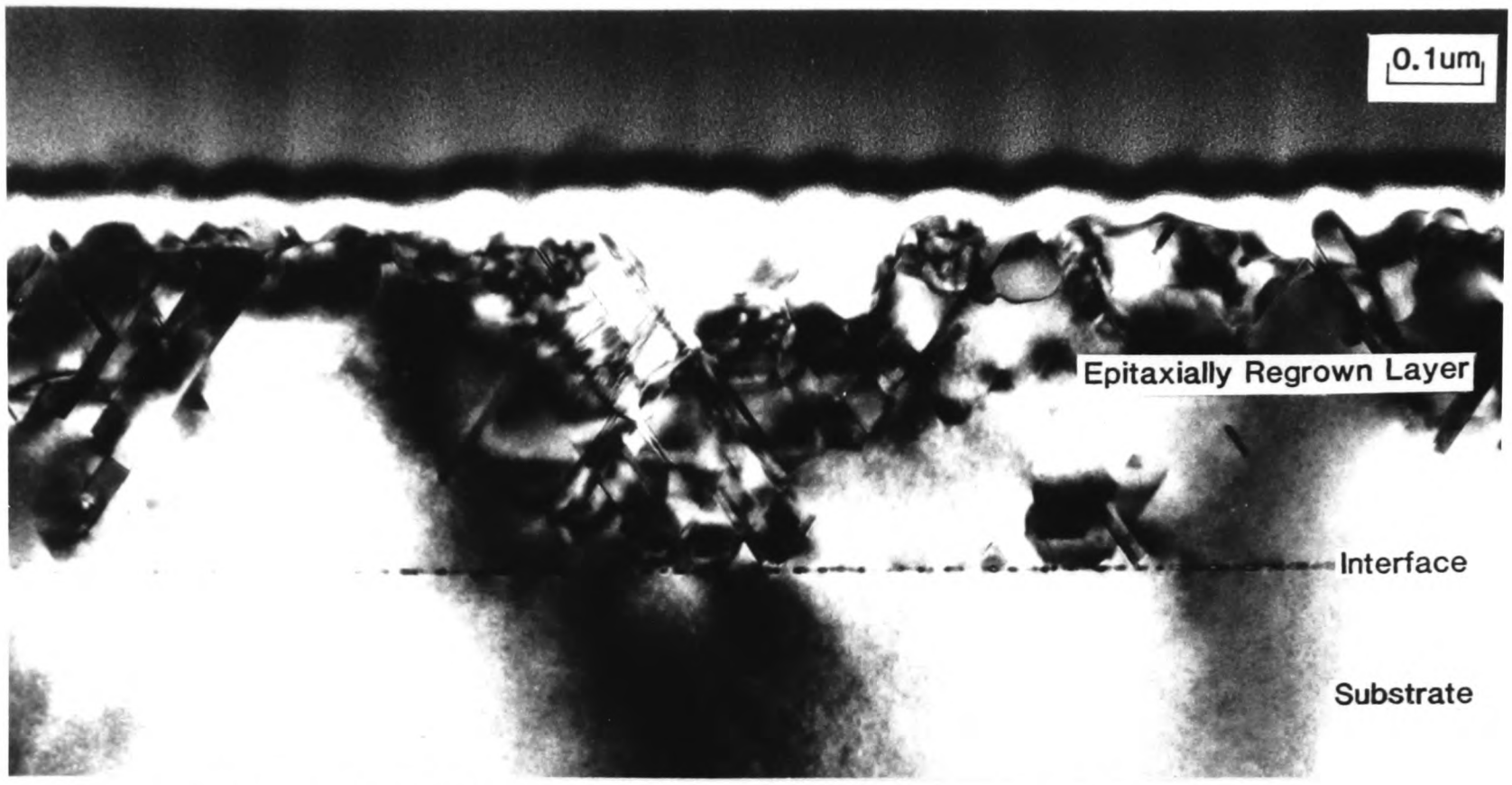
Sample 11 is the equivalent RCA cleaned sample. This sample exhibits an approximately uniform, continuous oxide layer 14Å thick. The substrate beneath the interface exhibits the two monolayer abrupt contrast over large regions (Fig.3.8).

Sample 6 has had the HF clean, pre-anneal, P implant and drive-in. The polysilicon layer contains regions of regrowth of approximately $0.17\mu m \times 0.5\mu m$ (fig.3.9); the white mark on the micrograph was caused by handling damage to the micrograph negative. The regrown layers contain a low density of twinned material. It is interesting to note that unlike some of the other samples examined, and despite the size of the regrown regions, the whole interface is not covered by regrown material. Once again, regrowth is observed to proceed up grain boundaries.

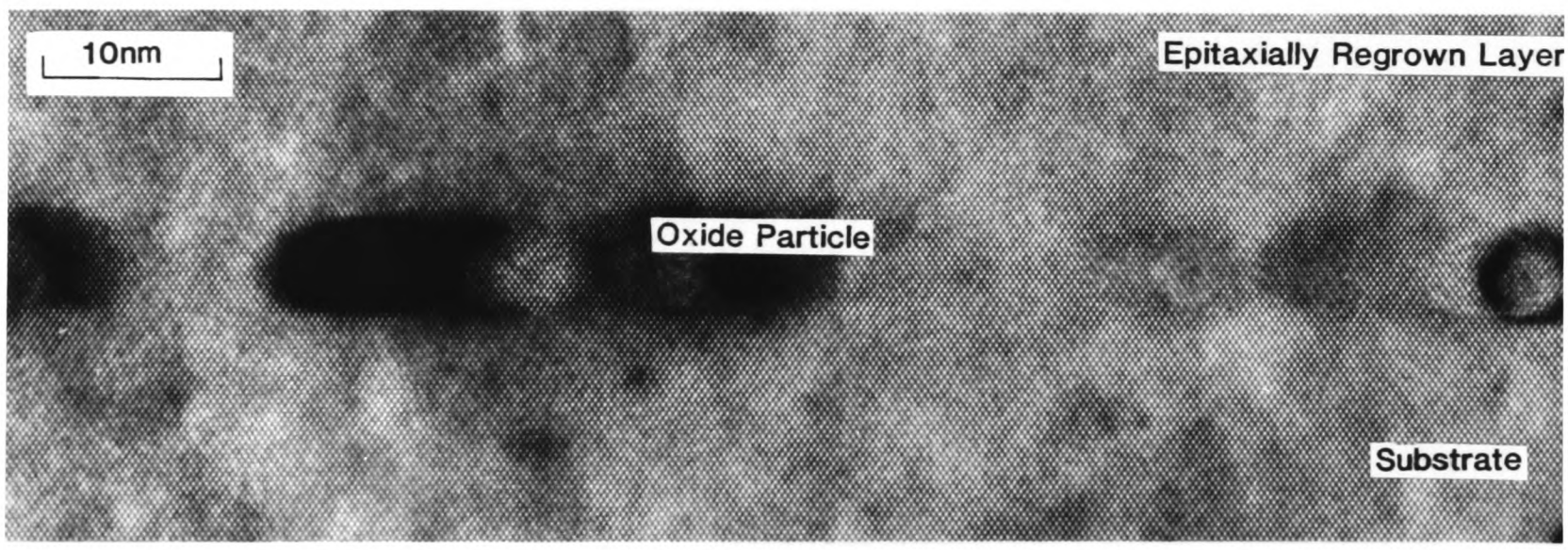
Sample 12 is the equivalent RCA sample. The polysilicon layer contains grains up to $0.2\mu m$ in size. Early stages of epitaxial regrowth are evident at the interface (fig.3.10), with dimensions of approximately 60Å×130Å. The amorphous layer has again formed pools at the periphery of the regrown regions.

A simple study was also carried out on the effect of varying the pre-anneal temperature. Three temperatures were used, 1000°C, 1100°C and 1150°. Pre-annealing was carried out on both HF and RCA cleaned samples, to generate six samples as detailed in the processing table (table 3.1). These results have particular relevance to chapter 4.

Sample 13A (HF, 1000°C pre-anneal) contains many small areas of regrowth approximately 70Å×175Å. Regrowth up grain boundaries can be clearly observed. The separation of regrown regions is typically 100Å (Fig. 3.11). The regrown regions appear to contain large numbers of microtwins.

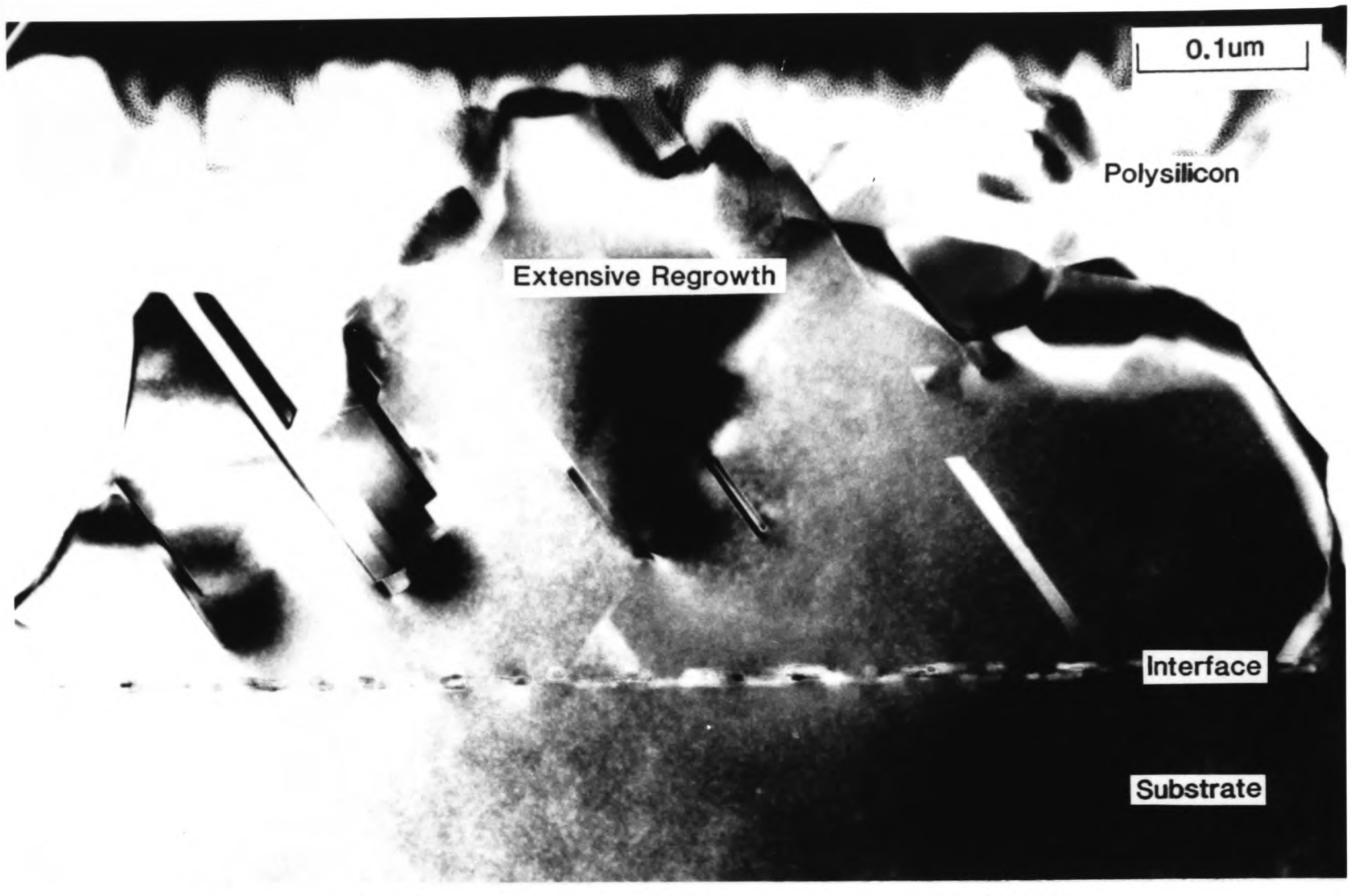


Sample13B Fig.3.13a XTEM

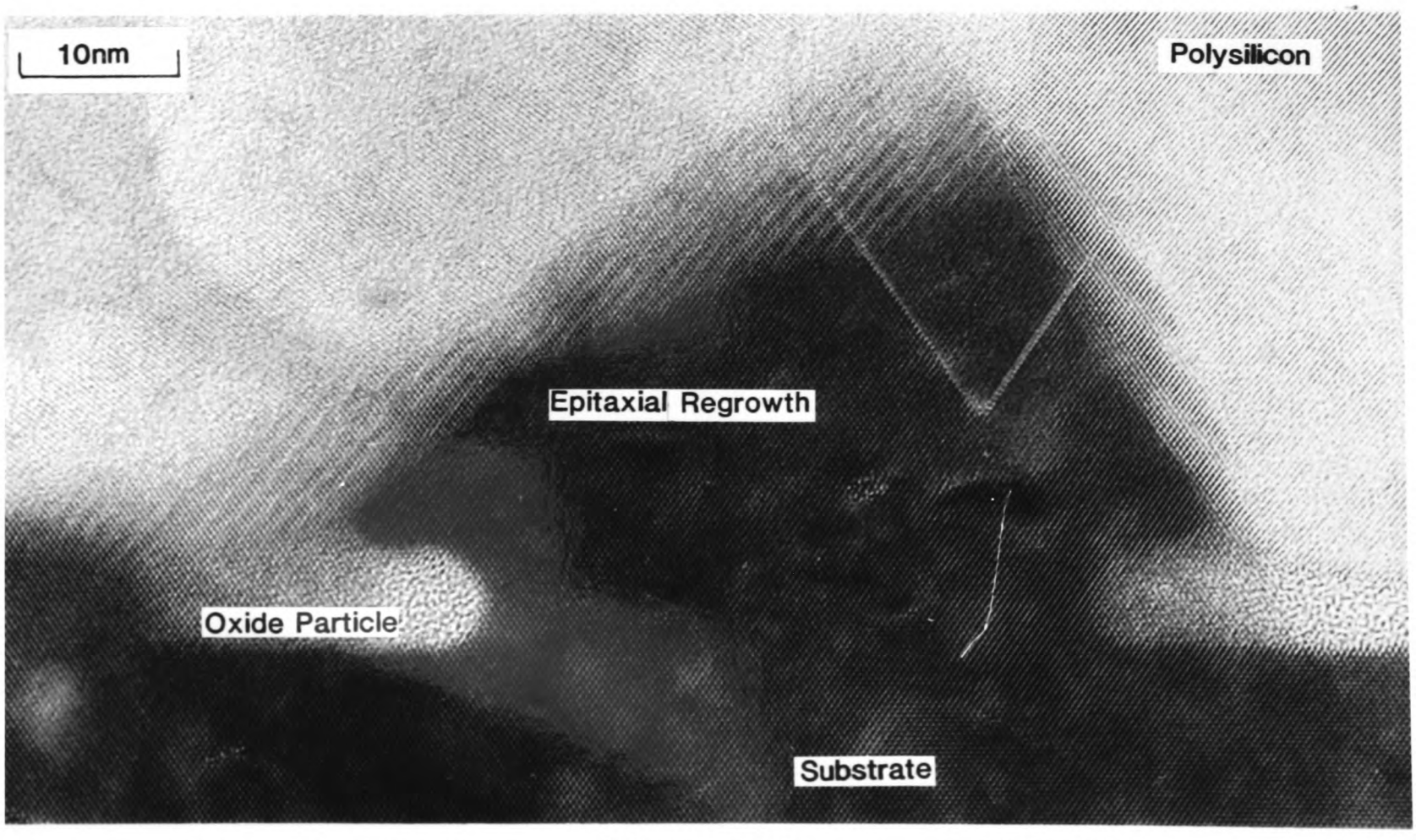


Sample13B Fig.3.13b XTEM

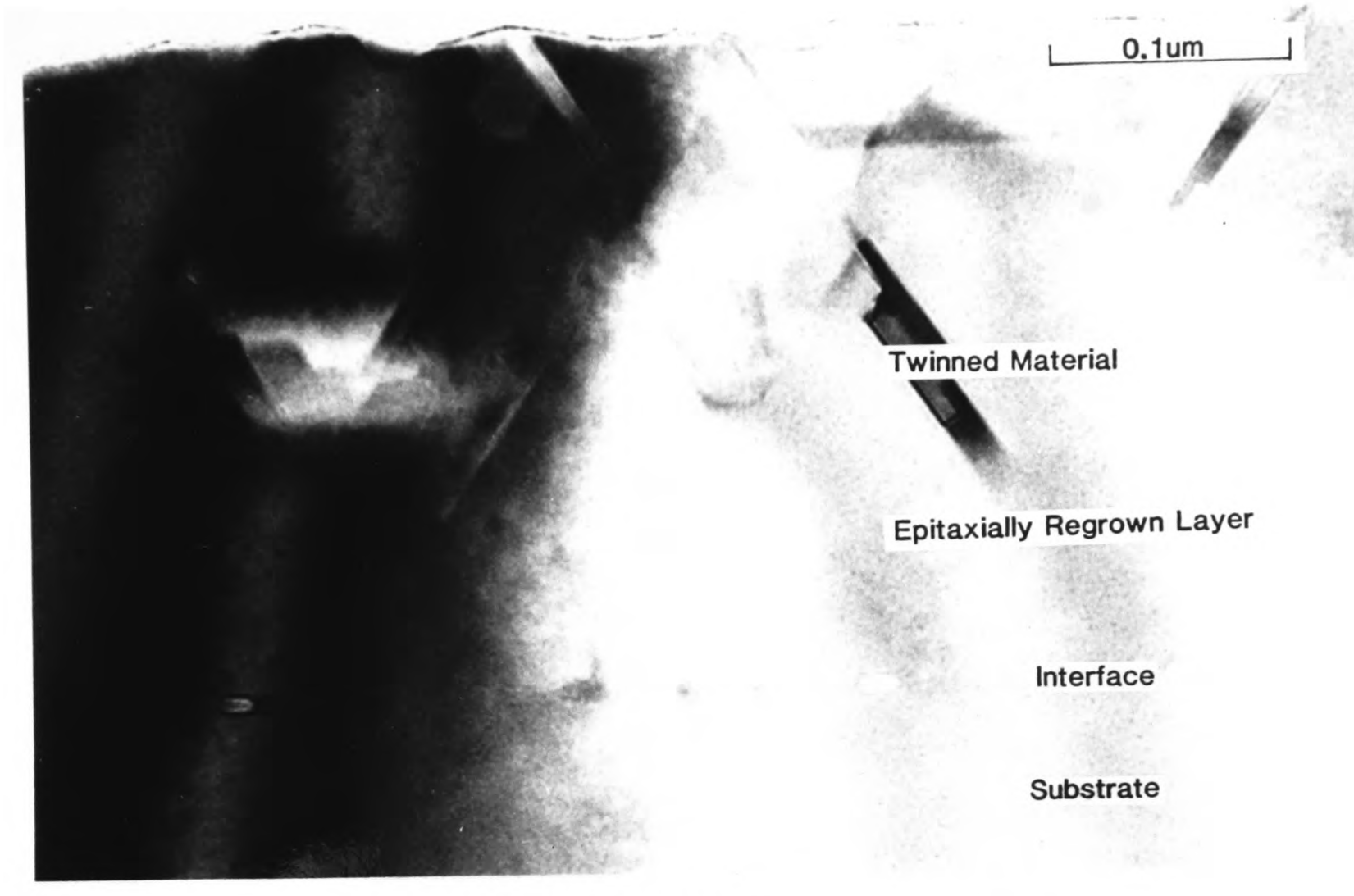
Fig.3.13c
Sample13B
Diffraction Pattern
of Twinned Material



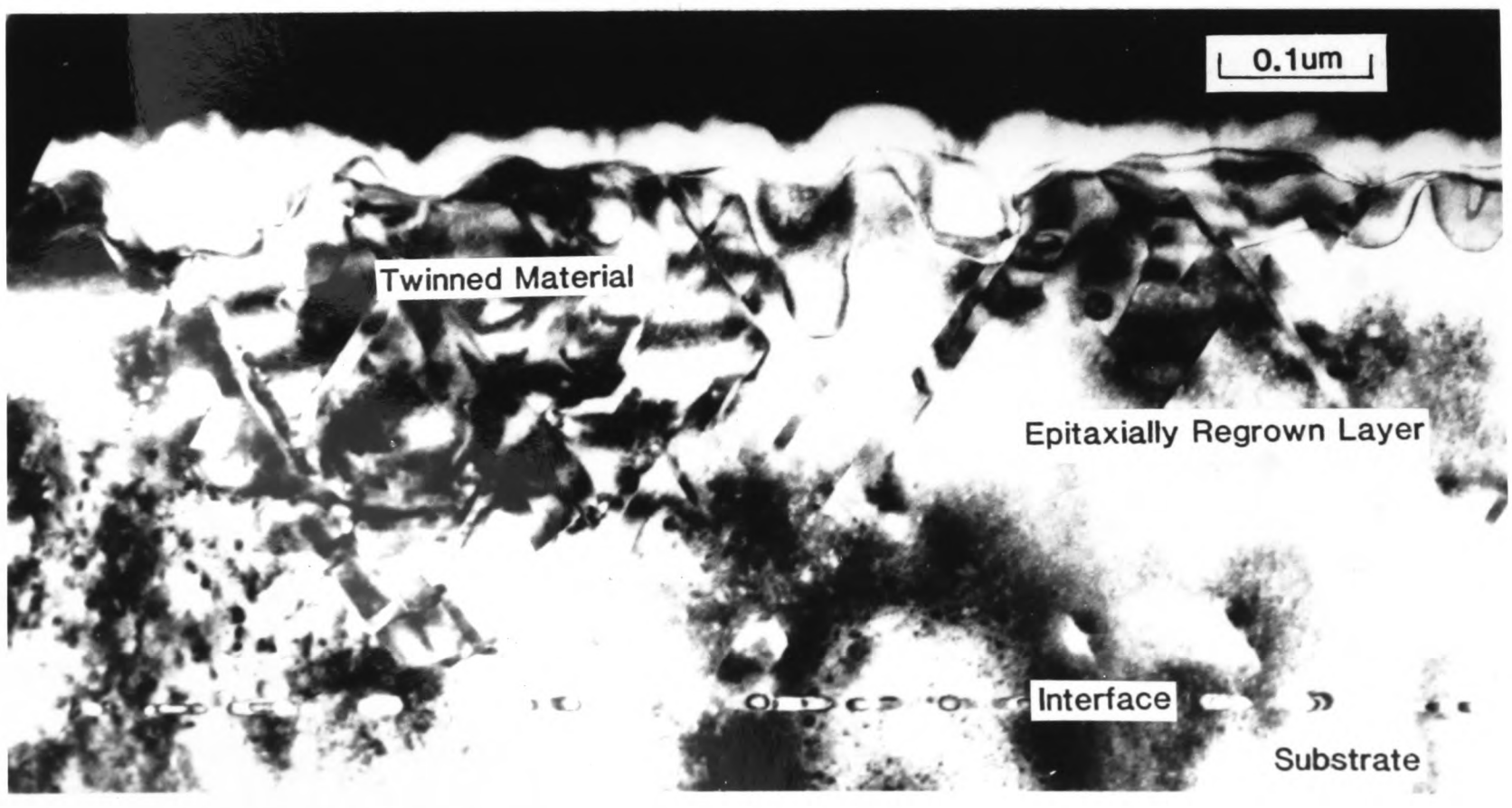
Sample 14B Fig.3.14a XTEM



Sample 14B Fig.3.14b XTEM



Sample 13C Fig.3.15 XTEM



Sample 14C Fig.3.16 XTEM

Sample 14A is the equivalent RCA cleaned sample. This sample exhibits small areas of regrowth typically 50Å×150Å (fig.3.12). The regions of regrowth are separated by approximately 200-1000Å. The substrate interface has begun to corrugate. Oxide thickness between regrowth regions is approximately 12Å, the apparently unchanged oxide thickness is indicative of the small displacement of oxide neccessary for the small number of observed regrowth sites.

In sample 13B (HF, 1100°C pre-anneal) the polysilicon layer has completely regrown. The regrown layer is largely defect-free for approximately half the layer thickness from the interface, and a surface layer containing more twinned material (fig.3.13a). The interfacial layer has formed into a planar layer of balls and dumb-bell like particles at approximately the depth of the original interface (Fig.3.13b). These particles are approximately 57Å thick, are facetted on the (100) and (111) type planes and are separated by approximately 300Å. A (110) diffraction pattern from the regrown area is shown in fig.3.13c. The extra spots due to the presence of twins can be clearly seen.

The equivalent RCA sample is 14B. Fig.3.14a is a diffraction contrast micrograph of the polysilicon layer. From this it can be seen that extensive epitaxial regrowth has occurred, but the layer is not yet completely regrown. The regrown regions contain a density of twins similar to that of 13B. The oxide layer has again formed balls and dumb-bell shapes. These are approximately 67Å thick. It is interesting to note the large extent of break-up of the amorphous layer, considering the sometimes limited regrowth as shown in fig.3.14b.

Sample 13C (HF, 1150°C pre-anneal) corresponds to complete epitaxial regrowth (fig.3.15). There are a small number of twins remaining, but the layer is largely defect free. The interfacial layer has formed facetted particles approximately 78Å thick, separated by 300-700Å.

The equivalent RCA cleaned sample (14C) also has an epitaxially regrown polysilicon layer (fig.3.16). This layer has a heavily twinned surface region but is relatively defect free nearer to the interface, it is very similar to sample 13B

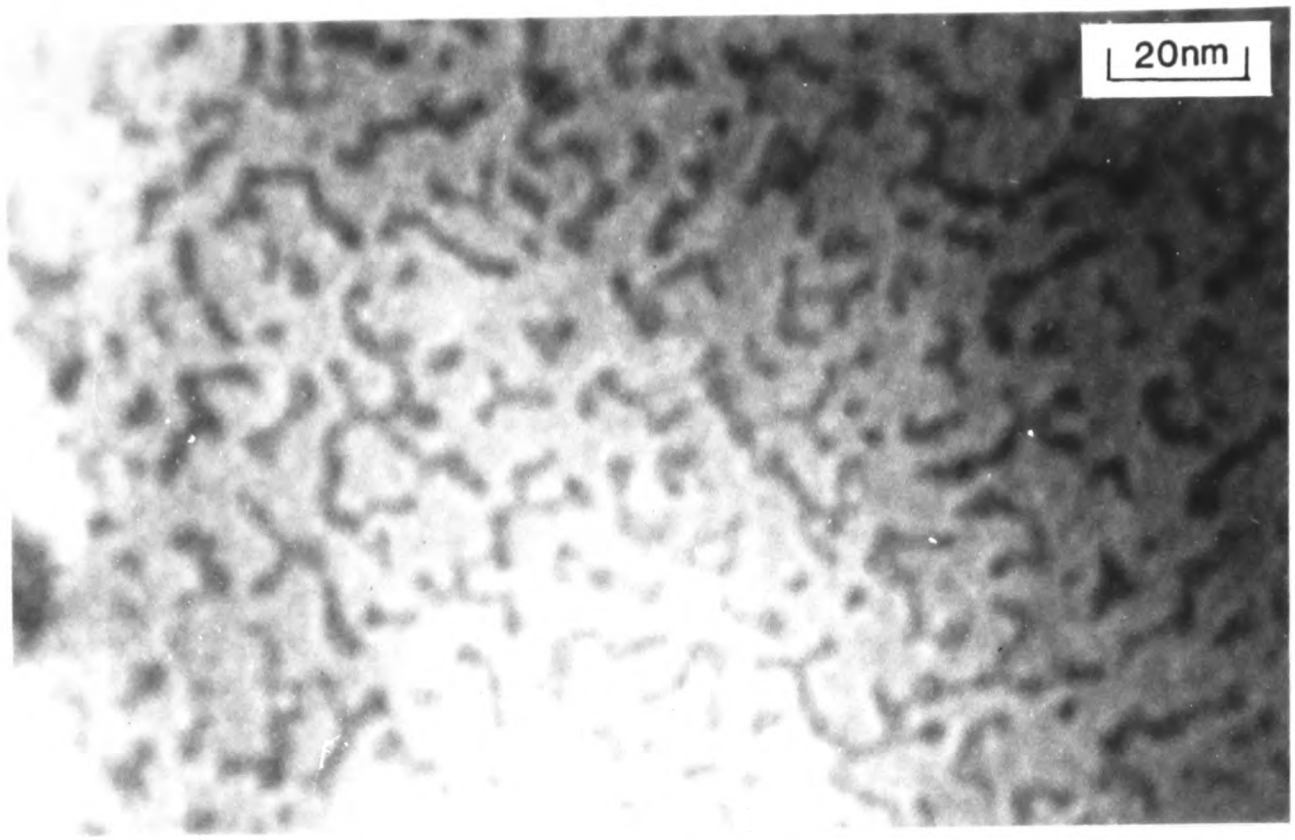


Fig.3.17 HF 950C pre-anneal, P doped, PVTEM

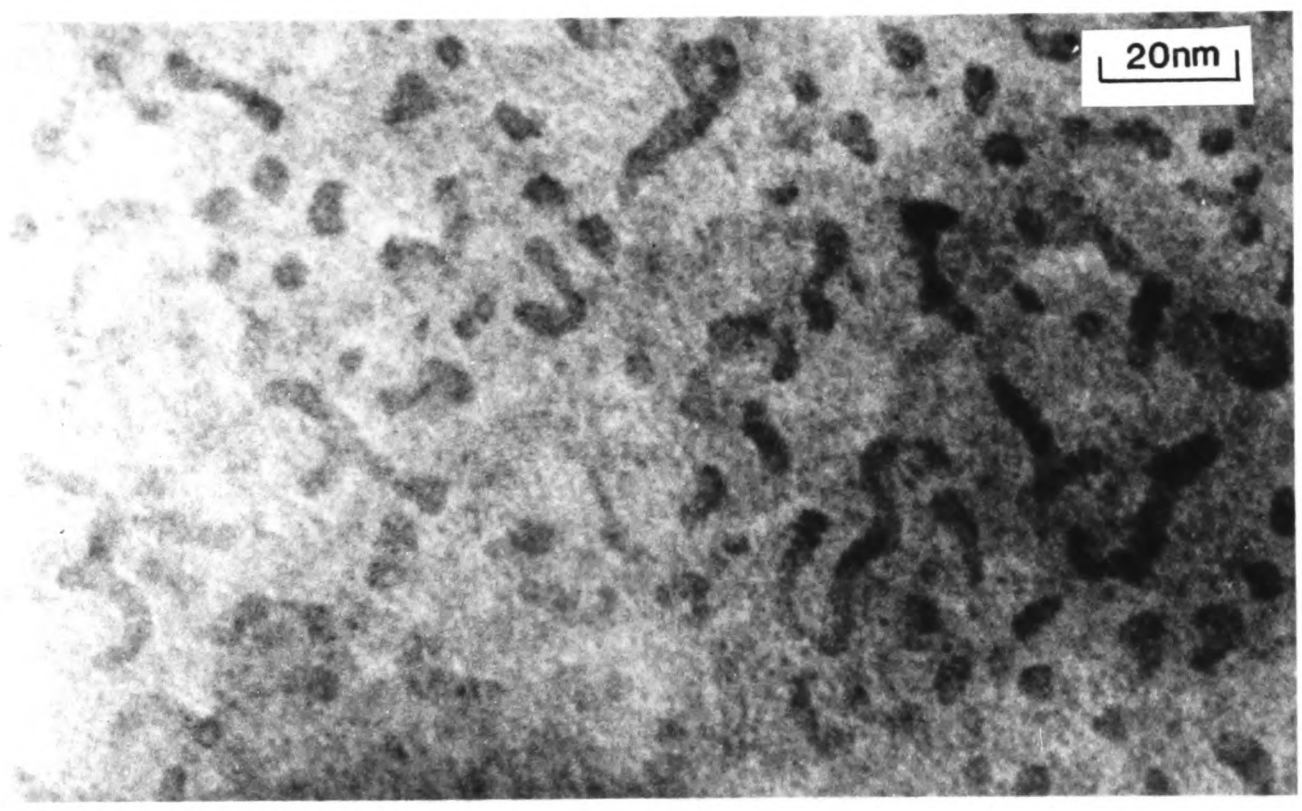


Fig.3.18 HF 1000C pre-anneal, P doped, PVTEM

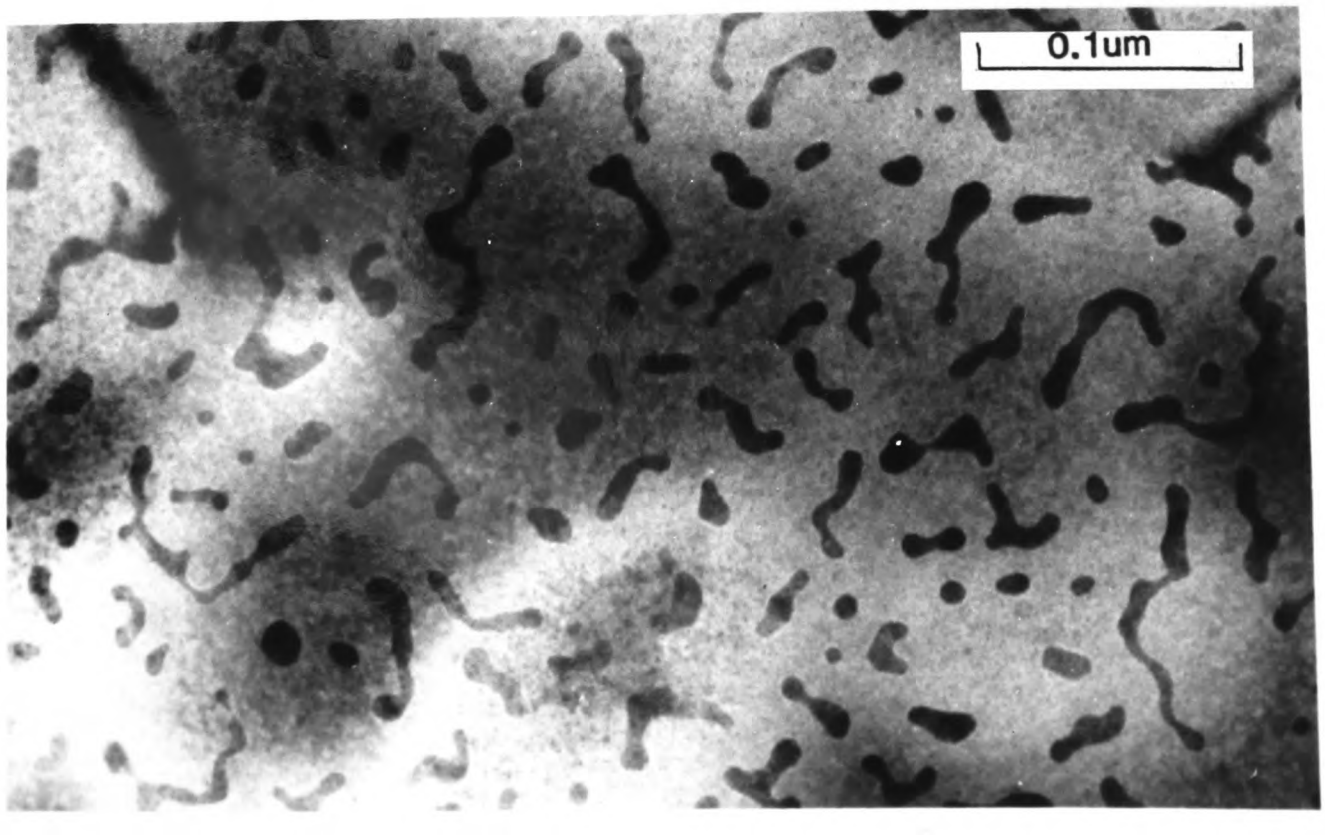


Fig.3.19 HF 1100C pre-anneal,P doped,PVTEM

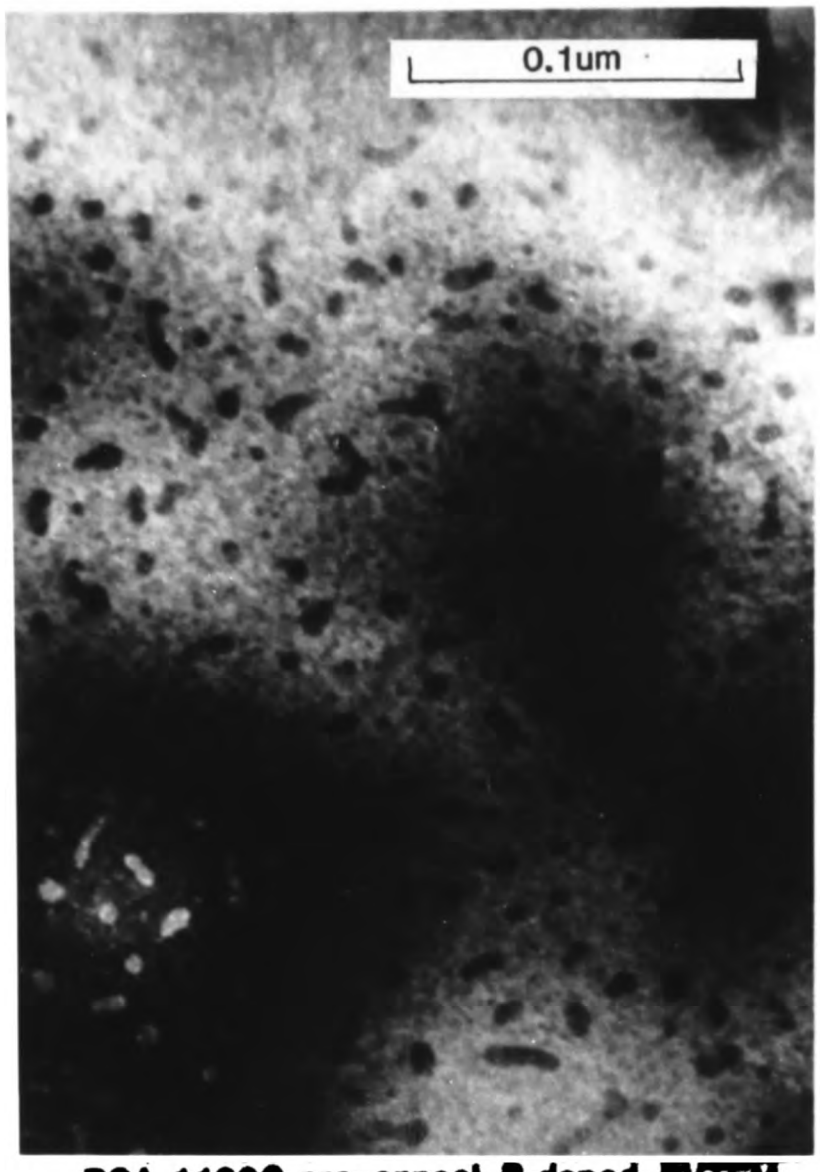
in general structure. The oxide layer in this case is facetted and approximately 85-90Å thick.

3.4.2 PLAN-VIEW TEM

The results that will be described in this section, were obtained from (001) type foils, prepared as described in the previous chapter. To obtain significant results, it was necessary to use samples with extensive regrowth of the polysilicon layer. The samples used are those described in chapter 4, where full processing details are listed (table 4.1). In this section, the samples will be described, as above, by the pre-clean, the dopant species and the pre-anneal temperature.

To obtain quantitative information about the oxide layer from the PVTEM micrographs, they were analysed using a commercial image processing package to carry out oxide particle size and distribution analyses. This technique is fully discussed in chapter 2. The technique was used to obtain measurements of the percentage of interfacial area covered by oxide, and to extract particle size distributions. The measurement of the fraction of oxide free area is independent of magnification whilst the particle size distribution is not. The micrographs were image processed at a magnification suitable to the hand tracing necessary for this procedure. For particle size analysis each set of areas was then corrected to a common magnification of 1 pixel=7.3Å as a linear measurement, i.e. 1 pixel=53Å^2. For measurement of the oxide covered area, a percentage was obtained corresponding to the total particle area, divided by the total analysed area of the image, as detailed in chapter 2.

The HF, P, 950°C pre-anneal sample is shown in fig.3.17. The oxide has broken up into a series of worm-like and dumb-bell like particles. These are about 70Å-250Å×40Å, and are separated by about 50-100Å. The surface area covered by these particles at the interface is 29%, as measured using the technique discussed in chapters 2 and 4. The thickness of the oxide particles, measured from

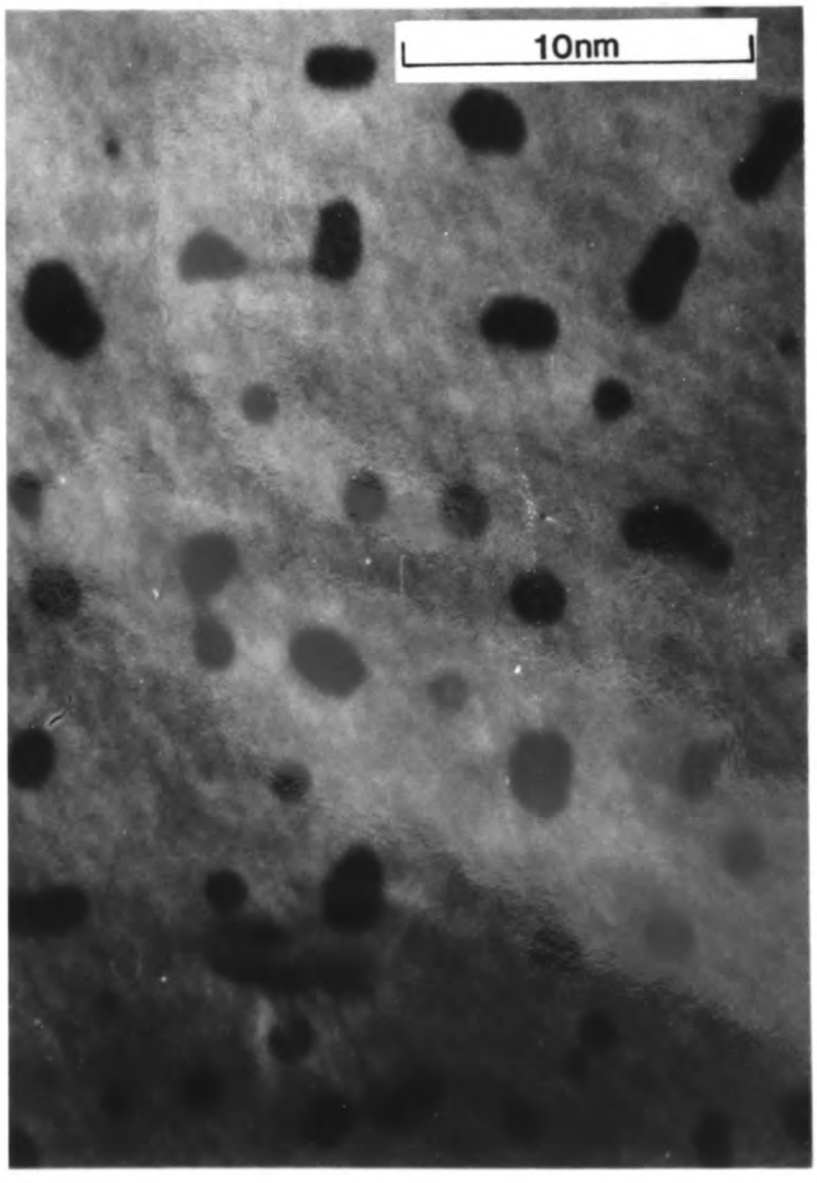


RCA 1100C pre-anneal, P doped, PVTEM

Fig.3.20a

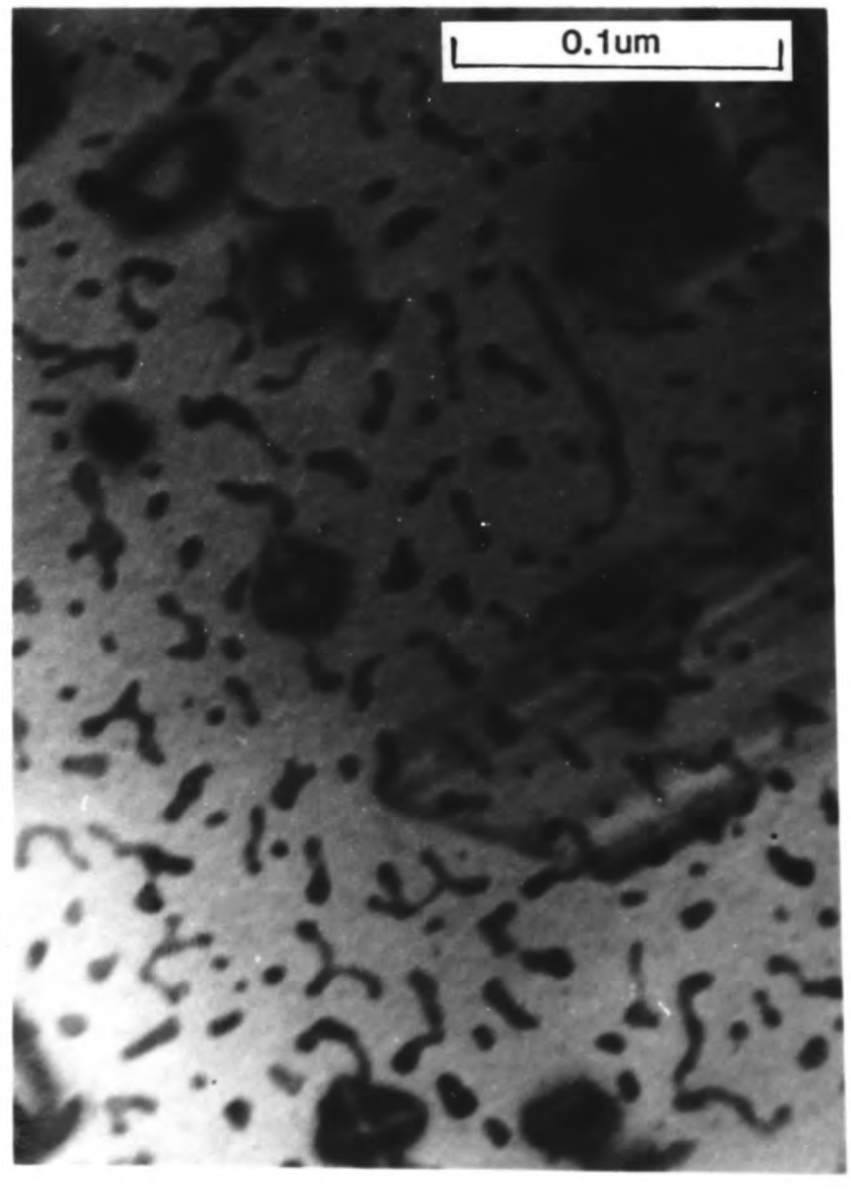
RCA 1100C pre-anneal, P doped, PVTEM

Fig.3.20b



RCA 1100C pre-anneal, As doped, PVTEM

Fig.3.22



HF 1100C pre-anneal,As doped

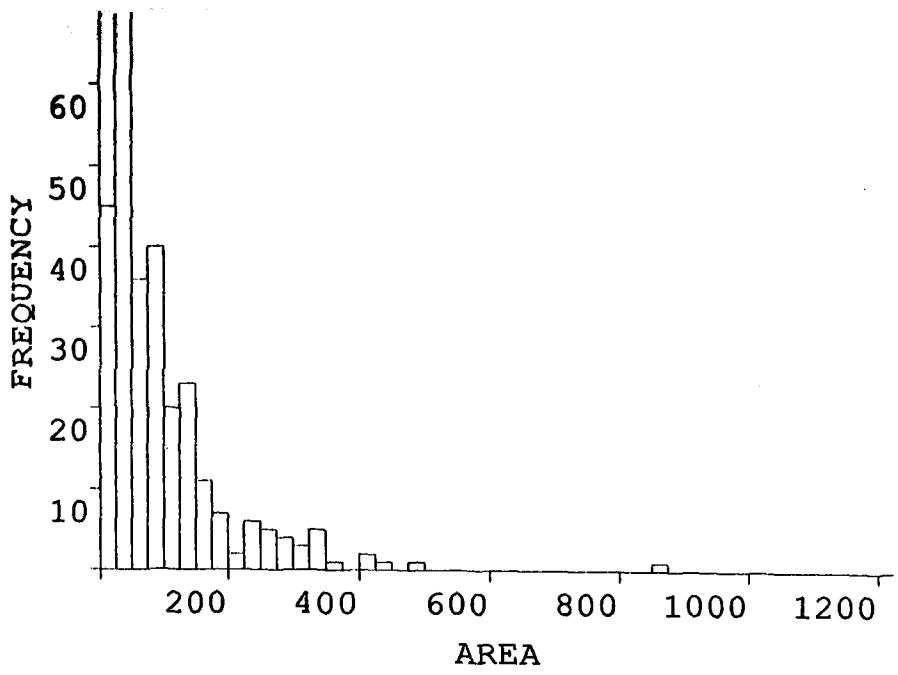
Fig.3.21

the HREM image is approximately 25-30Å. Fig 3.23a gives the histogram of particle size distribution for this sample. The total analysed area was approximately $0.47\mu m^2$. The average particle size was approximately $3800Å^2$, and the mode was approximately $2650Å^2$. The distribution of the particles fell away sharply from the mode with only one particle larger than $26500Å^2$.

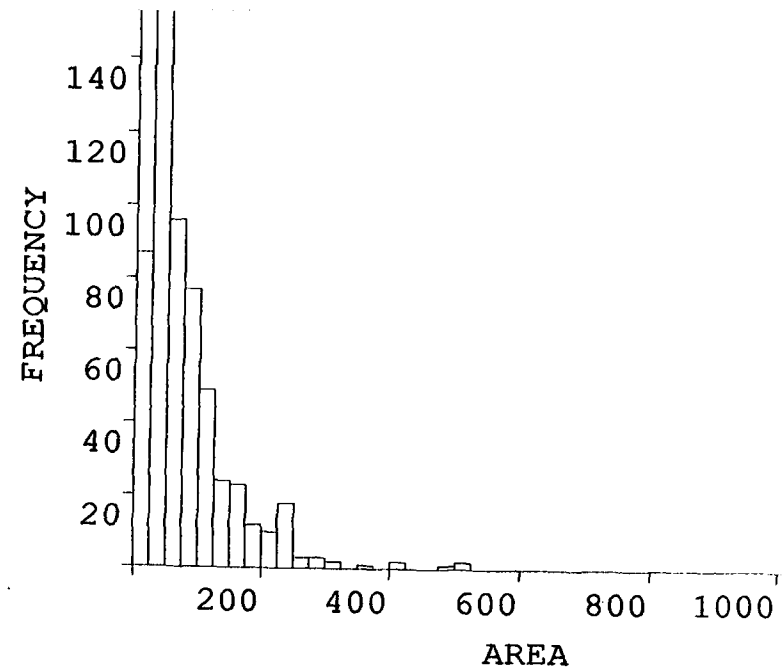
The HF, P, 1000°C sample is illustrated in fig.3.18. The shape and distribution of these particles is similar to the previous sample. The particles are approximately 70-220Å×40Å. The particle thickness is approximately 38-45Å. The interfacial area covered by these particles is approximately 20%. The corresponding histogram of surface area is given in fig.3.23b, the sampled area was $0.1\mu m^2$.

The HF, P, 1100°C sample is shown in fig.3.19. Once again the general shape and distribution of the particles is similar, however in this case, they are generally somewhat larger. The particles are approximately 70-900Å×57Å. The interfacial surface area covered by the sample is 12.4%. Fig.3.23c illustrates the corresponding histogram, taken from a total sampled area for this histogram of approximately $2\mu m^2$. Once again the mode is shared with the two previous samples, however there is a range of particles up to $66250Å^2$ with similar peak heights. This is indicative of the general shape of the curve which is spread out towards the higher end of the histogram, compared with the previous pair, having particles as large as $106000Å^2$. The average particle size has increased approximately four times to $14500Å^2$.

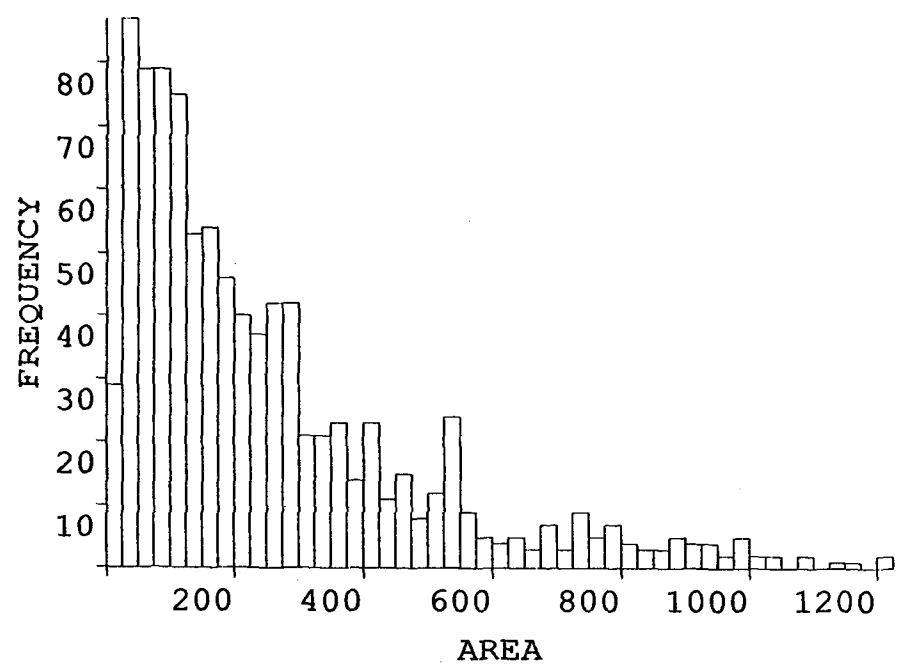
The RCA, P, 1100°C sample on the other hand has a demonstrably different oxide structure. Fig.3.20a reveals the oxide to consist of approximately equiaxed particles having dimensions of approximately 40-250Å. The area of the interface covered by oxide is 12.4%. A lower magnification image illustrates the particles surrounded by regrowth generated dislocations (fig.3.20b). This value is, apparently coincidentally identical to the value for the HF 1000°C sample.



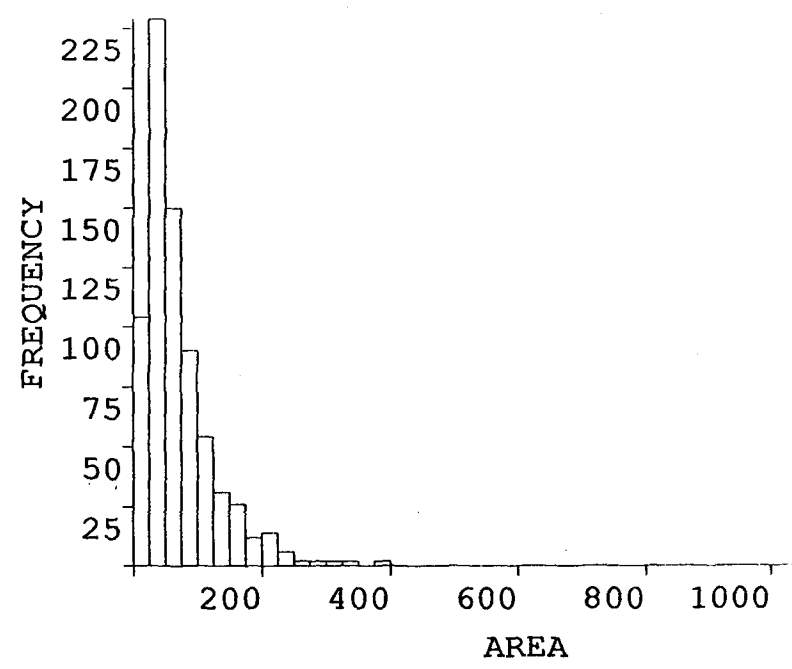
a)phosphorus HF 950C



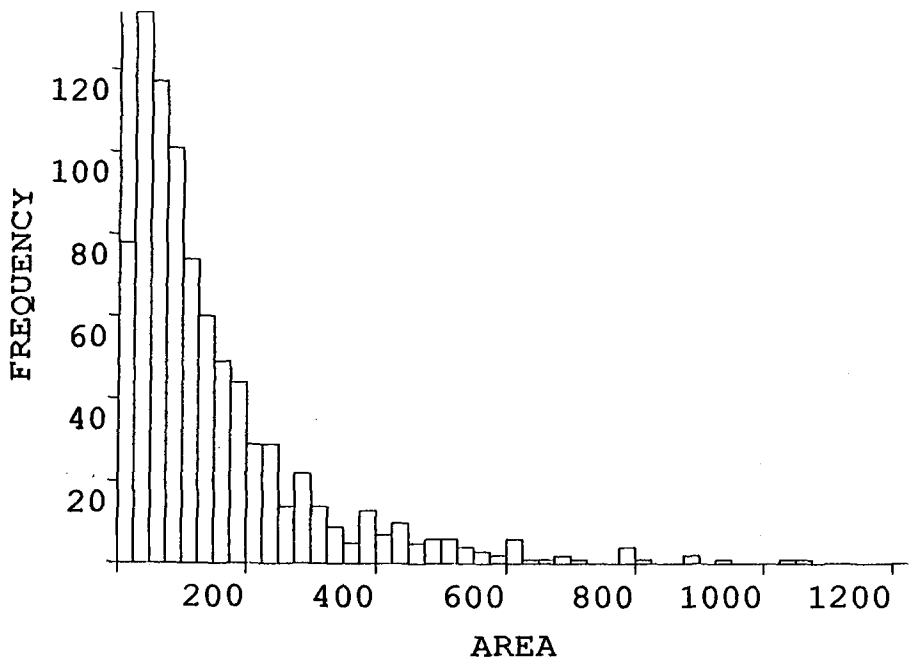
b)phosphorus HF 1000C



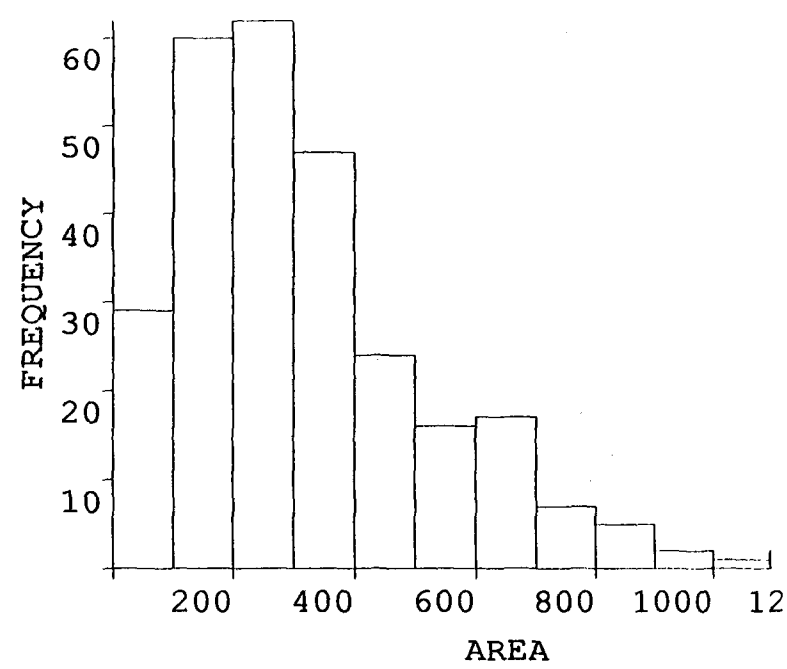
c)phosphorus HF 1100C



d)phosphorus RCA 110(



e)arsenic HF 1100C



f)arsenic RCA 1100C

Fig.3.23 Particle Surface Area Distributions

Careful examination of these particles shows that some of them are facetted. The corresponding histogram for this sample is given in fig.3.23d.

It is however more difficult to examine the oxide structure of the As samples in plan-view. It is only possible to examine the oxide in this way when the polysilicon layer has regrown away from the interface. This has only occurred in the As doped samples when the pre-anneal temperature is 1100°C. It can be seen from the pre-anneal results given above (13B and 14B), that this regrowth has already occurred during the pre-anneal. The HF, As, 1100°C sample is illustrated in fig.3.21, this shows the dumb-bell and worm-like structures typical of the P doped HF samples. The particles are 65-550Å×65Å, and are separated by about 70-130Å. The area covered by the interface is approximately 14.9%. The corresponding histogram of surface area distribution is given in fig.3.23e.

The equivalent RCA cleaned sample is shown in fig.3.22. The particles in this sample have a similar shape to the equivalent P doped sample, (i.e.they are approximately equi-axed) with dimensions of approximately 80-390Å. These particles are separated by about 130-230Å. The corresponding histogram of surface area distribution is given in fig.3.23f. The average particle area is greater than for any other sample measured

3.5 DISCUSSION

3.5.1 OXIDE BREAK-UP

Clearly, the two different interfacial oxides give rise to different shaped particles when broken up by the heat treatments discussed here. The HF clean oxide produces long worm-like particles when observed in plan view, whilst the RCA clean oxide results in particles which are equi-axed in the plane of the interface and therefore have a disc-like shape. This behaviour may be explained by considering the mechanisms resulting in break-up of the oxide. Firstly, break-up of the oxide layer is driven by the reduction in free surface area resulting from spherodisation of the oxide and secondly, the layer may be broken up due to displacement as the epitaxial regrowth pierces the layer. Epitaxial regrowth is driven by reducing both the total grain boundary area and the interfacial 'boundary' area, and this is the reason for the observed regrowth up grain boundaries which intersect the interface. The driving force for the epitaxial regrowth depends upon the surface energy of the particular grain boundary, clearly then regrowth will be more favourable up high energy grain boundaries. The epitaxial substrate covered by oxide 'sees' these boundaries as an image force[14] which drives local perforation of the oxide layer, this is the nucleation process for epitaxial regrowth through these oxide layers. The effect of the image force is inversely proportional to the square of the separation of the substrate and the grain boundary and therefore perforation will be favourable at more sites in the HF cleaned sample than in the RCA cleaned samples. From a comparison of the observed oxide balling behaviour and the description given above, it is clear that break-up of the HF layer results primarily from the contribution of the epitaxial regrowth (deduced from the non-circular projected shape of the particles), whilst by the same argument the extra thickness of the RCA oxide allows the reduction in oxide surface area to dominate in its break up.

If the surface area distributions are now examined, it is found that for the HF clean, P doped samples, the free area of the interface (area not covered

by oxide) gradually increases as the pre-anneal temperature increases. Growth of the particles occurs in the 950°C and 1000°C samples largely by thickening perpendicular to the interface, the planar shape and area distribution increasing only slightly. In the 1100°C sample the particles are coarsening in the plane of the interface, resulting in the shape of the distribution becoming extended towards the larger particle sizes. This suggests that for this sample, the driving force for the reduction of oxide surface area is beginning to influence the particle distribution, however the particle shapes still suggest that initial break-up is dominated by epitaxial regrowth. In comparison, the As doped sample has a similar thickness perpendicular to the interface, but the distribution is not as broad as for the phosphorus doped sample. The general shape of the particles is very similar to that of the HF, P doped samples, having the long worm-like structure typified by HF samples. As break-up in these samples is driven by epitaxial regrowth and regrowth is not enhanced by the arsenic in this sample, the oxide coarsening would be expected to be less advanced, as observed in the particle area distributions.

Fewer results are available for the RCA oxide, due to its effect on suppressing regrowth. The distribution of particles sizes in the P doped, RCA 1100°C sample peaks lower than for the equivalent As sample. The general particle shapes are similar to those of the P doped samples, clearly identifying the RCA clean sample. It was necessary to use a larger channel width for this histogram (fig.3.23f) as the smaller number of particles observed introduces too much statistical noise with a narower channel width. The oxide particles in the As doped sample are larger than for the phosphorus doped sample.

In general terms, the exposed area of interface is similar for both RCA and HF oxides after an 1100°C pre-anneal. The P doping however appears to result in smaller particles compared with As for the RCA oxide whilst the opposite is true for the HF oxide. As described in the previous section, the phosphorus enhances regrowth compared with the arsenic for the HF oxide, this results in a

larger number of regrowth sites (grain boundaries) at the interface, and hence a larger number of small oxide particles. In the case of the RCA oxide, enhanced regrowth is not observed, it appears however that P doping results in smaller oxide particles for the RCA samples than As.

3.5.2 EPITAXIAL REGROWTH

The progressive stages of the observed regrowth clearly follows the trends discussed in the introduction to this chapter, in addition, however, it is clear that epitaxial regrowth into the polysilicon proceeds largely up grain boundaries, irrespective of the dopant or of thermal processing. This is the expected behaviour for the case in which epitaxial regrowth is driven by a reduction in total surface area in the polysilicon (i.e. reduction in grain boundary area) and at the interface, as described previously.

Inspection of the pre-anneal only samples, reveals that the total area of the regrown regions for both the HF and RCA cases is similar, the primary difference being that there are 5-10 times as many regrowth sites in the HF cleaned sample. The similar sizes of these regions implies that they have been regrowing for similar periods of time; i.e. the appearance of gaps in the oxide in both cases occurs at similar times. The results of the 1100°C pre-anneal reveal, not surprisingly that the more regrowth sites that are present, the more polysilicon will be consumed during regrowth. A pre-anneal at 1100°C would be totally unsuitable with either clean for use in real devices, as it is imperative thatthe layer is polysilicon at the time of dopant implantation to provide a high diffusion rate dopant source for the substrate. An idealised situation might be to provide a polysilicon layer at the time of implantation (for the reasons described above), which then regrows epitaxially (to reduce emitter series resistance), whilst still retaining enhanced gain (which may be traded for a reduction in base resistance), but without increasing the thermal budget of the emitter (which would

result in a deeper junction). The results presented in this chapter suggest that such a situation may be possible, and this is investigated further in chapter 4.

The fully processed samples (2 and 8) illustrate clearly that, in the absence of dopant, the additional thermal budget associated with the emitter drive-in, does not produce significant structural change from that of the pre-anneal only.

The two doped samples which have received an RCA clean but no pre-anneal (9 and 11) show very similar structures in that there is no evidence of either oxide break-up or epitaxial regrowth. In the HF cleaned samples however, differences do occur and in this case, the P doped sample has begun to regrow.

The fully processed samples including pre-anneals, follow similar trends to those without pre-anneals. The RCA cleaned samples have a similar structure irrespective of dopant, to the undoped structure (8). Once again however, clear differences are observed in the HF cleaned structures (4 and 6). Comparing first the As doped and the undoped samples, the regrowth in the As doped case appears slightly greater. However, typical of both samples is an approximately even regrowth coverage over the interface. The P doped sample has undergone accelerated (or enhanced) epitaxial regrowth during the drive-in, resulting in a layer that is more than 50% regrown. The regrowth in this sample however takes the form of large islands, rather than the relatively even coverage of the As and undoped samples. This suggests that after pre-annealing, accelerated regrowth may only occur in a relatively small number of the available gaps. This appears to be confirmed by the behaviour of the regrown regions in the areas between islands, where regrowth is of similar dimensions to the other two samples. The centres of this island regrowth are separated by approximately 0.3μm or more.

It has been shown that the regrowth occurs by the reduction in grain boundary area, both in the silicon and at the interface. To achieve this reduction, the shape of the regrowth front must attempt to minimise its total surface area. To satisfy this condition, the regrowth front must be as nearly parallel to the

interface as possible. In the case of the RCA oxide, regrowth parallel to the interface is constrained by the larger volume of oxide which must be transported away from the regrowth region compared with the HF oxide. It is energetically unfavourable for the regrowth front to extend its aspect ratio vertically due to the increase in interfacial area, and so extended regrowth is constrained by the mechanism for break-up of the RCA oxide. Regrowth in the phosphorus doped, arsenic doped and undoped RCA samples is similar and therefore it appears that the driving force for the balling-up of the RCA oxide is independent of the presence of dopant for the conditions and dopants investigated here, and so enhanced regrowth is not observed for the P doped, RCA oxide sample.

3.5.3 ORDER IN THE OXIDE INTERFACIAL STRUCTURE

Unusual contrast was observed immediately below the interface in the RCA cleaned samples (unless extensive regrowth had occurred) of approximately two monolayers, the contrast being sharply defined. This contrast was not observed in any HF cleaned samples. The interfacial region between a silicon substrate and surface oxide layer has been the subject of investigation by a number of workers, detailed earlier in this chapter. The structure and electrical behaviour of thin oxides has become more important recently as devices have made use of reduced dimensions. Essentially, two important concepts arise from these investigations, namely stoichiometric fluctuation in the oxide layers, and ordering at the monocrystalline/polycrystalline interface. Grovenor and Cerezo[19] have observed that the stoichiometry of thin oxides can vary with the growth procedure. Specifically, they have investigated differences between hydrophobic (HB) 'native' oxides (HF clean) and hydrophilic (HL) 'native' oxides. The RCA produced oxide in this work falls into the latter category, although the detail clean used by Grovenor and Cerezo differed in detail from an RCA clean. They

observed that for the HB oxide, the stoichiometery varied approximately uniformly with layer thickness, the ratio of oxygen to silicon increasing as the layer thickness increased. For the HL oxide, the behaviour was different in that the layer was divided into two regions, the near interface region being approximately SiO, changing abruptly after approximately 7Å to a stoichiometry of SiO_2. This analytical procedure is unfortunately not suited to the examination of ordering in the oxide, as it is a surface analytical technique and surface reconstruction in a thin surface layer might confuse the results.

The second important concept proposed concerning the interfacial region is that a thin crystalline phase exists at the substrate/oxide interface. Three different experimental methods have been applied to this problem by different workers. Ross and Stobbs[20] applied a complex study of Fresnel fringes in diffraction contrast images and concluded that in thicker oxides, a separate crystalline phase exists at the substrate/oxide interface. Ourmazd[15] has investigated this using HREM microscopy and image simulation and suggests that a separate crystalline phase exists at the interface, which is identified as tridymite, as discussed above. Unfortunately insufficient description of the simulation technique is given to test its validity. Additionally, an apparent lack of the necessary projected symmetry of the investigated crystal in the actual images suggests that these images are not exactly on the pole. This would negate the value of any modelling and subsequent image matching to determine precise structure in the interfacial oxide.

Perhaps the most convincing evidence for the existence of an ordered layer at the Si/SiO_2 interface is the work of Fuoss et al [23]. In their work they have investigated the interface using glancing angle X-ray diffraction. They observe small crystalline regions of approximately 80Å-145Å in the plane of the interface, and these crystallites appear to bear some orientation dependence to each other. Their sampling area was very much larger than that of other techniques discussed here, being approximately $40mm^2$ corresponding to 10^{11}

particles. They conclude that the crystalline structure is similar to that of α-crystobalite, but lacks some of the weaker X-ray peaks and also the full rotational symmetry of α-crystobalite.

In the section on interfacial oxides, the difficulties associated with the use of HREM to interpret the oxide structure are discussed. Despite the attendant problems, it is possible to make a few simple observations on the present results. The contrast is abrupt and always of the same thickness. This suggests that the contrast does not arise from changes in the projected thickness of the crystalline region caused by interfacial roughness. The images are all taken from a position close to Gaussian focus where Fresnel fringing does not contribute significantly to the image. This was tested experimentally on samples similar to those investigated here, where the Fresnel fringes were not observed until a defocus at which the lattice contrast had nearly dissapeared. Symmetry appears to be retained in the regions of different contrast. This anomalous contrast is not observed in all places even within the small sample area of a single XTEM foil.

It is further interesting to note that this contrast is not observed in the HF oxide, nor is it observed in the thicker oxide balls under any imaging conditions.

3.6 CONCLUSIONS

In this chapter, the behaviour of the oxide interfacial layer and of the polysilicon layer in blanket structures similar to those of real polysilicon contacted devices have been investigated in the presence of a novel pre-anneal prior to dopant implantation. The structural changes have been examined at all stages of the process and also for a number of different pre-anneals, in the presence of a 'native' oxide (4-7Å thick) and also a chemically grown oxide (14Å thick), for P doped and As doped polysilicon layers.

It was observed that in the presence of the thinner oxide, the P doped polysilicon layer regrew faster than the As doped layer for the doping levels investigated here. The presence of the thicker purposefully grown oxide supressed this enhanced regrowth entirely. It was concluded that phosphorous in the concentration used, enhances regrowth of the polysilicon more than arsenic, whilst neither has a significant effect on the rate of break-up of the oxide.

The two types of oxide examined were found to result in characteristic, different particle shapes after being broken-up. It is concluded that the break-up of the thinner oxide is controlled by the epitaxial regrowth mechanism, whereas the driving force for the break-up of the thicker oxide is controlled largely by the reduction in surface area of the oxide resulting from it balling up, the driving force for epitaxial regrowth having little effect. The observed shape of the oxide particles when viewed in the plane of the interface is explained by these mechanisms. The exposed area, for the 1100°C pre-annealed and phosphorus doped samples, was found to be similar for both types of oxide.

Epitaxial regrowth into the polysilicon layer, was observed to proceed up grain boundaries, irrespective of the thermal treatments or the presence or absence of any dopant species. Examination of the pre-annealing technique suggests that a pre-annealing temperature of between 1000°C and 1100°C for ten minutes would enable a useful device to be constructed. The possibility of producing a device having polysilicon layer at the time of implantation, allowing a shallow junction, but a single crystal contact and broken oxide at the end of processing (resulting in a reduced series resistance) without increasing the emitter depth, is proposed. In the next chapter, the possibility of fabricating such a device is investigated.

Unusual, abrupt contrast was observed in the Si substrate, close to the Si/oxide interface under HREM imaging conditions. Such contrast has been previously observed by other workers. This contrast has been ascribed to the

presence of a thin layer of crystalline oxide. No successful HREM image simulation of this region has yet been carried out, however the present work indicates strong circumstantial evidence for some real interface effect, such as an ordered oxide or sub-oxide layer. Other evidence suggests that this is a reasonable supposition. However, perhaps the most important conclusion, combining the present work with that of other workers is that the nature of the interface region is strongly dependent on the precise, and at present poorly understood differences in oxide fabrication conditions, in addition to the oxide layer thickness.

References

[1] . S. Duncan, M.C. Wilson, P.C. Hunt and D.J. Bazley. Symposium on VLSI Technology, San Diego pp.87-88.1988

[2] . T. Sakai, Y. Kobayashi, H. Yamauchi, M. Sato and T. Makino. Jap. J.Appl.Phys., 20, suppl.20-1, pp.155-159, 1980.

[3] . G.R. Wolstenholme, N. Jorgensen, P. Ashburn, G.R. Booker. M J.Appl.Phys. 61(1), pp.225-233, 1987.

[4] . H.C. de Graaff and J.C. de Groot. IEEE. Trans.Electron.Devices. ED-26, pp.1771-1776, 1979.

[5] . N. Jorgensen, J.C. Barry, G.R. Booker, P. Ashburn, G.R. Wolstenholme, M.C. Wilson, P.C. Hunt. IOP Conf. Ser. 76(11), pp.471-476,1985.

[6] . M.Y. Ghanriam and R.W. Dutton. Appl.Phys.Lett. 51(8), pp.611-613, 1987.

[7] . B.Y.Tsaur and L.S. Hung. Appl.Phys.Lett., 37(7), pp.648-651, 1980.

[8] . A. Neugroschel, M. Arienzo, Y. Komen and R.D. Isaac. IEEE Trans. Electron Devices, ED-32, pp.807-816, 1985

[9] . G.L. Patton, J.C. Bravman and J.D. Plummer. IEEE Trans. Electron Devices, ED-33, pp.1754-1768, 1986

[10] . D.W.Greve, D.L. Chen and A.M. Gruzman. BCTM, Minneapolis, pp.150-153, 1987.

[11] . B. Benna, T.F. Meister and H. Schaber. Solid State Electronics, 30, pp.1153-1158, 1987.

[12] . I. Ohdomari, T. Mihara and K. Kai. J.Appl.Phys., 60(1), pp.3900-3904, 1986.

[13] . N. Jorgensen, Private Communication.

[14] . K.Y. Ahu, U. Gosele and P. Smith. Mat.Res. Soc. Symp.Proc., Vol 107, 1988.

[15] . A. Ourmazd, D.W. Taylor and J.A. Reutschler. Phys.Rev.Lett., 59(2), pp.213-216, 1987.

[16] . J.W. Matthews and A.R. Blakeslee, J. Crystal Growth, 27, p.118, 1974.

[17] . B.R. Weinberger, G.G. Peterson, T.C Eschrich and H.A. Kranski. J.Appl.Phys., 60(9), pp.3232-3234, 1986.

[18] . M.A. Taubenblatt, PhD thesis, Tech. Rep. G-740-1, Stanford Electron. Lab., Stanford University, 1985.

[19] . C.R.M. Grovenor and A. Cerezo. J.Appl.Phys., 65(12), pp.5089-5095, 1989.

[20] . F.M. Ross and W.M. Stobbs. Surface and Interface Analysis, 12, 33-44, (Wiley) 1988.

[21] . A. Albu-Yaron, J.C. Barry, N. Jorgensen, G.R. Booker, P. Ashburn and P. Ward. British Association for Crystal Growth, Annual Conference, Lancaster, UK, 1984.

[22] . G.R Wolstenholme, PhD thesis University of Southampton, 1988.

[23] . P.H. Fuoss, L.J. Norton, S. Brennan and A. Fischer-Colbrie, Phys.Rev.Let. 60(7), pp.600-603, 1988.

Chapter 4

EFFECT OF PRE-ANNEALING ON BIPOLAR CHARACTERISTICS

4.1 INTRODUCTION

The polysilicon contacted emitter has become the forefront technology in the development of high speed bipolar silicon devices for VLSI applications. Since the structure was first reported [1], considerable effort has been devoted to the understanding of the mechanisms controlling the operation of such devices.

Polysilicon contacted devices have several advantages in VLSI applications. The principal driving forces for development have been the self-aligning nature of the structure, and the ability to reduce the total area of the depletion region and hence control the associated capacitance. This in turn reduces the internal RC time constants, and enables higher switching speeds to be obtained. Additionally, gain enhancements of up to ten times those of comparable conventional transistors have been reported [2].

This gain enhancement has been shown to be due to suppression of the base (hole) current as opposed to enhancement of the collector current [3]. An explanation for this effect was presented by de Graaff and de Groot [2] in which they attributed the hole supression effect to the presence of a thin interfacial oxide at the polysilicon/monosilicon interface. It was proposed that this oxide presented a higher tunnelling barrier to holes than electrons.

Ning and Issac proposed a second theoretical model in which they attributed the supression of the base (minority carrier or hole) current to reduced mobility in the polysilicon [4]. Patton [5] has observed a strong blocking effect at the polysilicon/monosilicon interface which is strongly dependent on the doping level. One possible explanation proposed by Patton is the existence of a 'high-low' barrier at the monocrystalline/polycrystalline silicon interface. Patton suggested that the observed segregation of dopant [6] [7] (arsenic in the case of Patton's work) to interfaces has a high concentration of donor states associated with it. This assumes of course that the segregated dopant is electrically active. It was suggested that this would present a thermionic barrier to carrier transport. Neugroschel et al [8] also observed that the segregation of arsenic to the polysilicon/monosilicon interface was important in depressing the hole current. They suggested that minority carrier transport is dominated by a 200-300Å thick highly disordered layer in the polysilicon immediately above the polysilicon/monosilicon interface.

More recently attempts have been made to unify these models, principally by Eltoukhy and Roulston [9] and by Yu [10]. Ashburn and Soerowirdjo [11] experimented with devices fabricated with and without interfacial layers and concluded that the minority carrier current (hole current) suppression was controlled by interfacial layer tunnelling when the layer was continuous, and by the polysilicon transport properties when it was discontinuous.

To study transport effects in the polysilicon layer, several authors have investigated the effects of reducing the layer thickness [5][8] [12] [13]. These results show that the base current is independent of polysilicon layer thickness for layers greater than approximately 500-1000Å thick. Ashburn [14] has carried out similar experiments on devices closely similar to those examined here. He observed no change in base current with polysilicon layer thickness for layer thicknesses of 0.1μm or greater. However for a layer thickness of 0.04μm a $\times 4$ increase in base current was observed.

4.2 EXPERIMENTAL DATA

As stated in chapter 3, it is generally desirable to break up the oxide in a controlled manner to ensure reproducibility from device to device and from batch to batch and, for many applications to bring the emitter resistance down to a level usable in VLSI applications [15]. To ensure this, the devices are subjected to a relatively high temperature annealing treatment ($\geq 950°C$) which has a detrimental effect on the depth of the emitter in the single crystal silicon. Additionally, it would be desirable to epitaxially regrow the polysilicon layer fully, as this would further reduce the emitter resistance. Polysilicon presents a significantly higher resistance to current flow than monocrystallineSi without defects. In practice, epitaxial regrowth has not yet been utilised in real devices.

Reducing the interfacial resistance by annealing can not only result in a larger area of depletion region by increasing the sidewall component, but also results in the loss of some or all of the depressed base current effect. If some of this gain enhancement were retained, this could be traded for higher doping levels in the base. This reduces the base resistance, and as the base resistance falls, the rate at which the input capacitance can be charged increases. This produces an increase in f_{mosc}, the maximum oscillation frequency.

This investigation examines the changes in device electrical characteristics, resulting from the implementation of the pre-annealing technique discussed in chapter 3. In particular, the results of this investigation are used to examine the various contributions to the hole current supression and explain the progressive gain reduction which occurs as the pre-annealing temperature increases. These pre-anneals were carried out on various samples, immediately following the deposition of the polysilicon but before dopant implantation as detailed previously. This technique produces devices of closely similar doping profiles whilst allowing the structure of the polysilicon and interface to be varied.

Base oxidation (25' dry O2 1100C)
Base implant (2E13 cm-2, B, 40keV)
Base drive-in (30', dry O2, 950C)
HF etch until hydrophobic
Interfacial layer clean (HF or RCA)
Emitter polysilicon deposition
Pre-anneal (10', dry N2 at 900C
950C
1000C
1100C)
Emitter implant (1E16 cm-2, P, 50keV)
Emitter drive-in (60', 850C, wet O2)

Table 4.1

Phosphorus implanted samples are examined in this work, and the structures of these samples are investigated using cross-sectional TEM (XTEM), and plan-view TEM (PVTEM) in conjunction with image processing. By comparing these with electrical data obtained by Graham Wolstenholme at Southampton University, the implication of these results on mechanisms for gain enhancement in polysilicon transistors are considered

4.2.1 TEM OBSERVATIONS

The pre-annealed samples examined in the present work were grown on 3" diameter (001) n-type Czochralski silicon wafers. The processing history of the samples is given in table 4.1. Subsequent to the identical base treatments, the wafers were cleaned by one of two methods, either an RCA or an HF wafer surface pre- clean. The details of both the RCA and HF clean techniques are detailed in chapter 3. The RCA clean leaves a deliberately grown surface oxide approximately 14Å thick. The HF clean on the other hand is designed to remove any native oxide. It has been shown previously however, that a thin, non-uniform oxide approximately 0-8Å thick remains after this surface treatment [16]. Immediately following the appropriate treatment, a layer of undoped polysilicon approximately 0.4μm thick was deposited on the wafer in a commercial LPCVD reactor at 610°C.

The pre-anneals were all carried out by furnace annealing in a dry nitrogen ambient. The samples were annealed for 10 minutes at a variety of temperatures from 900°C to 1100°C prior to an emitter implant, in a similar manner to those discussed in chapter 3. They were then examined by both XTEM and PVTEM. All cross-sections were (110) type, and all plan views were (100) type. All HREM imaging was carried out in JEOL 200CX TEM operating at 200kV. The bulk of the diffraction contrast imaging was carried out in a Phillips CM12 TEM operating at 120kV.

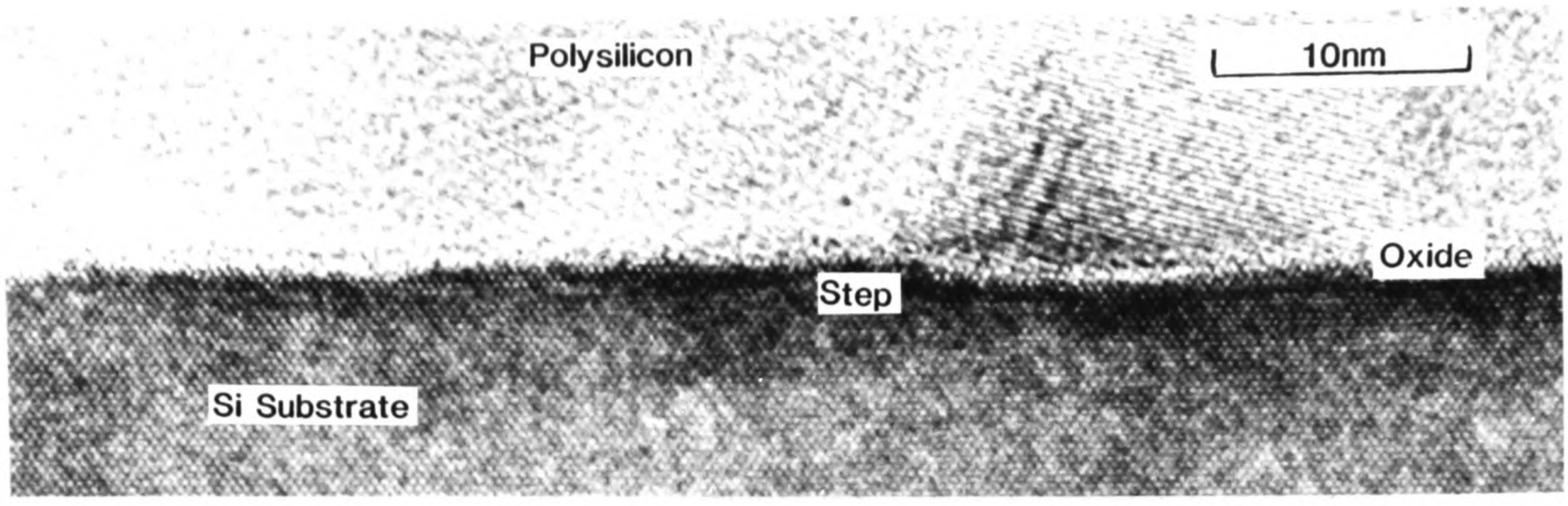


Fig. 4.1 RCA, drive-in XTEM

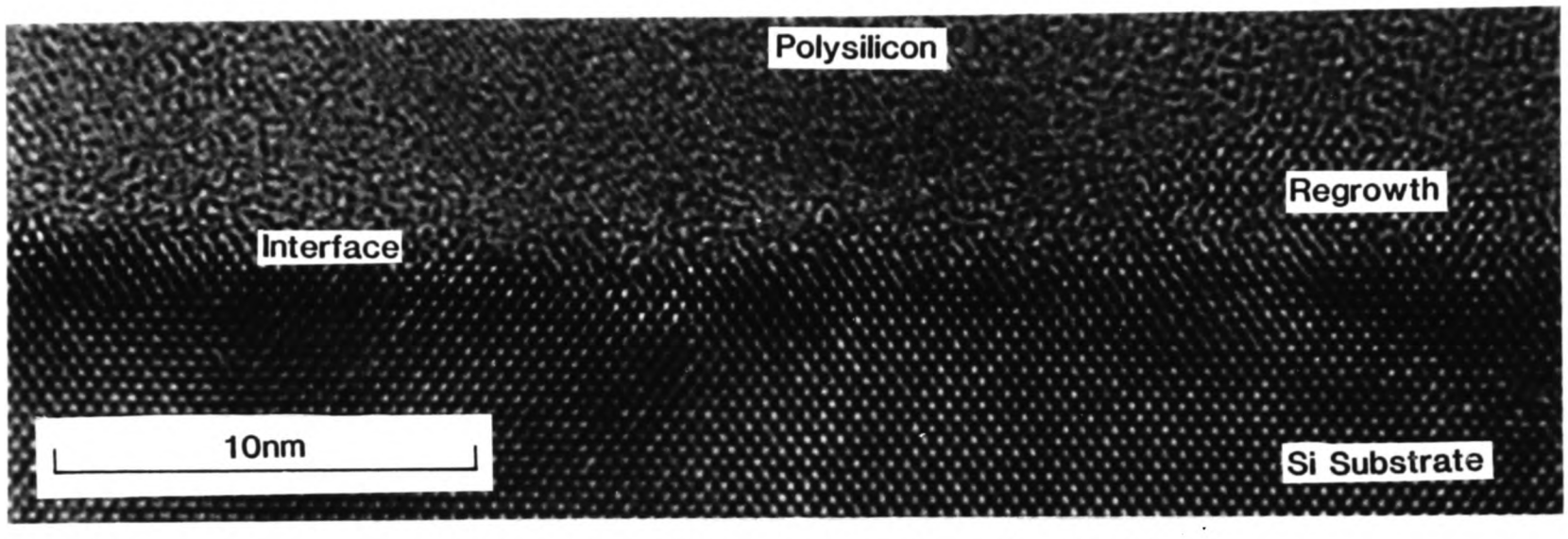


Fig.4.1 HF, drive-in XTEM

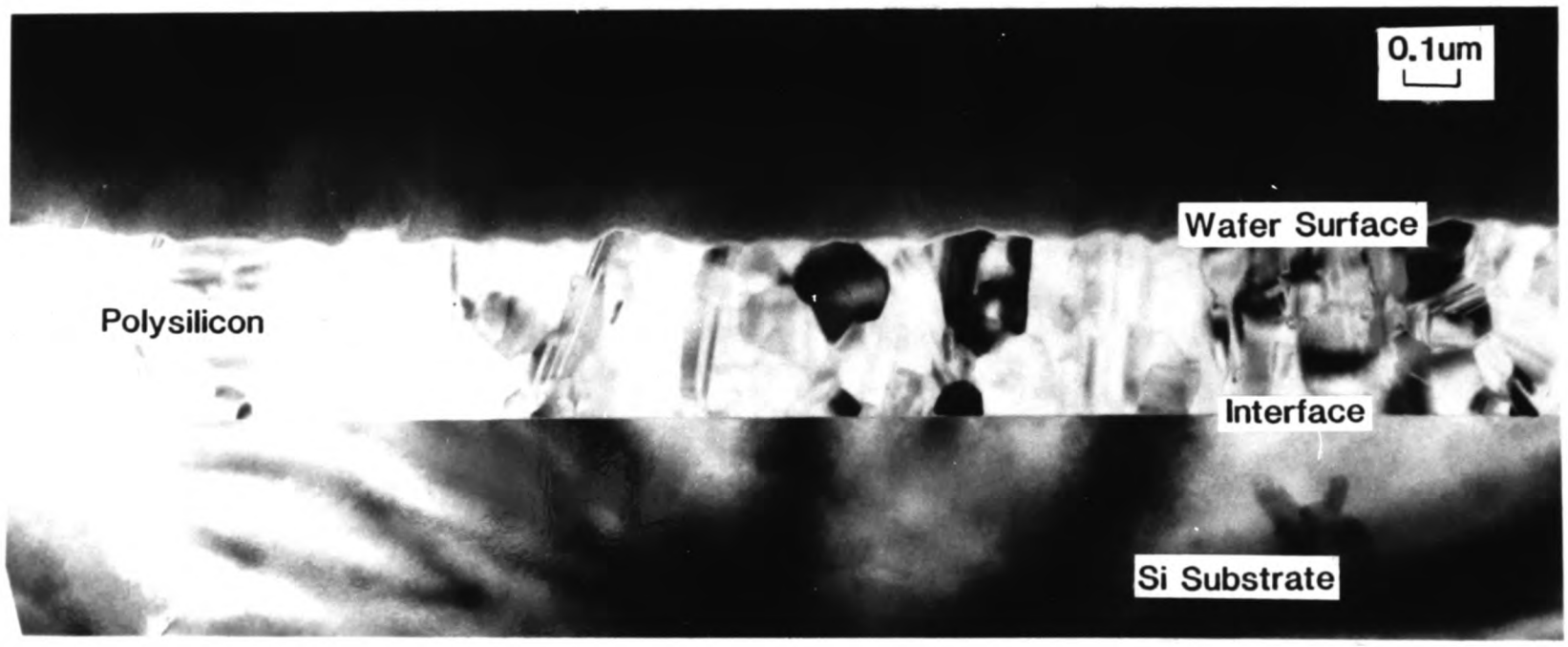


Fig.4.2 RCA, drive-in XTEM

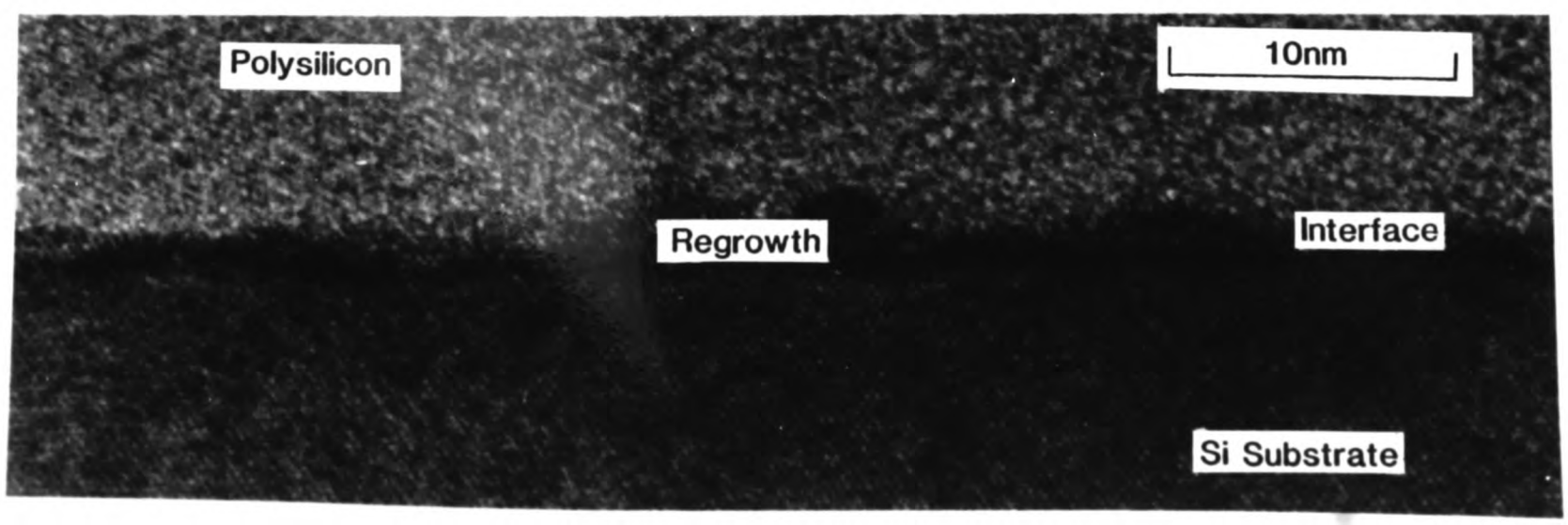


Fig.4.3 HF, 900C XTEM

HREM micrographs of the two control slices which received no pre-anneal, but all other stages of processing, are illustrated in fig.4.1. The RCA cleaned control sample shows a uniform oxide layer 14±4Å thick with no evidence of epitaxial regrowth or oxide break-up. The oxide in the HF cleaned control sample on the other hand is extremely irregular, varying in thickness from approximately 8Å down to zero and shows some evidence that early stages of epitaxial regrowth have begun. The accuracy with which the oxide thickness can be measured is considered later in this chapter and is not dealt with further at this point.

Fig.4.2 shows the RCA control sample. From this it can be seen that the polysilicon region is composed of equiaxed grains having a mean size of approximately 1000-2000Å although the range of sizes is large, and a few grains span the thickness of the whole layer. As the polysilicon layer was grown by LPCVD at 610°C, where the temperature range for amorphous growth is less than approximately 580°C, it might therefore be expected to show a typical, low temperature columnar structure [17]. The equiaxed grains observed are typical of annealed polysilicon [18] and are a consequence of the later stages of thermal processing, i.e. the emitter drive-in. This structure is an important consideration in view of some results presented later to explain the behaviour of minority carriers in the single crystal emitter.

The HF sample annealed at 900°C already exhibits considerable interfacial roughness (fig.4.3) and indeed small areas of epitaxial regrowth can be observed where the (111) lattice fringes are seen to cross through the interface region. These areas of regrowth penetrate approximately 17Å into the polysilicon region with a maximum lateral spread of 100Å. At this stage, the oxide has only thickened slightly to approximately 12Å and is becoming more uniform in thickness in the regions not pierced by epitaxial regrowth. An approximate measure

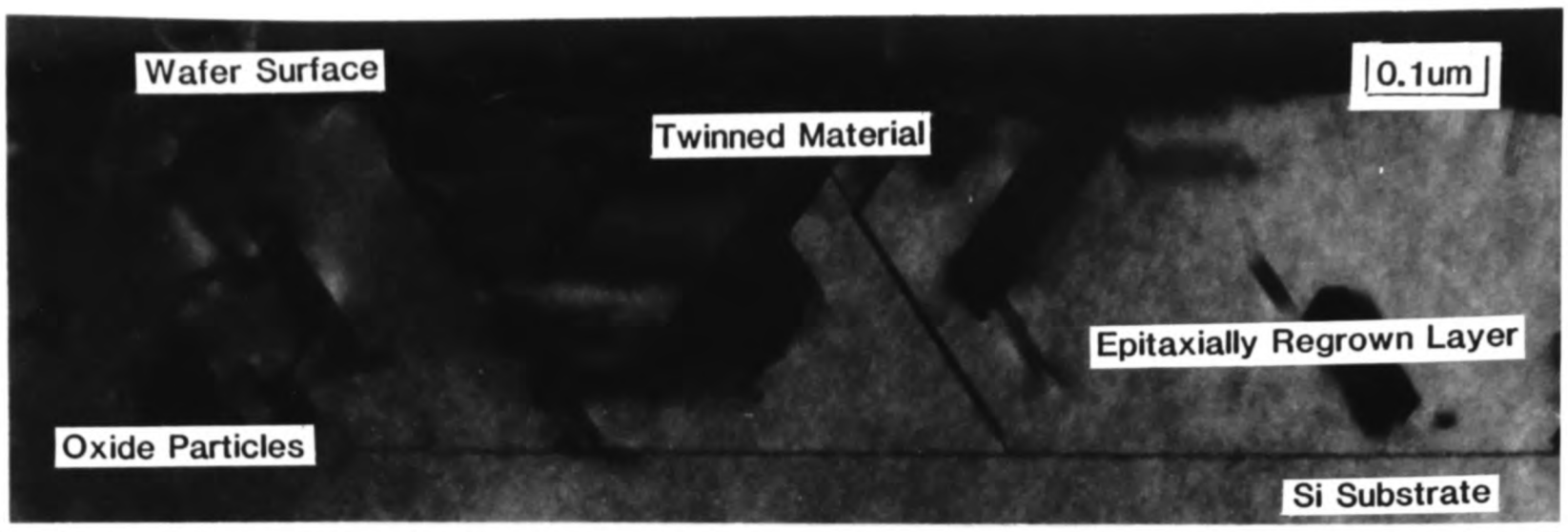


Fig.4.4a HF, 950C XTEM

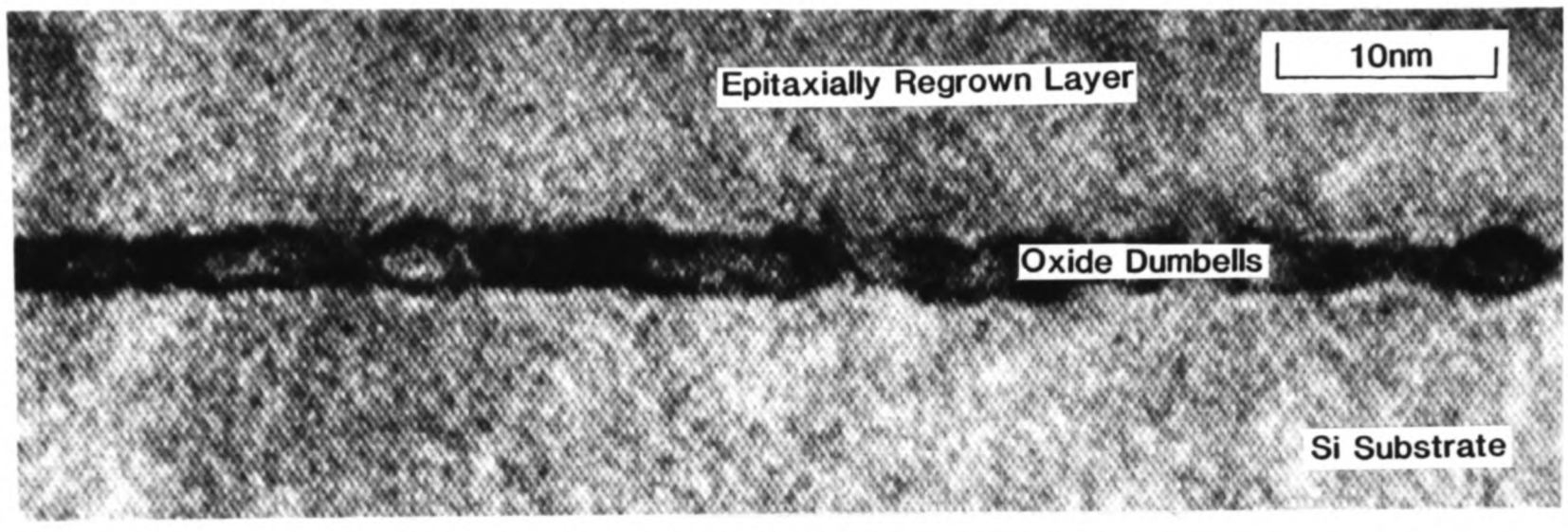


Fig.4.4b HF, 950C XTEM

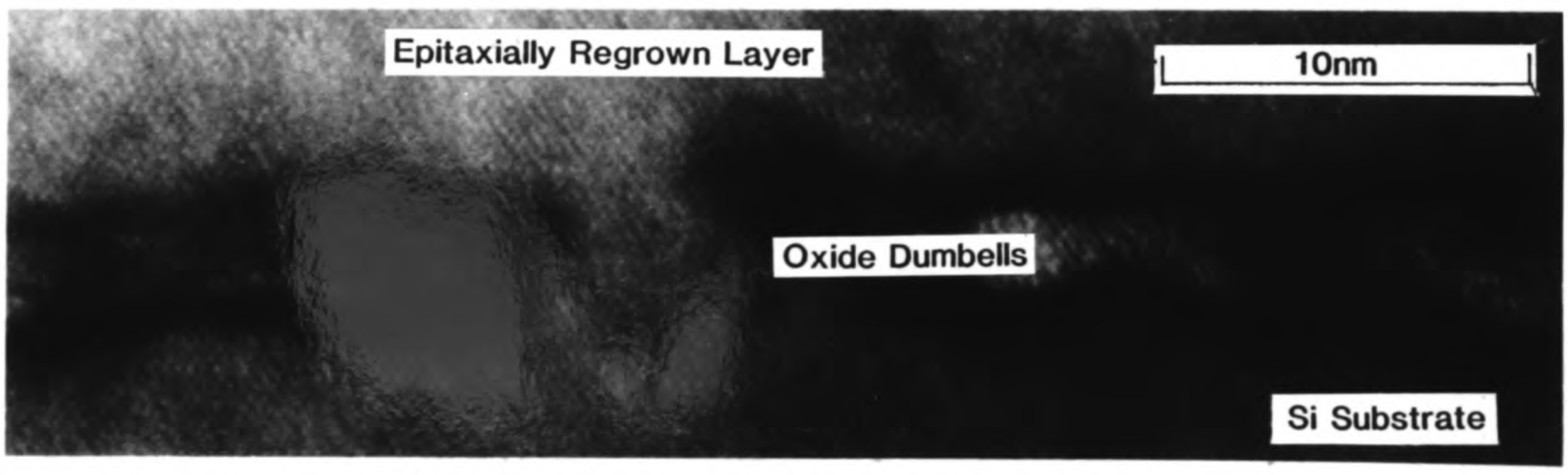


Fig.4.5a HF1000C XTEM

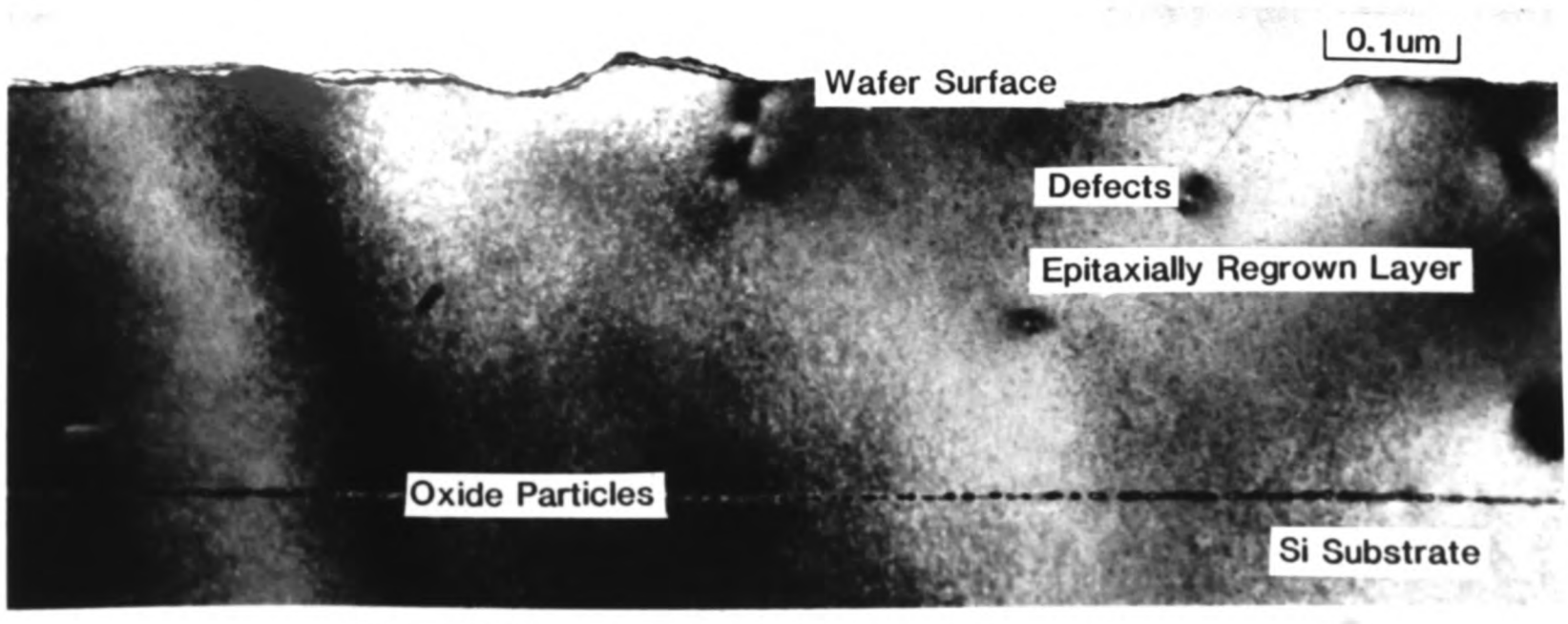


Fig.4.5b HF, 1000C XTEM

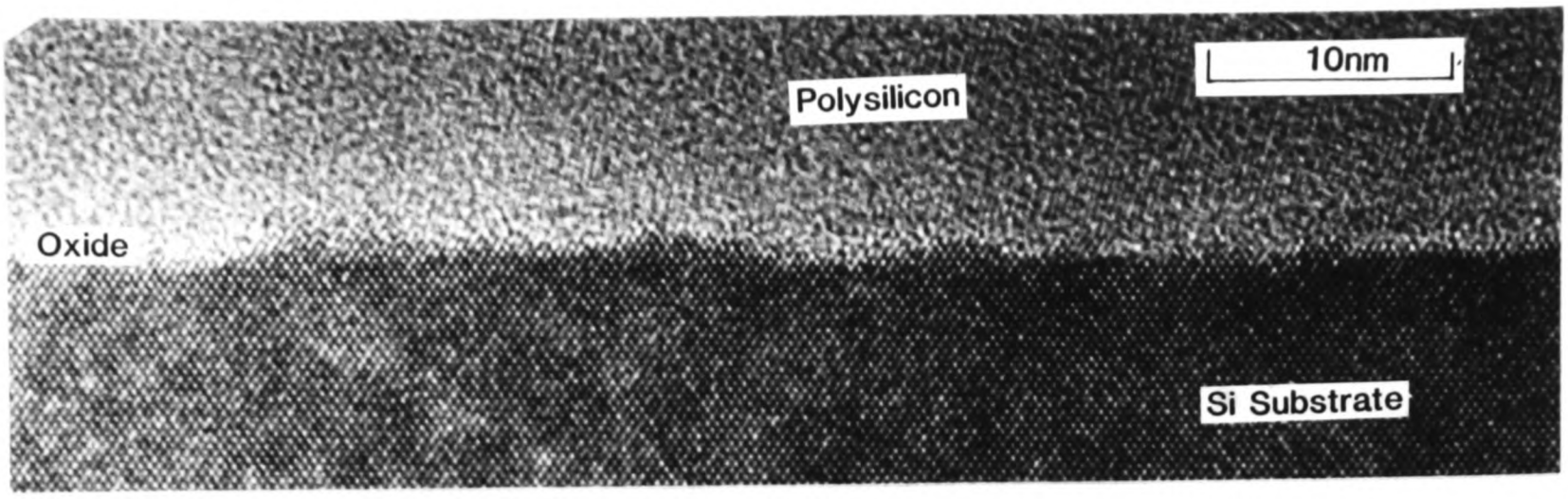


Fig.4.6a RCA, 900C XTEM

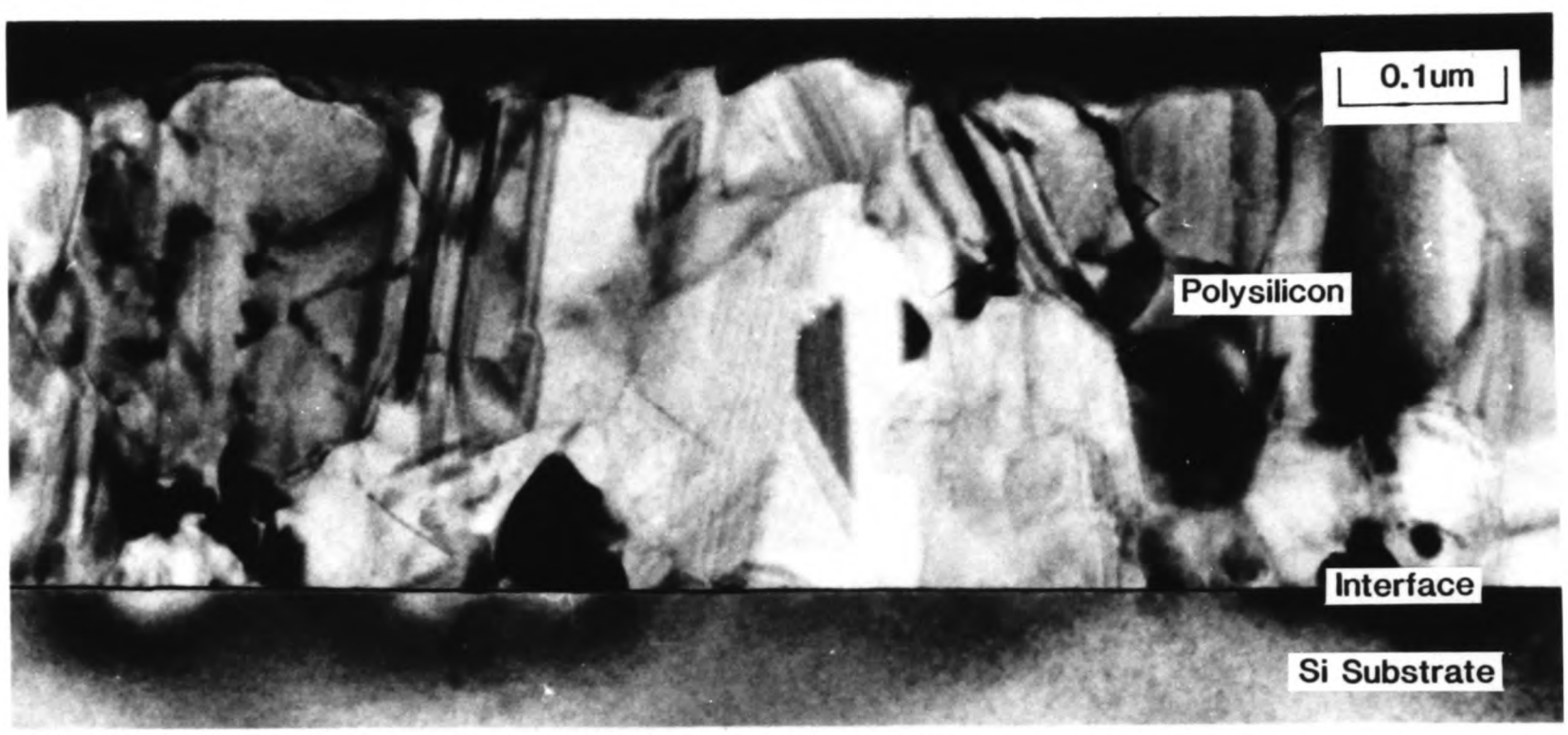


Fig.4.6b RCA, 900C XTEM

Fig.4.7a RCA, 1000C XTEM

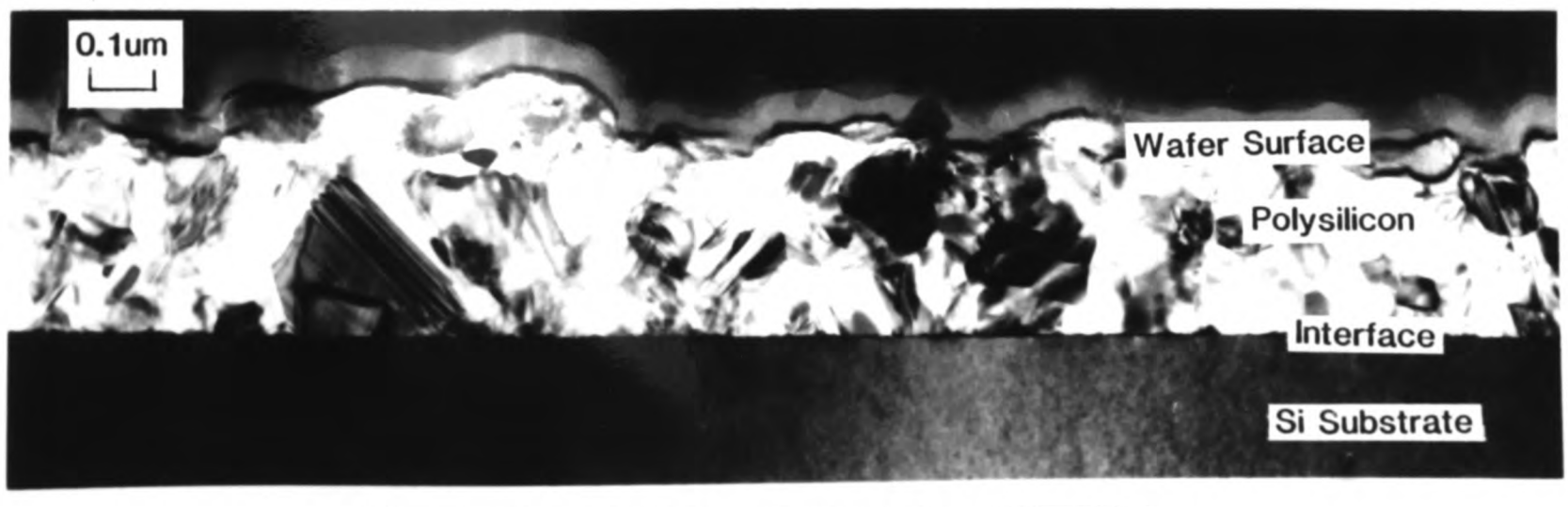


Fig.4.7b RCA, 1000C XTEM

of the percentage of interface uncovered may be obtained by taking the percentage of measured regrowth linearly along the micrograph and squaring this figure. For this sample a figure of approximately 16% was calculated.

In the 950°C HF cleaned sample, the polysilicon layer has epitaxially regrown to produce a region containing a variety of growth induced defects including dislocation loops and twins (fig.4.4a). The oxide layer has thickened to form both balls and extended 'dumb-bell' shapes having a thickness of ≈25-30Å (fig.4.4b). The wafer surface above the regrown region is seen to be rough with an amplitude of ≈150Å. These surface contours are believed to be closely associated with the original, irregular surface shape of the polysilicon layer, which results from the peaks of the surface grains extending locally beyond the bulk.

Examination of the HF sample pre-annealed at 1000°C confirms this trend. This polysilicon layer has also completely regrown and the balled up oxide particles have formed a discontinuous planar layer of thickness 38-45Å at approximately the depth of the original interface (fig.4.5a). Examination of the epitaxially regrown layer clearly shows that the twinned material associated with this region in the 950°C sample does not occur, and the layer consists mainly of unfaulted material (fig.4.5b). The density of crystalline defects (both dislocations and loops) is substantially lower.

The RCA samples by comparison follow a similar pattern, but each stage of the regrowth occurs at a higher pre-anneal temperature. In the sample pre-annealed at 900°C, HREM images (fig.4.6a), show surface roughening of the substrate at the interface of ≈9Å with a maximum lateral spread of ≈70Å. The area of interface uncovered, measured using linear measurements from the cross-sectional image, as for the HF 900°C sample, is approximately 6%. There is a slight increase in average grain size (fig.4.6b) over the control sample (no pre-anneal), to approximately 1500Å. Owing to the large spread in observed grain-size, it is however difficult to measure this accurately.

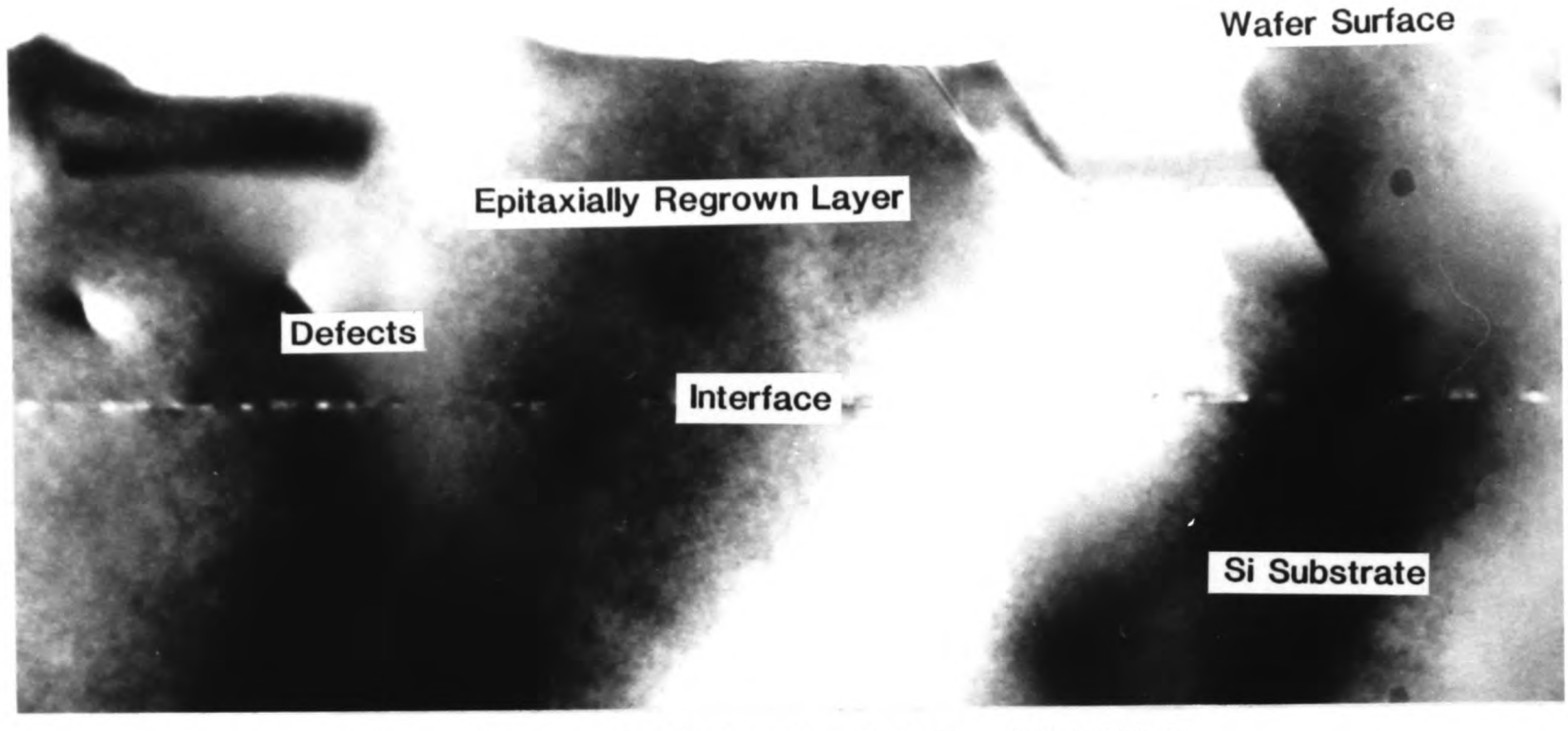


Fig.4.8a RCA, 1100C XTEM

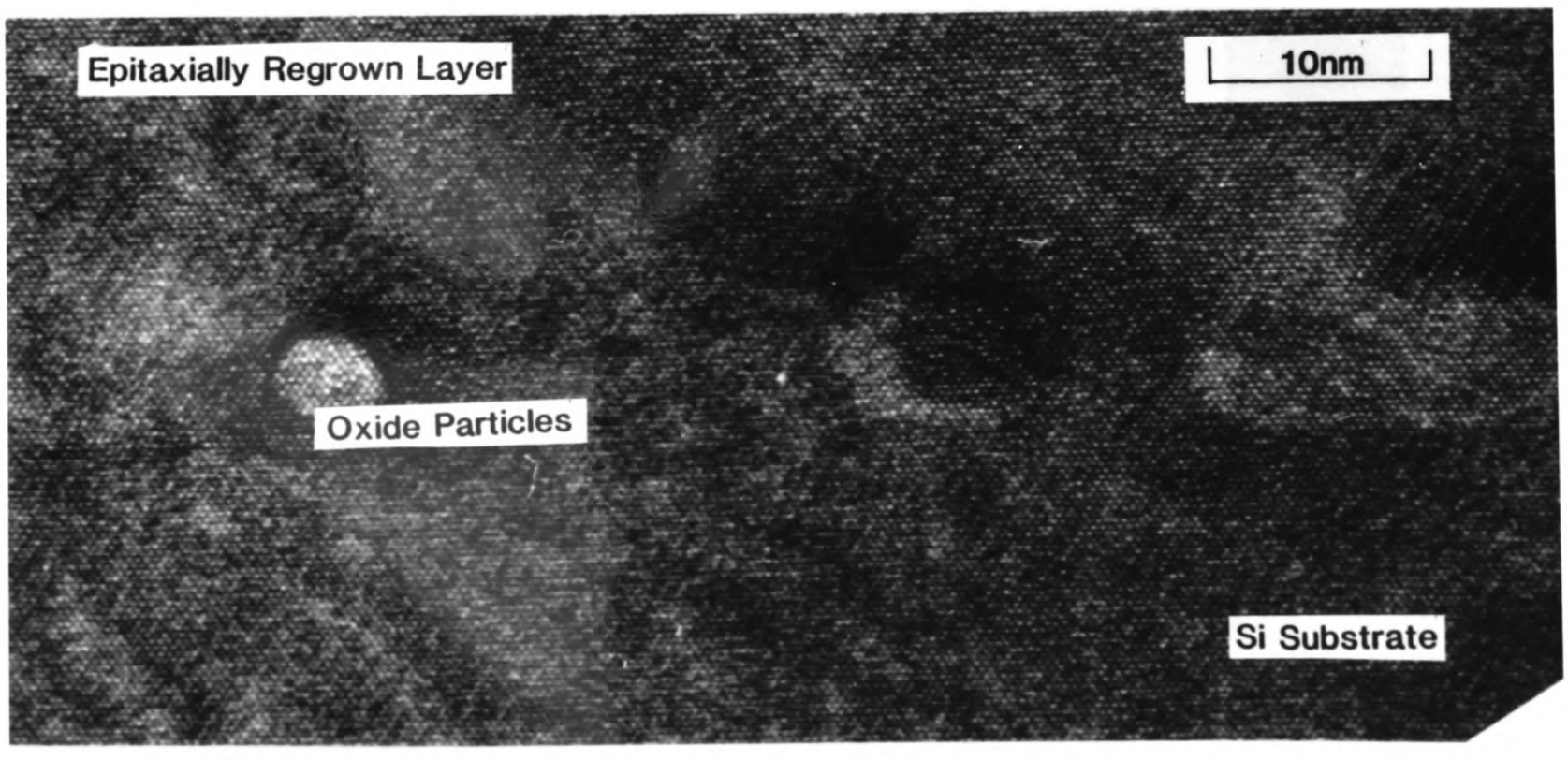


Fig.4.8b RCA, 1100C XTEM

Sample	Oxide	Regrowth	Grain size
RCA, 900C	continuous(14A)	no	1200A
RCA, 950C	broken (19A)	yes(35A)	1200A
RCA, 1000C	broken (22A)	yes(40A)	1200A
RCA, 1100C	balled-up (55A)	yes(170A)	2000A
HF, control	broken (8A)	yes(25A)	1200A
HF, 1000C	broken (50A)	yes(150A)	1200A

TEM results from arsenic doped samples (after Jorgensen)

Table 4.2

It is not until a pre-anneal of 1000°C that the onset of epitaxial regrowth can be clearly observed in the RCA samples (fig.4.7a). In the HREM micrographs of this sample the regions of regrowth can be seen to pierce the interfacial region by up to 100Å with a maximum lateral spread of 300Å. More typically areas of regrowth of ≈60Å by ≈200Å are observed. Any change in the grain size of the layer which epitaxy has not yet consumed, is similar to that observed in the 900°C case (fig.4.7b).

It is necessary to examine the RCA clean sample pre-annealed at 1100°C before complete epitaxial regrowth of the polysilicon layer is observed. Diffraction contrast (fig.4.8a) reveals that the regrown layer contains a number of dislocations and a low density of twins compared with the HF clean 950°C sample. In this respect the twin density is similar to the HF 1000°C sample. The oxide has formed balls and dumb-bells of approximately 52-58Å thickness (fig.4.8b). These balls show evidence of facetting.

The results of similar work carried out by Norman Jorgensen on arsenic implanted samples are included for comparison in table 4.2. These results follow broadly similar trends, in that the HF cleaned samples break up and regrow more rapidly than the RCA cleaned samples. However there is a definite difference in the degree of regrowth between the As implanted samples and the P implanted samples.

For the P^+ doped, HF clean sample, a pre-anneal at 950°C is sufficient to cause complete regrowth of the polysilicon layer. For the As^+ doped samples, a pre-anneal of 1000°C is still insufficient to produce a similar degree of regrowth. The RCA cleaned, As^+ doped sample after the 1100°C pre-anneal has still not completely regrown and comparison with results from the previous chapter for the RCA cleaned sample pre-annealed at 1100°C with no further processing shows the two structures to be similar. This suggests that the As at this doping level has had no enhancing effect at all on epitaxial regrowth. This confirms

the observations made in the previous chapter concerning the relative effects on structure of the presence of these two dopants at the given concentrations.

To try and relate the structural data to the electrical characteristics of real devices in the present work, two sets of samples were fabricated simultaneously. One set was blanket samples, that is, fully processed wafers but with no lithographic steps. This results in a processed wafer containing a base which has been driven in, and the polysilicon layer deposited. From this step on, the wafer received the stated pre-anneals, implants and drive-ins but was not patterned for metallization. These blanket samples were generated and used for the above TEM analysis and the simultaneous processing is designed to reproduce as closely as possible the morpholgies generated in the real devices which constitute the other, fully processed set, fabricated for electrical characterisation at Southampton.

There are however two potentially significant areas of variance between the two sets of samples. Firstly, and potentially most important, are the local stresses generated in real devices and the effect these may have on the morphology of the emitter region. The devices have an active emitter area of $7 \times 7\mu m^2$. Some results are presented in chapter 5 which suggest that significant stress does exist in real devices. Secondly, the effect of metallization on the regrowth characteristics must be considered,as the blanket wafers received no metallization. However, the effects of the relatively low temperatures associated with metallization are not expected to significantly affect the device structure.

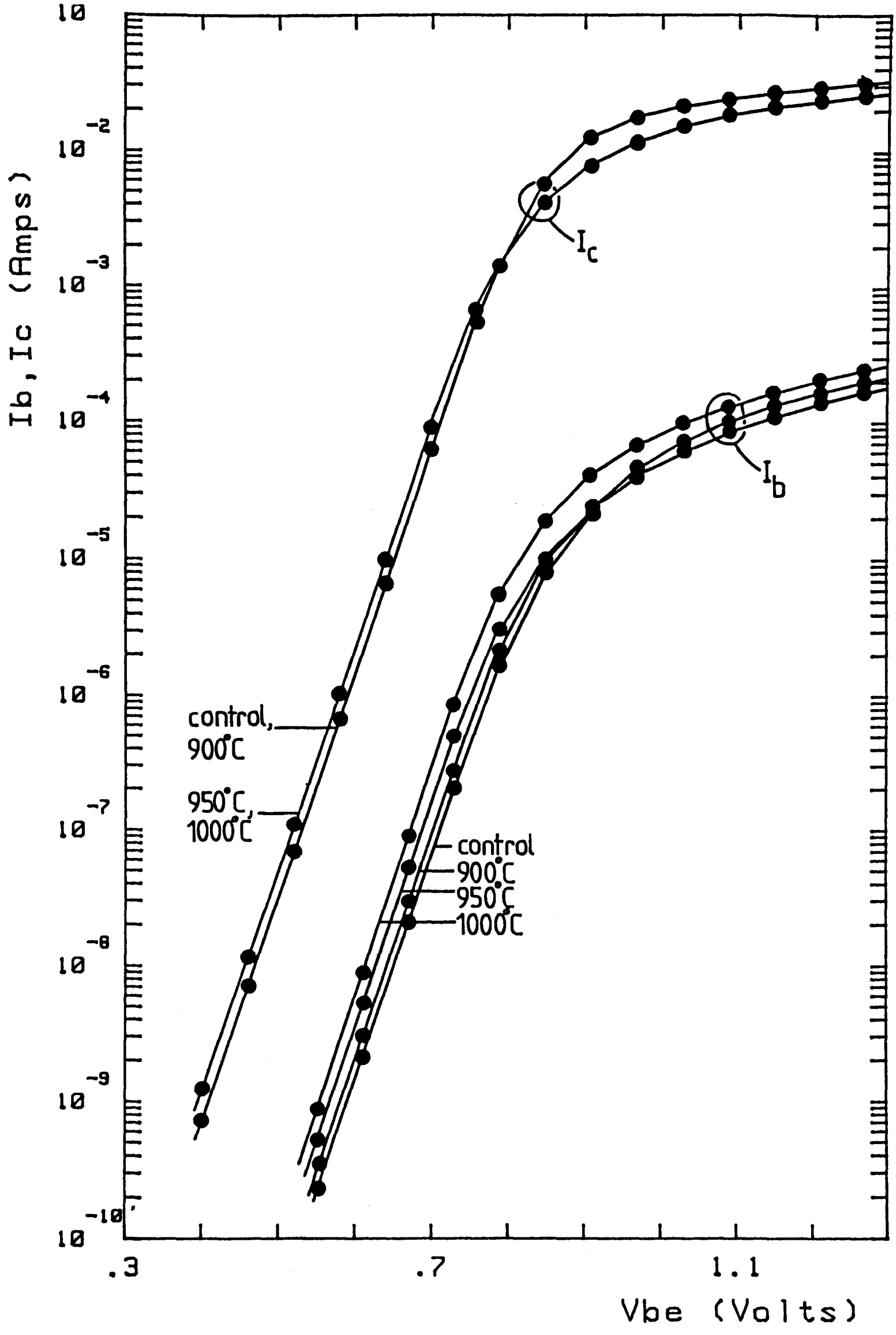


Fig 4.9a Gummel plots of phosphorus doped RCA clean polysilicon emitter transistors given various pre-anneals from 900C to 1000C

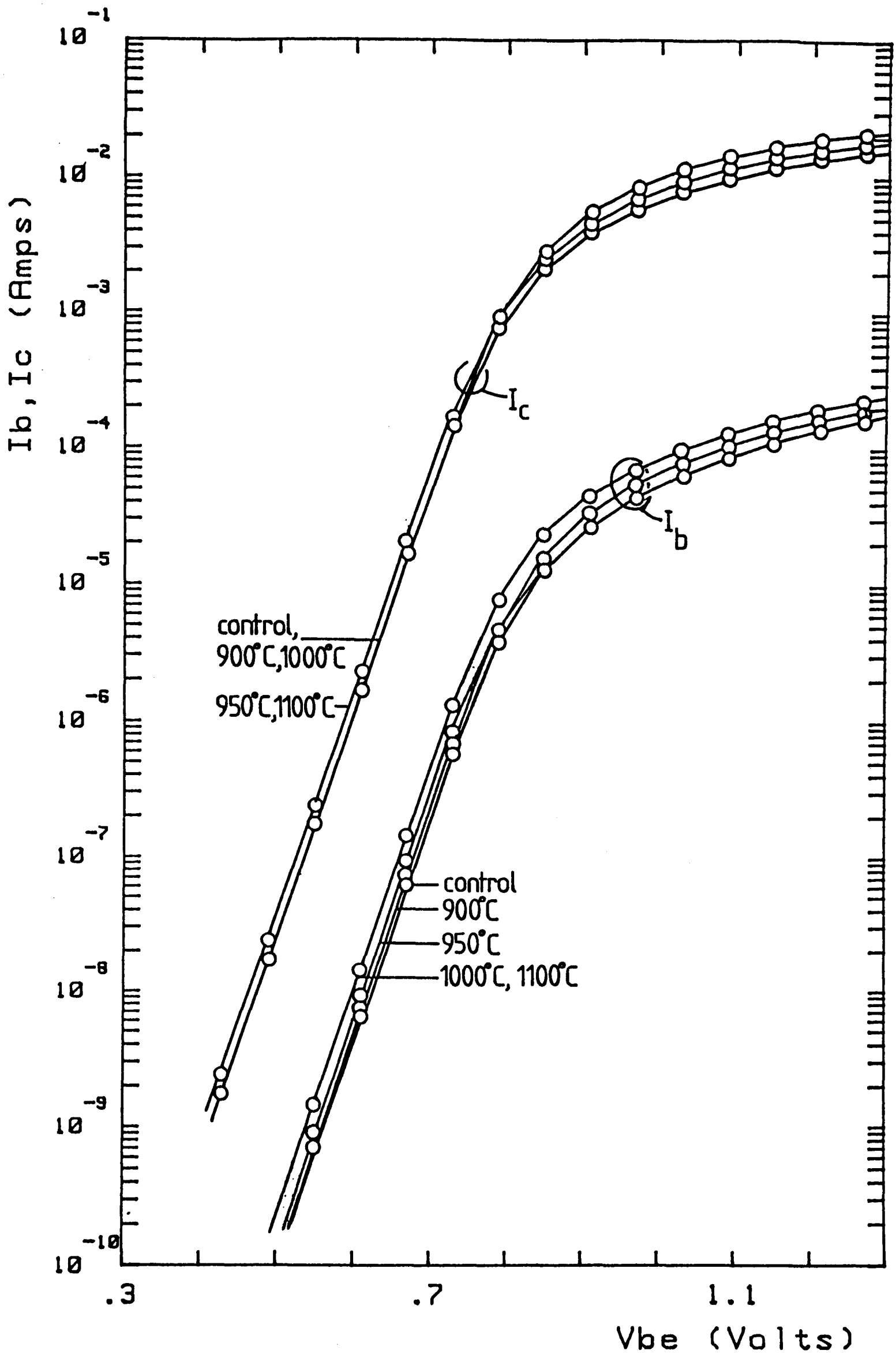


Fig 4.9b Gummel plots of phosphorus doped HF clean polysilicon emitter transistors given various pre-anneals from 900C to 1100C

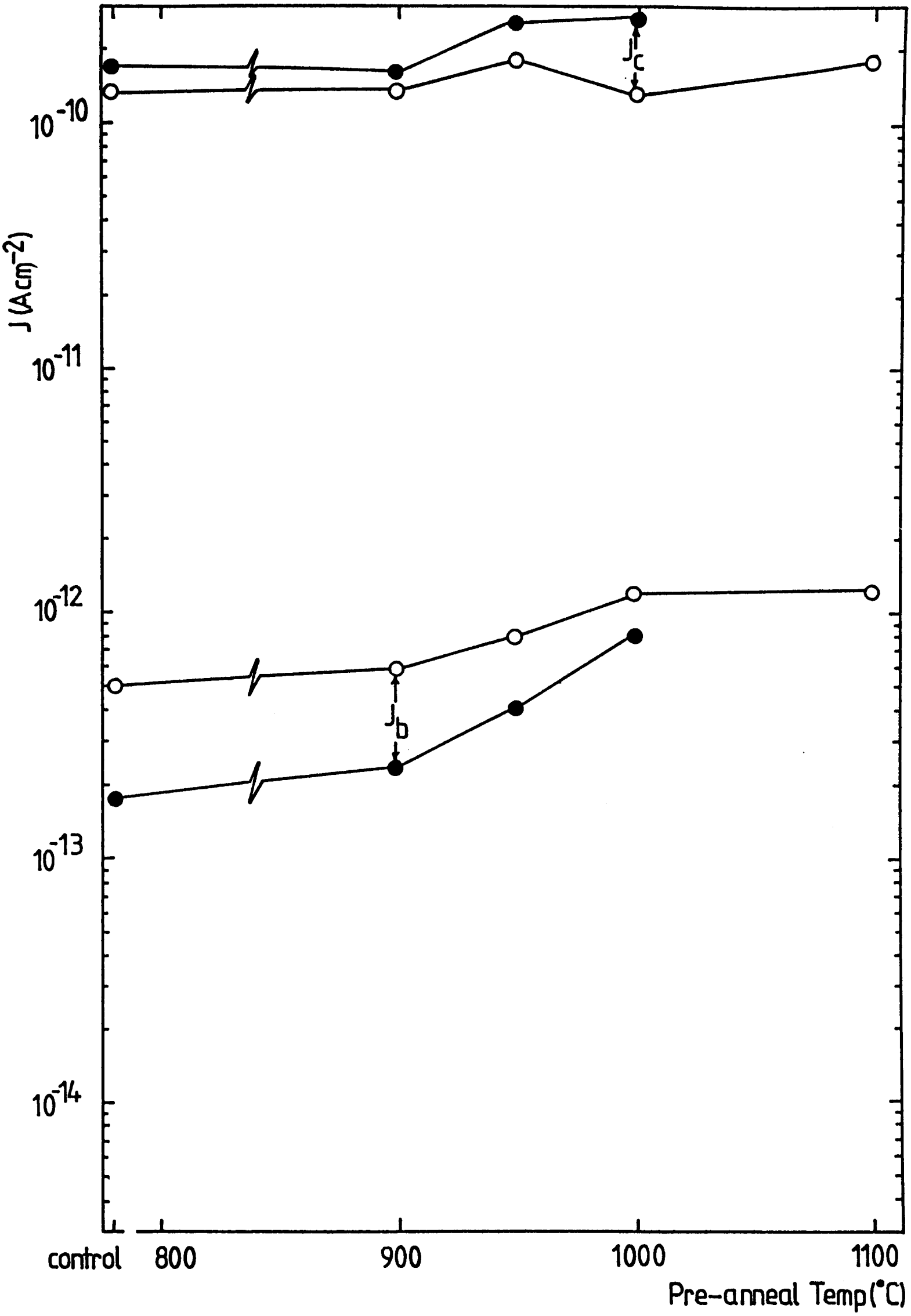


Fig.4.10 Saturation base and collector current densities versus pre-anneal temperature for phosphorus RCA(●) and HF(O) polysilicon emitter transistors

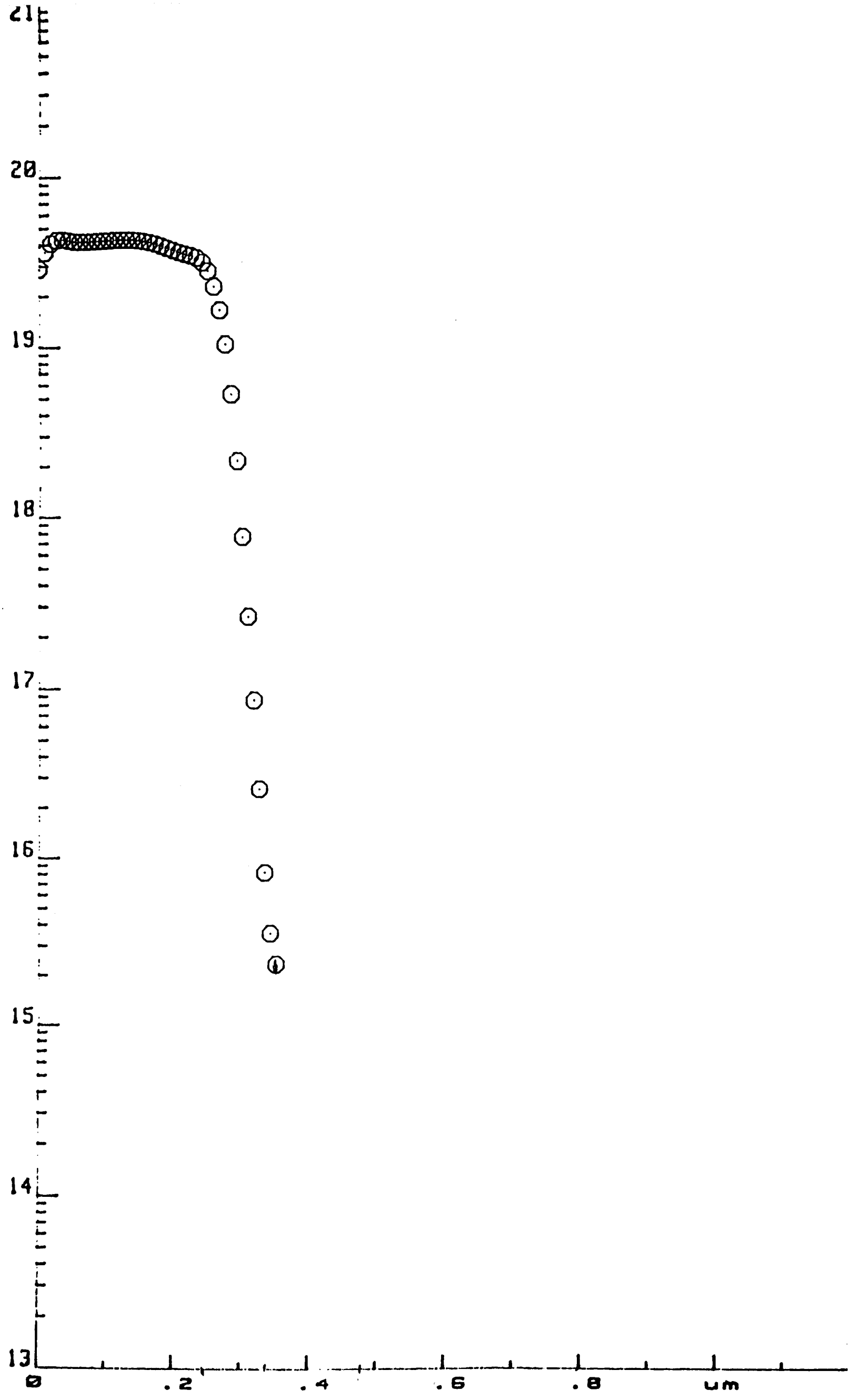


Fig.4.11 Spreading resistance of emitter region in HF clean no pre-anneal phosphorus doped device

4.2.2 ELECTRICAL RESULTS

Fig.4.9a and 4.9b shows a series of Gummel plots for the pre-annealed samples. A Gummel plot shows the values of $\ln i_B$ (base current) and $\ln i_C$ (collector current) vs. linear V_{BE} (base/emitter voltage) [19]. From this, β can be calculated from the ratio of I_c to I_b for any given V_{BE}.

As the pre-anneals progress, it can be seen that for any given value of V_{BE}, i_C remains relatively constant, whilst i_B is progressively greater as the pre-anneal temperature increases. This is more clearly revealed in the saturation current densities vs. pre-anneal temperature curves of fig.4.10. The enhanced gain of the samples before the oxide begins to break-up can be seen as a factor of 2.5 difference in i_B for the HF dip samples across the anneal range and a factor of ten for the RCA samples. The plot shows very little difference in base current between the HF 1000°C sample and the HF 1100°C sample, which leads us to the conclusion that a stable state with regard to the morphology controlling the enhanced gain has been reached. It might also be expected that the RCA sample would reach a similar value of i_B after a sufficiently high temperature or sufficiently long pre-anneal.

4.2.3 DOPANT PROFILES

One dimensional doping profiles of the emitter regions were obtained from spreading resistance measurements of the blanket wafers (see chapter 6). These results correspond to the electrically active dopant. Fig.4.11 illustrates the profile for the HF control sample with no pre-anneal. The profile consists of a flat region doped to approximately $5\times10^{19}\text{cm}^{-3}$. This region is approximately 0.3μm across. This is followed by a steep decrease over the next 0.1μm to a doping concentration of $1\times10^{16}\text{cm}^{-3}$, beyond this point is the base, which is not shown here for clarity. The flat region corresponds to the polysilicon layer

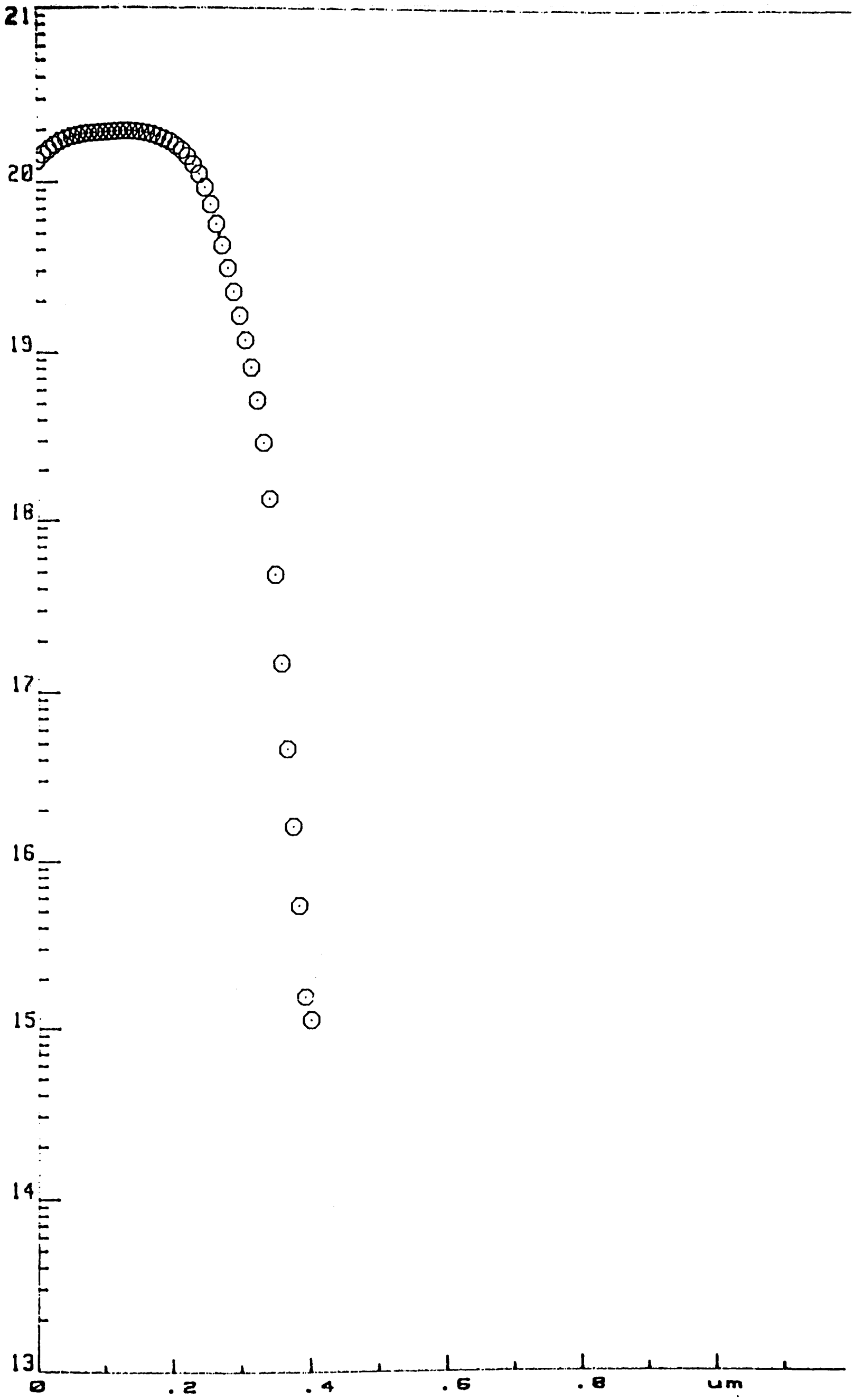


Fig.4.12 Spreading resistance of emitter region in HF clean 1000C pre-anneal phosphorus doped device

and appears uniformly doped due to the rapid diffusion of the dopant atoms down grain boundaries during the emitter drive-in. This layer then acts as a diffusion source into the bulk silicon. Because the diffusion rate is much lower in bulk silicon, the profile decreases rapidly in the bulk material. It is important to clarify that the emitter itself is generally defined as the region from the e/b junction to the monosilicon/polysilicon interface. The polysilicon itself is defined as a contacting material.

Fig 4.12 corresponds to the HF 1000°C pre-anneal sample. This is a sample in which complete epitaxial regrowth has occurred (fig.4.5b). The shape of this profile is similar to that of fig.4.11, but there are important differences. It is clear from the TEM results that this sample has epitaxially regrown, but has the SIMS profile of a polysilicon contacted sample. This occurs because most of the epitaxial regrowth does not occur during the pre-anneal. Thus the subsequent implant takes place into polysilicon and diffuses rapidly down the grain boundaries to produce the typical flat profile. The polysilicon then regrows rapidly due to P enhancement of the regrowth as detailed in chapter 3. This occurs because the pre-anneal breaks up the interface and initiates the regrowth. Regrowth does not accelerate until after the P has diffused, and so a doping profile typical of a polysilicon contact is obtained in the ultimately single crystal material. The measured peak concentration is $2\times10^{20}\text{cm}^{-3}$, and the width of the highly doped region (based on $5\times10^{19}\text{cm}^{-3}$ as the end point) is 0.35μm. Both of these values are larger than the corresponding values for the HF control sample.

It is well known that dopant species can segregate to grain boundaries[6]. As the regrowth proceeds, the dopant atoms are left behind by the regrowth front, which is known as breakaway. Breakaway only occurs when the speed of the regrowth front exceeds the rate of diffusion of the dopant. This can lead to a sudden increase in the speed of movement of the regrowth front as the dopant atoms may have a pinning effect on the boundary. These dopant atoms must be accommodated in substitutional sites in the lattice in which they become

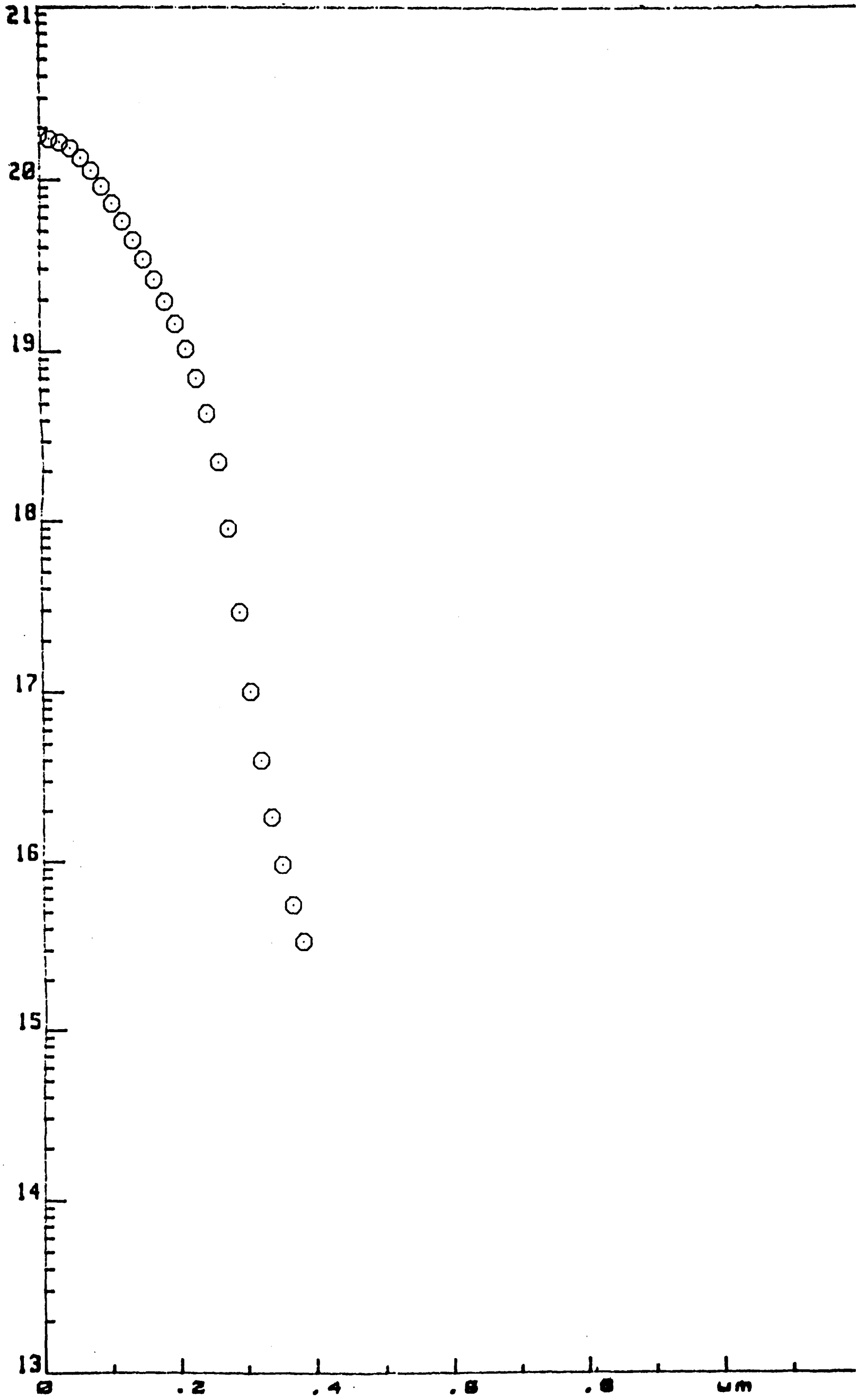


Fig.4.14 Spreading resistance of emitter region in RCA clean 1100C pre-anneal phosphorus doped device

electrically active. Thus the concentration of electrically active dopant increases, as is observed. This could be confirmed by comparison with the corresponding SIMS profile which should remain largely unchanged. SIMS profiles were not available for this sample however. It can be concluded therefore that this sample was polysilicon after the pre-anneal and epitaxially regrew during the subsequent drive-in. This again confirms the TEM observations of chapter 3 in which the HF sample pre-annealed at 1000°C (with no subsequent processing) was found to contain only small areas of regrowth.

Fig.4.14 is the spreading resistance doping profile for the RCA 1100°C sample. The profile of this sample is significantly different from the two previous examples in that it no longer has a flat initial profile but starts to decrease immediately, but at a much shallower gradient. This is typical of a single crystal implanted sample, and shows none of the enhanced diffusion typical of polycrystalline material. Thus it may be concluded that the polysilicon layer in this sample undergoes complete epitaxial regrowth during the pre-anneal. Once again this largely confirms the observations of chapter 3, in which the layer was observed to be almost completely regrown.

4.2.4 OXIDE BEHAVIOUR

Two TEM techniques are used to examine the oxide in this investigation. Cross-sectional HREM, detailed above, and PVTEM. The latter is utilised to examine the shape of the balled up oxide and, in conjunction with image processing, to measure the area of the interface with remaining oxide layer coverage. The morphology of the balled up oxide is discussed in chapter 3 on the basis of these results. In previous work, calculation of this area was performed by measurement from cross-sectional HREM micrographs [20]. This cannot be considered an entirely satisfactory method for producing this data. Firstly, the

Sample	% free area
Phosphorus, HF 950C	70.9
Phosphorus, HF, 1000C	80
Phosphorus, HF, 1100C	87.6
Phosphorus, RCA, 1000C	49.7
Phosphorus, RCA, 1100C	87.6
Arsenic, HF, 1100C	91.8
Arsenic, RCA, 1100C	85.1

Break-up of Oxide

Table 4.3

TEM foil may be thick compared with the oxide particle thickness, which means that observations might be made through several oxide particles at once. Additionally, use of XTEM allows an examination of only a very small area of the interfacial region. For both these reasons, measurements carried out by this technique were considered unsuitable.

On the other hand, PVTEM allows the plane of the interface to be examined over an area of up to $100\mu m^2$. Preparation of these foils was discussed in chapter 2. The object was to produce quantitative data concerning the interface, specifically the exposed emitter window area. This information was extracted from the micrographs using a commercial image processing package in the manner described in chapter 2. The results are given in table 4.3.

This technique is only suitable for examining the behaviour of the interface when the balling up of the oxide has reached a fairly advanced stage, certainly at least 50%. This and the accuracy of the technique were discussed in chapter 2. For this reason the method is unsuitable for examination of the pre-annealed As samples as examined by Norman Jorgensen [21] except at the highest temperatures. Results measured from the As implanted HF and RCA samples pre-annealed at 1100°C using the same technique are included in the table for comparison.

4.3 DISCUSSION

4.3.1 THE INTERFACIAL OXIDE

The 'native' oxides formed by HF cleans can vary significantly in thickness depending on the precise details of the cleaning process. Variations in the maximum thickness of the HF oxide layers from 4Å to 8Å have been reported[21]. The nature of the interfacial layer is complicated, and was discussed in detail in chapter 3. Certain points however are worth restating. Even assuming that the interfacial layer contains only oxygen and silicon, there is a strong body of evidence that the stoichiometry of thin oxides of different thicknesses may vary significantly. Additionally, there is the question, as yet unresolved, as to whether an ordered region of oxide or sub-oxide exists on the substrate surface, as described by Ross and Stobbs [22] and by Ourmazd [23]. The presence of a sub-oxide and/or an ordered region of oxide, may have a significant effect on the electrical behaviour of the oxides, specifically the tunnelling barrier height. With approximately one or two monolayers of sub-oxide, the difference between the HF oxide and the RCA oxide, may be considerable.

Jorgensen[20] observed that modifications in the detailed processing conditions used in the HF clean process resulted in a change in the oxide thickness (both by HREM and by oxide balling measurements). This type of variation illustrates the potential for the production of different interfacial structures by different workers and may well be one of the reasons for the confusion in the literature over the exact nature of the interfacial region and the differing electrical properties obtained by different workers on nominally similar samples.

4.3.2 GAIN ENHANCEMENT

The most important consideration when investigating the mechanisms of 'gain enhancement' in polysilicon contacted devices is that it is caused by supression of minority carrier current flow in the emitter, i.e. hole transport in the n-type emitter region in all reported cases[2][3][4]. The injected minority current density from the base into the emitter, J_p in the one dimensional case, is given by the diffusion expression [24];

$$J_p = -qD_{pe}\left(\frac{dp_e}{dx}\right)$$

where q is the charge on an electron, D_{pe} is the diffusion coefficient of holes in the emitter and p_e is the hole concentration in the emitter. x has its usual meaning of distance, in this case away from the emitter side of the emitter/base depletion region, such that at this position, x=0. This expression is simplified (it ignores bulk recombination and the built in retarding field due to the impurity profile gradient) but illustrates the general effect. Thus, assuming D_{pe} is constant, the hole current is controlled to a first approximation by the minority carrier concentration gradient at the emitter side of the emitter/base space charge region. To produce enhanced gain over an entirely, single crystal emitter by modifying the level of the injected minority carriers in the emitter, it is necessary to reduce $\frac{dp_e}{dx}$ at the emitter edge of the depletion region. Assuming a single crystal region in the first part of the emitter, supression of the minority carrier current may be acheived by a discontinuous change in minority carrier concentration close to the emitter/base junction, where "close" means that the distance is short compared with the minority carrier diffusion length in single crystal silicon. This is the means by which $\frac{dp_e}{dx}$ can be reduced, in the simplest case of no bulk recombination to;

$$\frac{dp_e}{dx} = \frac{pe_{(0)}}{W_E}$$

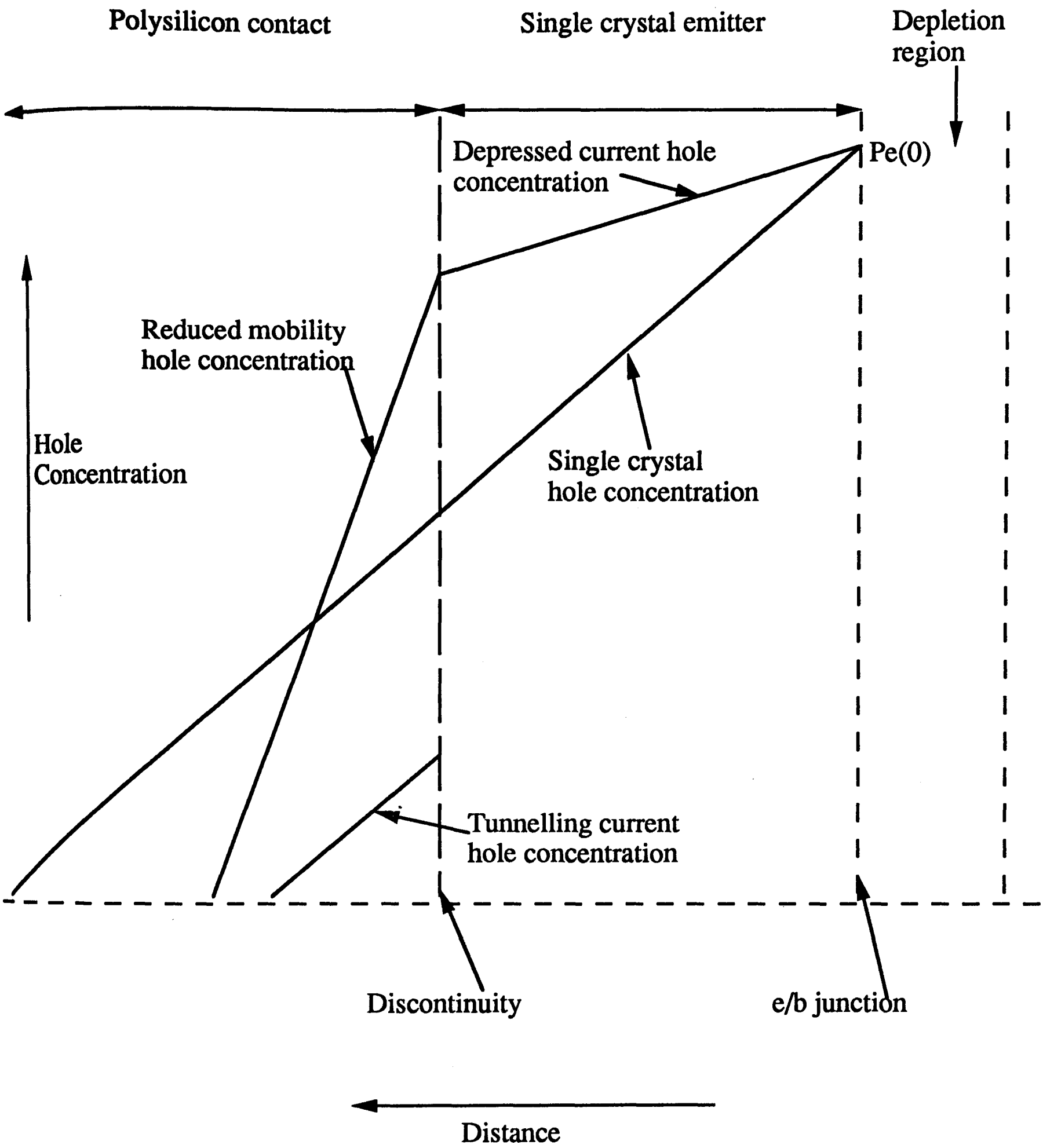


Fig.4.16 Effect on hole current of a discontinuity in the emitter close to the e/b junction

where W_E is the width of the emitter and $pe_{(0)}$ is the hole concentration at the emitter side of the emitter/base junction, which is determined by V_{BE}. The minority carriers injected into the single crystal emitter accumulate under the discontinuity. The minority carrier concentration at the emitter side of the emitter base depletion region, pe_0, is controlled largely by the bias conditions on the device; thus raising the minority carrier concentration within the single crystal region reduces $\frac{dp_e}{dx}$. If the discontinuity were not close to the depletion region, the presence of such a discontinuity would be irrelevant as the excess minority carrier concentration would decay to zero due to the normal recombination processes in the single crystal emitter or at the ohmic contact, and therefore $\frac{dp_e}{dx}$ would be unchanged.

Fig.4.16 shows qualitatively how minority carrier concentration varies with distance in the case of a bipolar device containing such a discontinuity. From this it can be clearly seen that the right hand (monocrystalline) side of this discontinuity should not contain a high density of recombination centres or it would act as a minority carrier sink in a similar manner to a metal contact, i.e. it should have a low surface recombination velocity (S_p). This would result in a high value of $\frac{dp_e}{dx}$. The discontinuity is shown here as being local and planar, although this need not be the case[4]. It must however be "close" to the junction. There are three different models to explain this gain enhancement effect, all three rely implicitly on reducing $\frac{dp_e}{dx}$ at the emitter edge of the emitter/base depletion region.

As described earlier, when gain enhancment was first considered by de Graaff and de Groot, they proposed that the layer of 'native' oxide formed on the single crystal surface during or immediately following the wafer pre-clean before polysilicon deposition, was acting as a tunnelling barrier to carrier transport and that the barrier height to electrons X_e was very small, whilst the barrier height to holes was significantly greater. Ng and Card [25] has indeed shown experimentally that the barrier height to electrons is significantly lower for thin

oxides than for thick oxides. Typically for thick oxides X_e is of the order of 4eV. They have shown that for thin oxides this may decrease to as little as 0.4eV. Ng and Card have also shown that the barrier height of holes (X_h) in thin oxides is significantly greater than X_e and is of the order of 1eV. This variation in the electrical properties of silicon oxide with thickness must be considered in terms of the changing nature of the oxide as discussed in the chapter 3. The hole tunnelling current in such a device J_{PT} has been modelled by Yu and Dutton [8] as;

$$J_{PT} = q\left(\frac{kT}{2\pi m_h^*}\right)^{\frac{1}{2}} \frac{exp - b_h}{(1 - c_h kT)}(p_{IR} - p_{IL})$$

where

$$b_h = \frac{4\pi\delta}{h}(2m_h^* X_h)^{\frac{1}{2}}$$

and

$$c_h = \frac{2\pi\delta}{h}\left(\frac{2m_h^*}{X_h}\right)^{\frac{1}{2}}$$

m_h^* is the effective mass of the holes, p_{IR} and p_{IL} are the hole concentrations on the right and left hand side of the discontinuity respectively, based on the convention shown in fig.4.16. δ is the barrier thickness and X_h is the barrier height to holes. Eltoukhy and Roulston[7] have modelled this behaviour including the voltage drop across the oxide (V_{ox});

$$J_{PT} = A_h T^2 e^{\frac{-(E_g - 11)}{kT}} \frac{e^{-b_h}}{1 - c_h kT}\left[exp\left(\frac{qV_R}{kT} + \frac{q\phi_R}{kT} + \frac{qc_h V_{ox}}{2}\right) - exp\left(-\frac{qV_L}{kT} + \frac{q\phi_L}{kT} - \frac{qc_h V_{ox}}{2}\right)\right]$$

where b_h and c_h have the same meaning as above, A_h is the modified Richardson's constant for holes, $11 = (E_{fn} - E_c)$, V_R and V_L are the heights of the potential barriers in the depletion region adjacent to the right and left side of the interfacial layer respectively, for a given voltage.

$$\phi_R(\phi_L) = \frac{E_{fp} - E_{fnR}}{q}\left(\frac{E_{fp} - E_{fnL}}{q}\right)$$

at the right (left) of the interfacial layer. E_{fp} is the quasi-Fermilevel for holes, E_{fnR} is the quasi-fermi level for electrons in the monocrystalline region, E_{fnL}

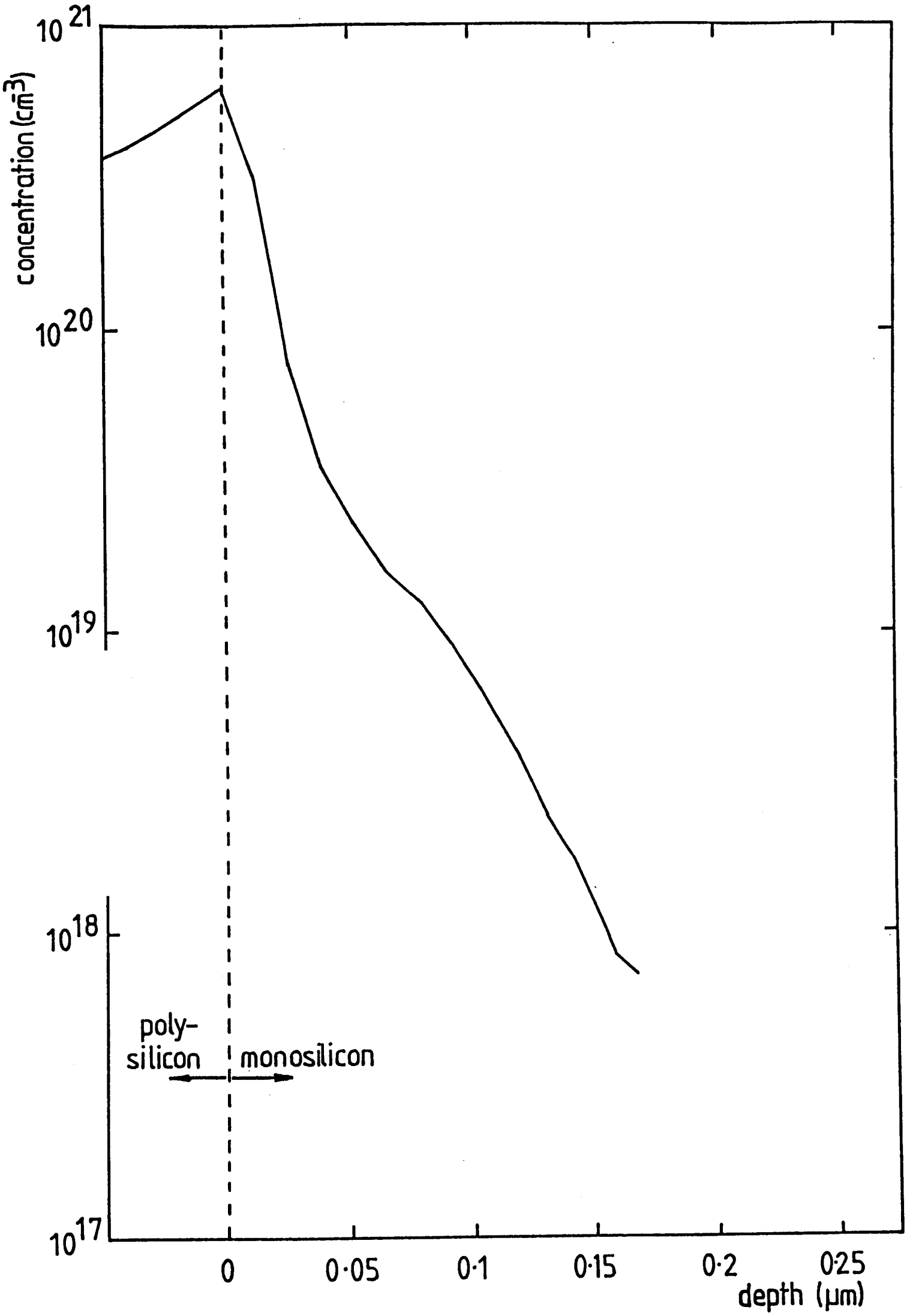


Fig.4.17 Phosphorus SIMS profile of the monocrystalline emitter of an RCA device given an emitt er drive-in at 900C for 10 minutes

is the quasi-Fermi level for electrons in the polycrystalline region. In this type of device it is important to bear in mind that if there is a continuous layer of oxide, there must also be a polysilicon layer. Thus the hole current tunnelling through the oxide will then be affected by the transport properties of the polysilicon. Thus oxide tunnelling can never be viewed in isolation in this type of device with regard to carrier transport.

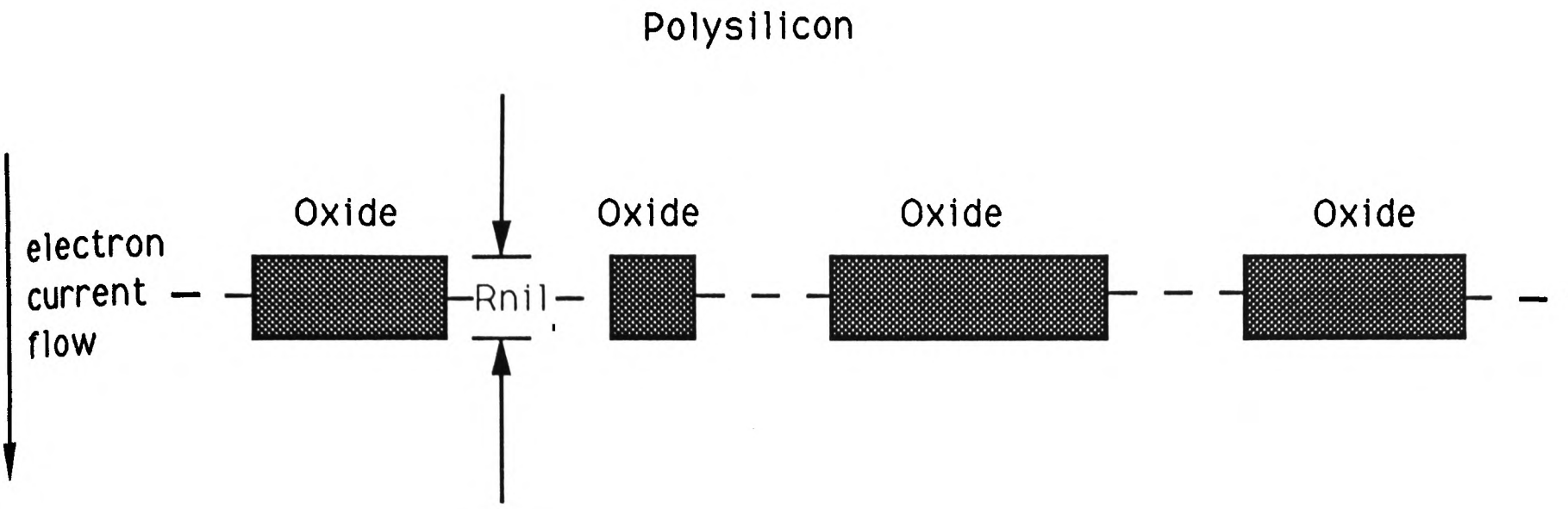
An alternative explanation for the enhanced gain is the presence of a 'high-low' barrier at the interface caused by the segregation of active dopant atoms to the interface[5]. The segregation of As and to a lesser extent P to grain boundaries is well documented. Fig.4.17 shows the SIMS profile of a phosphorus doped sample after a pre-anneal at 900°C, a peak at the depth of the interface (as measured from the TEM results) can be clearly seen. Reference to the spreading resistance profiles in fig.4.11 shows that no corresponding peak in the donor concentration is observed. Measurements of the chemical and active donor concentration of samples by Cuthbertson and Ashburn [26], show a similar peak in the chemical concentration at the interface. The corresponding Hall profiling measurements of donor states show a uniform doping concentration in the polysilicon and no discernable peak in the vicinity of the interface as for the present work. Indeed, if the dopant is located at pentavalent sites at the interface, this would not be surprising. Such a situation would however lead to a reduction in the density of interfacial traps. This strongly suggests that for the dopants investigated (As and P) the segregated dopant is not electrically active. On this basis, it is difficult to lend credence to the concept of a high-low barrier at the interface.

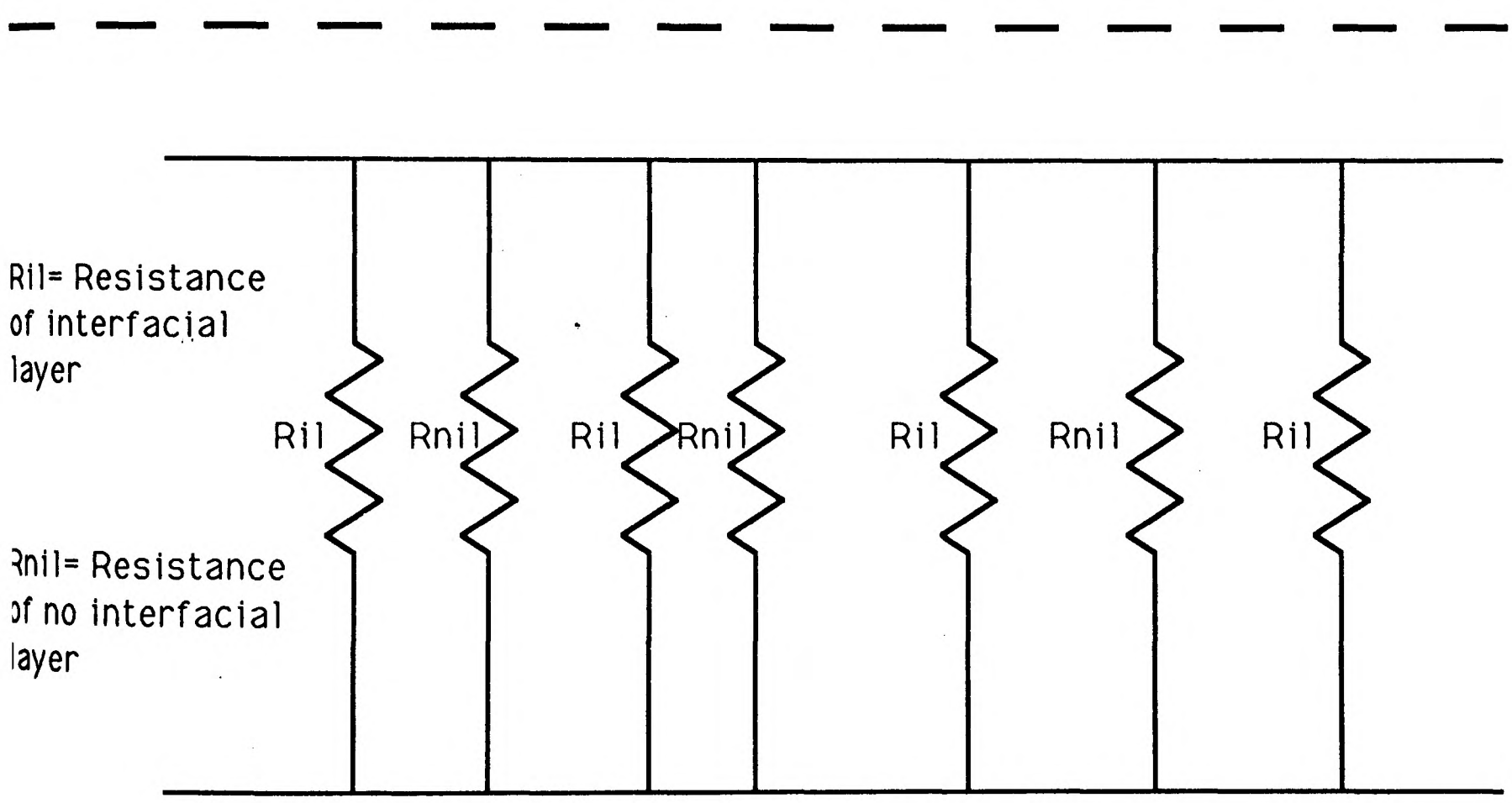
The third explanation for enhanced gain, by Ning and Isaac, proposes that reduced mobility in the polysilicon alters the hole diffusion coefficient D_{pe}. Neugroschel et al[8] have modelled the structure of the polysilicon layer as having a disordered region of material (a large number of small grains) within 200-300Å of the interface. Whilst this structure is typical of as grown LPCVD polysilicon, it

is shown in this work that this grain structure is not observed in fully processed devices. The explanation of the mobility reduction, for example by Mandurah, is given as being due to transport effects at the grain boundaries largely due to the observation that the mobility is further reduced as the grain size decreases. The reduced mobility has been modelled by Mandurah [27], by assuming that the grain boundary is composed of a disordered layer of finite thickness δ. Transport properties in this region are then assumed to be those of amorphous silicon (which has a lower mobility than monocrystalline material). Theoretical work carried out by Paxton and Sutton however [28], suggests that a grain boundary does not contain a wide disordered region, but that it consists of a dislocation array at the boundary and that the dangling bonds at the boundary reconstruct with other bonds at nearby sites. Bourret [29] has carried out HREM observations on some grain boundaries in silicon and does not observe a wide disordered region, but a highly localised boundary. The concept of grain boundary width was an expedient creation for the electrical modelling of grain boundaries. It has little physical significance however and it is not realistic to use it in this fashion.

It is important to realise that the gain enhancement observed by Ning and Isaac is only of the order of $\times 2$ or $\times 3$, whilst de Graaff and de Groot observed a gain enhancement of approximately $\times 10$. In their work, Ning and Isaac have observed that the base current is independent of polysilicon layer thickness down to 1000Å or less. But they record a sharp increase below this thickness. Their measurements are carried out on as grown layers, rather than etched back layers. Other workers[5][9][10][11] observed a similar dependence on polysilicon thickness.

If we examine the polysilicon layer using annular dark field imaging, it is possible to observe the size and shape of individual grains by setting them in a strongly diffracting condition. Fig.4.7(b) illustrates such an image from the RCA 1000°C sample. This shows no apparent difference in the grain structure of the near interface region apart from the small effects of epitaxial regrowth. The





Model of Majority Carrier current Flow

Fig.4.18

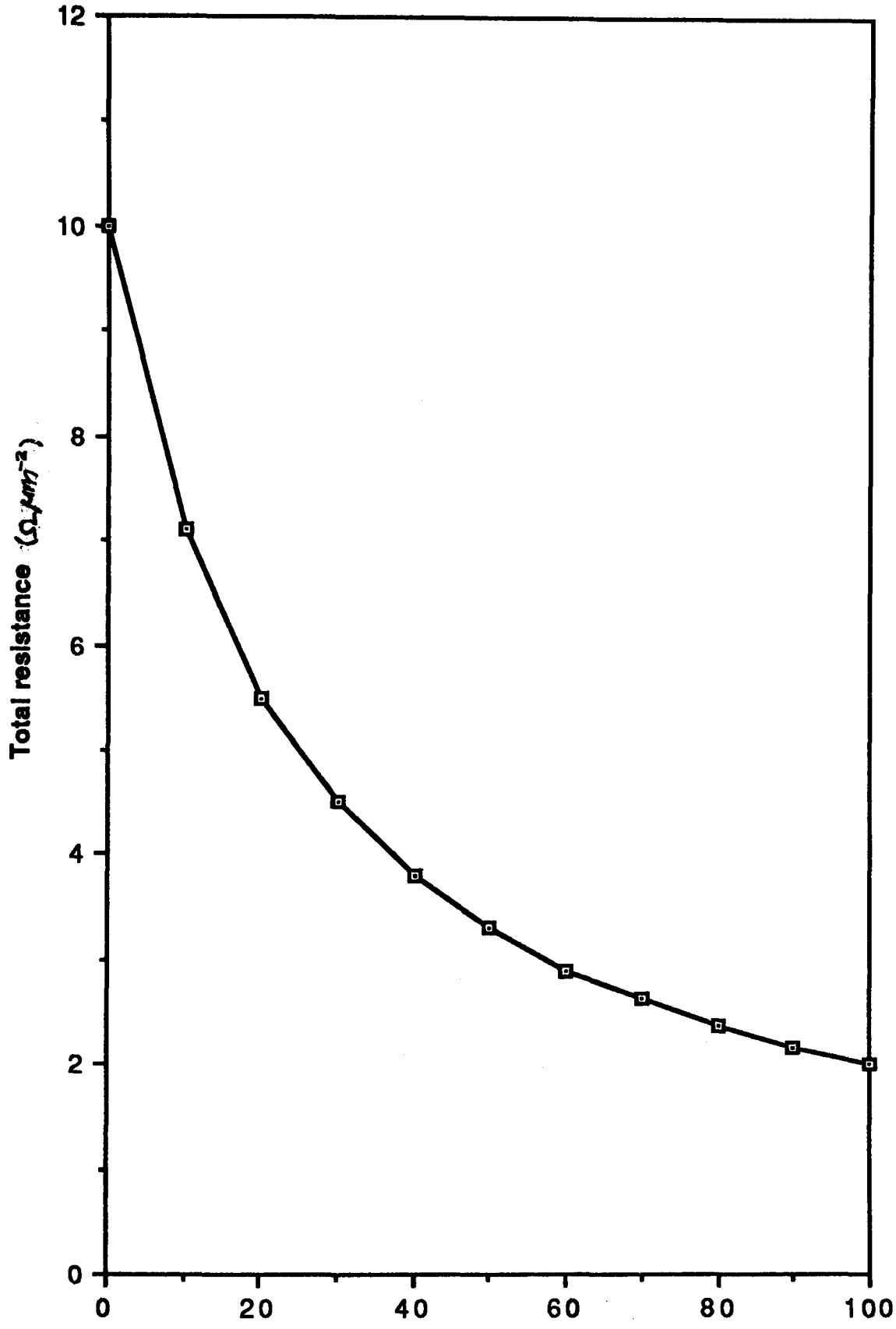
observed grain size varies considerably, but in general is $\geq$800-1500Å. As-grown LPCVD polysilicon generally consists of a disordered region of small grains near the interface and a layer of columnar grains which may contain a high density of micro twins above. Such a structure does not exist in processed polysilicon. After the normal thermal events associated with the production of real devices (such as emitter drive-in), the polysilicon grains recrystallise into an equi-axed randomly orientated structure. That is, the grains have similar dimensions in all directions and do not exhibit any preferential orientation particularly relative to the growth direction. It has been shown by a number of workers that all minority carrier activity occurs within the first 1000Å of the polysilicon beyond the interface, therefore if any barrier effect exists in the polysilicon layer, it must occur at the interfacial boundary and possibly also the first grain boundary only.

In the following section, the mechanisms described above will be compared with the experimental data presented earlier in this chapter, and the relative contributions of the differing mechanisms will be considered

4.3.3 MAJORITY CARRIER TRANSPORT IN THE EMITTER

It is important to know the degree of break-up of the interfacial region when modelling the carrier transport. The measured values, obtained as described above and in chapter 2, are given in table 4.3. Where it was not possible to measure the areas from PVTEM results, the figures were extracted by measurement from cross-sectional micrographs as detailed in the results section. The technique used for each measurement is detailed in the table. Later in the text further reference is made to the percentage break up of the oxide layer. In all cases, this refers to the percentage of the interfacial area which is no longer covered by oxide.

A simple resistive model of transport in the emitter of a device is given in fig.4.18 for the case of a discontinuous layer. In this model, R_{il} is the resistance



Total interfacial resistance vs. interfacial layer break-up

Fig.4.19a

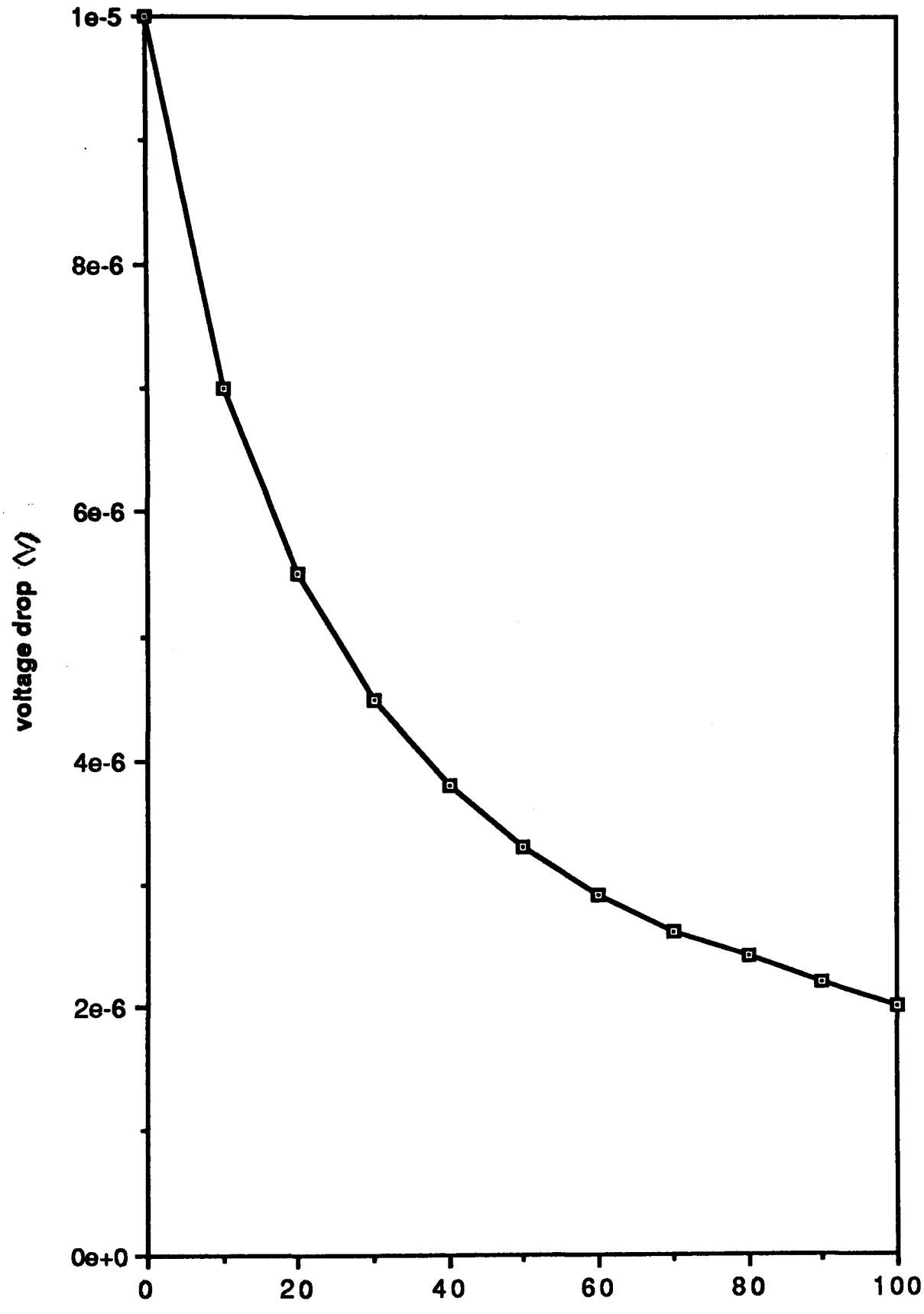
of the interfacial layer and R_{nil} is the resistance of no interfacial layer, i.e. of the gaps. The resistor networks produce a simple, one dimensional model for the analysis of majority carrier transport.

In such a device, the potentials set up throughout the emitter are controlled primarily by the majority carrier current flow as this is at least one hundred times greater than the minority carrier component. The model will be used to simulate emitter currents and thus derive the potential across the interfacial region.

Wolstenholme has shown experimentally [30] that for thin oxides (≤ 20Å) the tunnelling current of the majority carriers may be expressed as approximately ohmic. This confirms the theoretical calculations of Cuthbertson and Ashburn[26]. The modelled RCA interfacial layer resistance was measured by Wolstenholme to be approximately $10\Omega\mu m^2$. The polysilicon resistance, taken from the same reference is $8\Omega\mu m^2$ for P doping and a layer thickness of 0.3μm, for the polysilicon thicknesses used here.

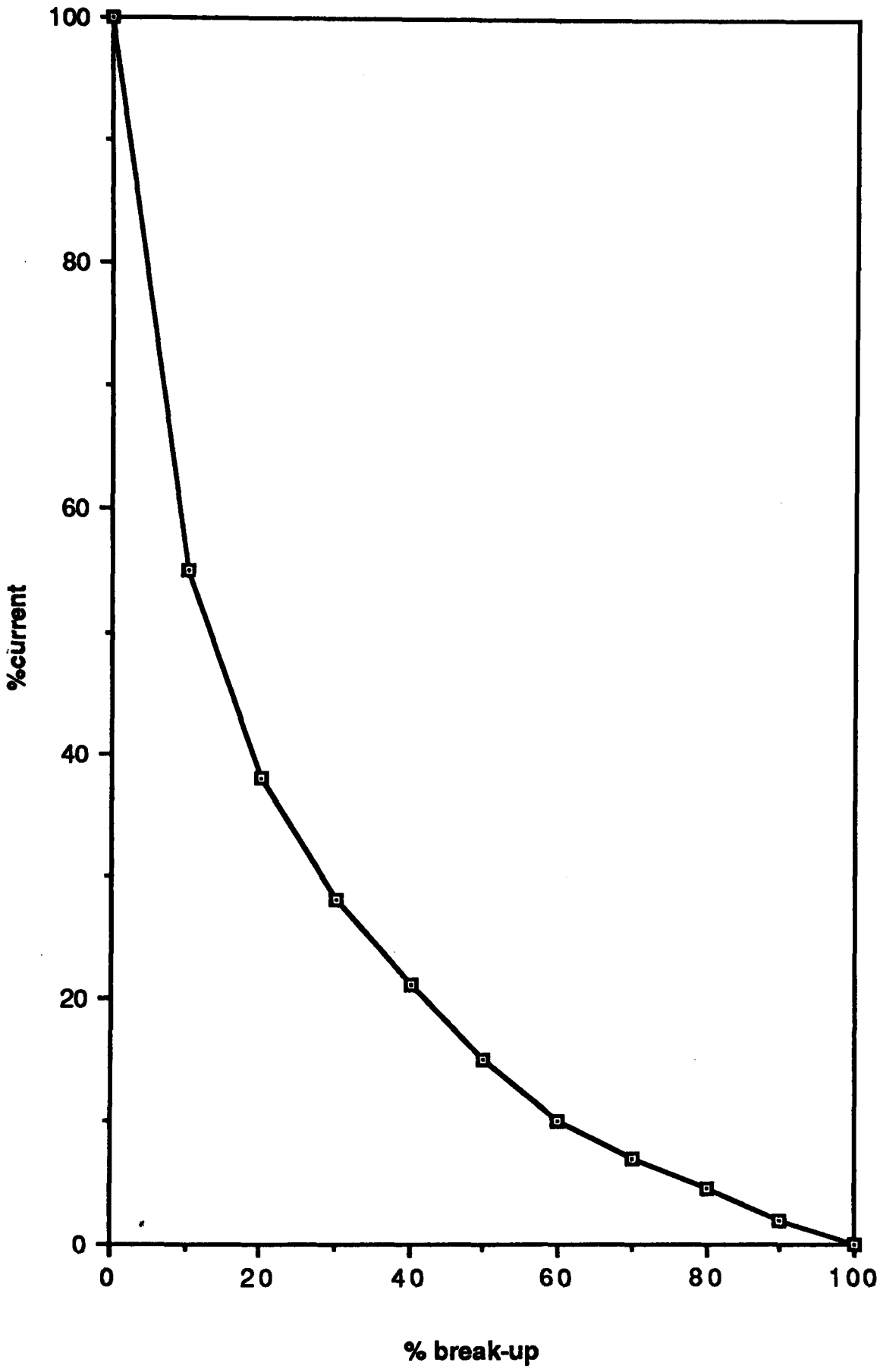
The difference in resistivity between polysilicon and monosilicon is largely due to the grain boundaries, the regrowth that pierces the gaps in the layer does not contain any boundaries and thus might be expected to have a resistivity closer to that of monocrystalline material. Assuming a grain size of approximately 1000Å, the majority carrier current flow would have to pass through an average of four grain boundaries. If the majority of the resistance of the polysilicon layer is caused by these boundaries, the total resistance may be divided approximately into four equal parts. This results in values of approximately $2\Omega\mu m^{-2}$ per boundary for P doping. These values have been assumed for the resistivity of the interfacial regions with no layer coverage (P_{nil}) and the results developed on this basis.

Fig.4.19(a) illustrates the results produced by the model for total interfacial resistance during break up of the interfacial layer. An intermediate current density of $100 A cm^{-2}$ has been used in the model to extract the potential drop



Voltage drop vs. break-up

Fig.4.19b



% current through oxide vs. break-up

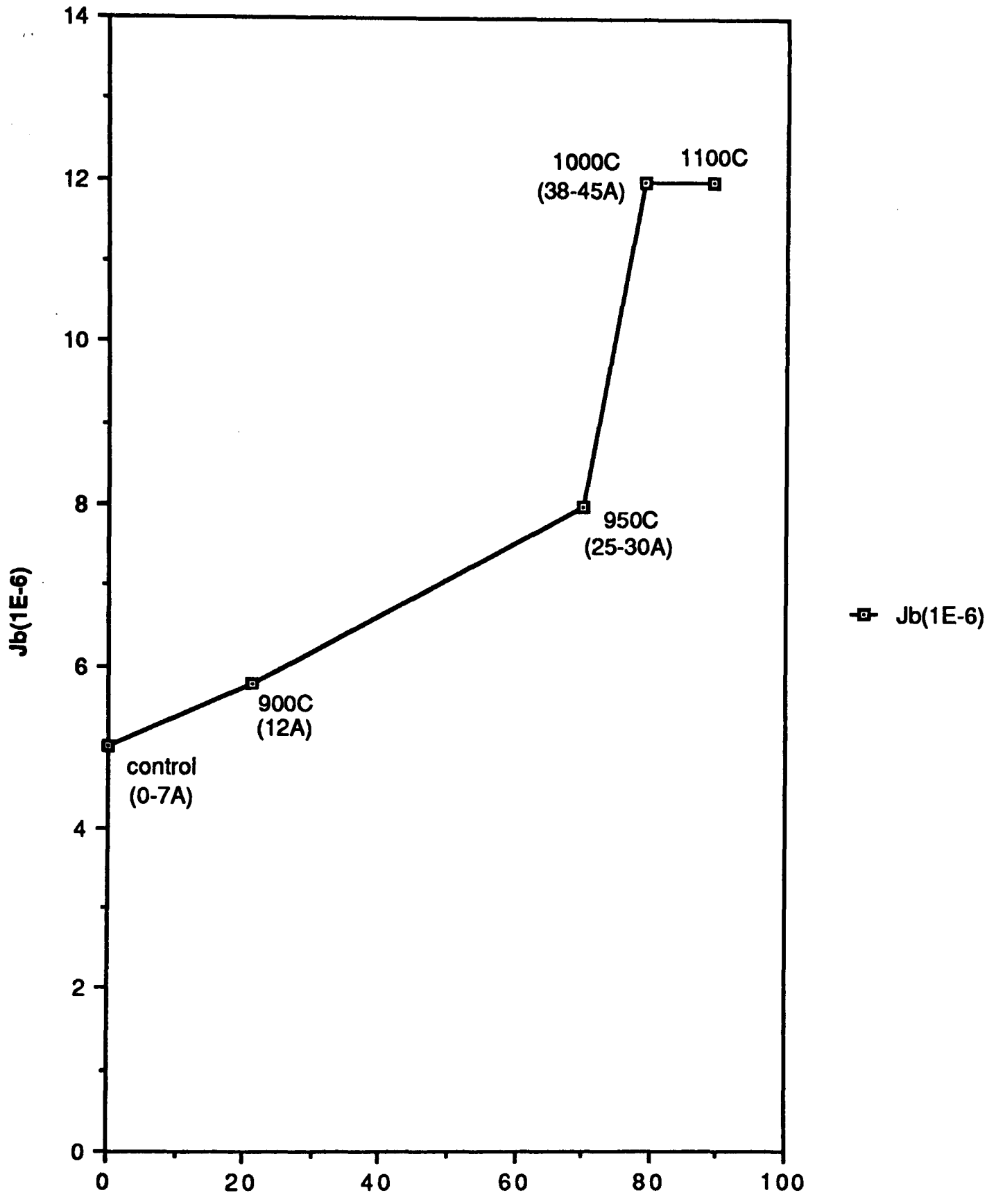
Fig.4.19c

across the interfacial region (V_{ox}) resulting from the effect of the majority carrier current flow. The oxide is assumed to be of constant thickness. Fig.4.19(b) shows the resulting potential drop as a function of the break up of the oxide layer. Fig.4.19(c) uses this to calculate the division of J_{Epoly} (the electron current density at the polysilicon boundary) between the two pathways. It is important to realise that the change in conduction pathways would have little effect on J_E, as V_{ox} is a small fraction of V_{BE}, typically, V_{BE} might be 0.7V, whilst V_{ox} is of the order of 10^{-6}-10^{-4}V. The emitter would appear smaller, but this would only be apparent in situations where the current density was high enough to result in emitter crowding. Thus a picture has been established of the variation of V_{ox} with interfacial layer break up. This model involves a number of simplifications. The most important of these is that the increasing thickness of the oxide results in a very large increase in the resistance of the remaining oxide. The equation for hole tunnelling current of Eltoukhy and Roulston given above has a similar form when applied to majority carriers, and the tunnelling current for majority carriers decreases exponentially with increasing oxide thickness. Thus the actual percentage of majority carrier current passing through the gaps is much larger than that shown by the model, and the corresponding curves would move downward and to the left. At high current levels a lateral potential drop will be generated under the oxide, similar to the effect of emitter crowding. A potential would then be generated across the oxide which would result in an increase in majority carrier tunnelling, however this would only occur in situations of high current density.

4.3.4 MINORITY CARRIER TRANSPORT

Although the model developed above is not sophisticated nor unique, it is included to illustrate clearly the difference in behaviour between minority and majority carrier behaviour and to model the voltage drop across the oxide. It is tempting to construct a similar model for minority carrier current flow. Knowing approximately the voltage drop across the oxide and the hole current density, it might be imagined that an ohmic expression could be derived. It is neccessary to consider the behaviour of the minority carriers in more detail to understand why this is not so. When the oxide is sufficiently thick, tunnelling through this layer significantly reduces the hole current. It might be imagined that in the limit of an extremely thick oxide, and an extremely thin emitter, the hole current would fall to zero. However this ignores the effect of recombination at the oxide surface. In the limit this will control the base current along with recombination in the single crystal emitter. As the oxide becomes thinner, the tunnelling current increases. The total base current therefore is the result of recombination in the single crystal emitter, at the oxide interface and beyond the oxide (if tunnelling occurs). It is neccessary therefore to use an expression for surface recombination velocity as described in chapter 1. This term will then include all the different recombination effects and allow a useful expression of base current.

Consideration of the structural data, in conjunction with data for spreading resistance presented above, allows the conditions to be established which the hole current suppression mechanism must fulfill. Firstly, it must be stressed again that the reduction in minority carrier current flow arises from a reduction in the hole concentration gradient at the emitter side of the emitter/base junction. This must result from some discontinuous change in the hole concentration in the emitter. Other experimental evidence detailed above has shown that the base current is independent of polysilicon thickness down to 500-1000Å or less.

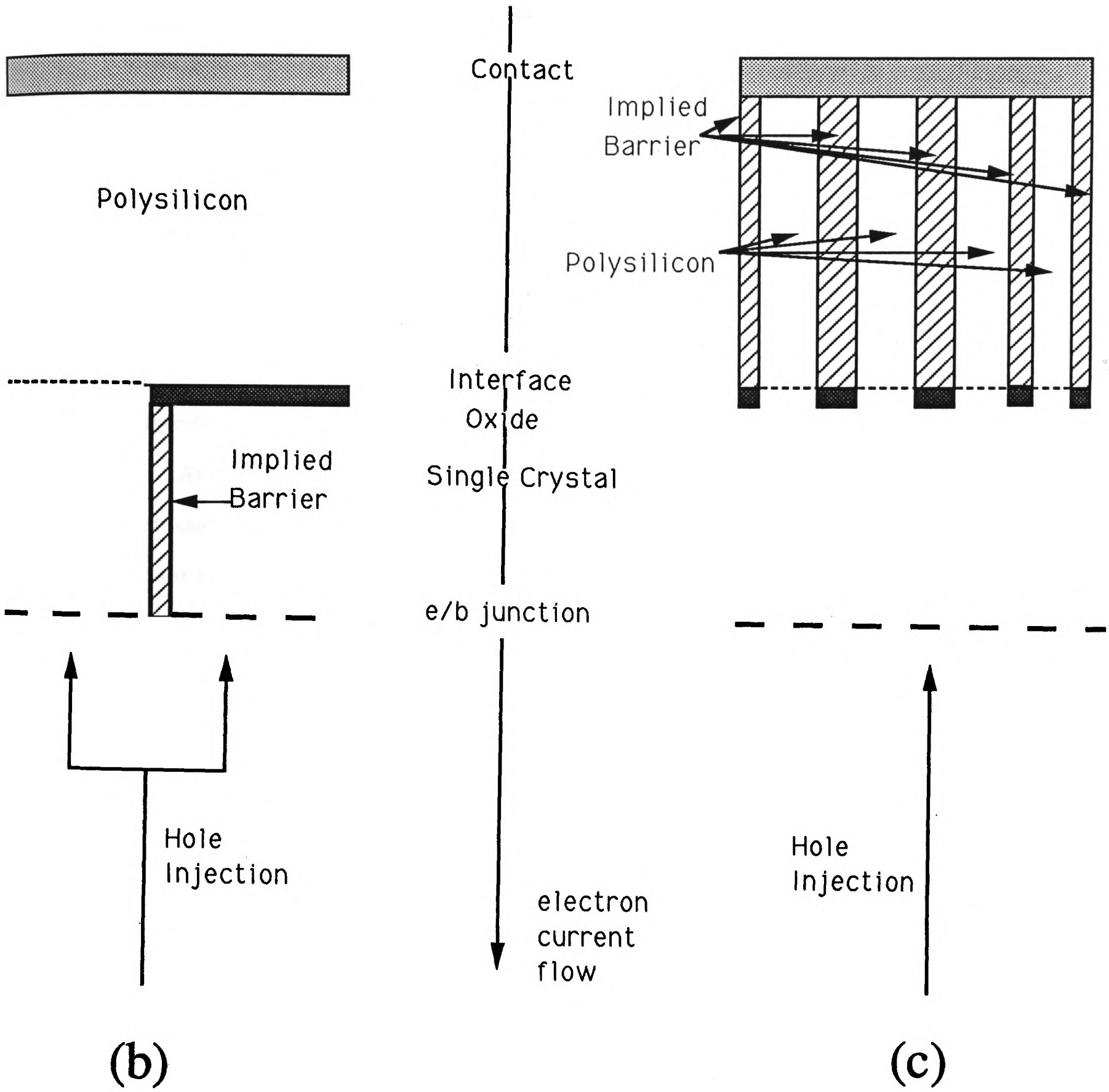


Phosphorous doped, HF device

Fig.4.21

$$Sp = Sp_{il}\frac{A_{il}}{A_{EW}} + Sp_{nil}\frac{A_{nil}}{A_{EW}}$$

(a)



(b) (c)

Fig.4.23

Therefore it is possible to state unambiguously that the gain enhancement controlling mechanism/s are associated with a region between the emitter/base junction and the first 500-1000Å of the polysilicon. Within this area there are four regions. The single crystal emitter region, the interfacial layer, the interface itself and the first 500-1000Å of polysilicon. These are considered in turn below. Fig.4.10 shows a series of plots of the saturation base current density against pre-anneal temperature for the RCA cleaned samples. These can now be used in conjunction with the measured values of exposed interfacial area (i.e. the area that is no longer covered by oxide) to produce a plot of base current density J_B against exposed area. This plot is shown in fig.4.21. J_B is plotted here on a linear scale. The graph serves to impress the stges at which the actual rise in current is greatest; the continuous use of logarithmic plots can mask this.

According to de Graaff and de Groot, gain enhancement can be strongly influenced by differential oxide tunnelling. Thus an accumulation of holes occurs under the oxide on the single crystal side of the interface, where the concentration is defined as accumulated when it is higher than the equilibrium concentration at the same distance from the junction in a similar, but entirely single crystal emitter device. The apparent resistance of the oxide for holes is therefore higher than the resistance of the oxide for electrons.

An attempt was made by Wolstenholme to model the minority carrier behaviour in devices with broken discontinous interfaces by measuring the effective surface recombination velocity at the interface for the balled up oxide and intact oxide cases, and multiplying these two values by their respective surface areas as detailed in fig.4.23(a). This model however has certain deficiencies. Firstly, it is in effect creating a device of the type illustrated in fig.4.23(b) and 4.23(c). It can be seen from this that the single crystal emitter is effectively divided in two by an implied barrier which is impenetrable to holes and has negligible effect on recombination. In effect this would mean that holes arriving under the oxide do not have the option of transport through the gaps in the layer, that is, they

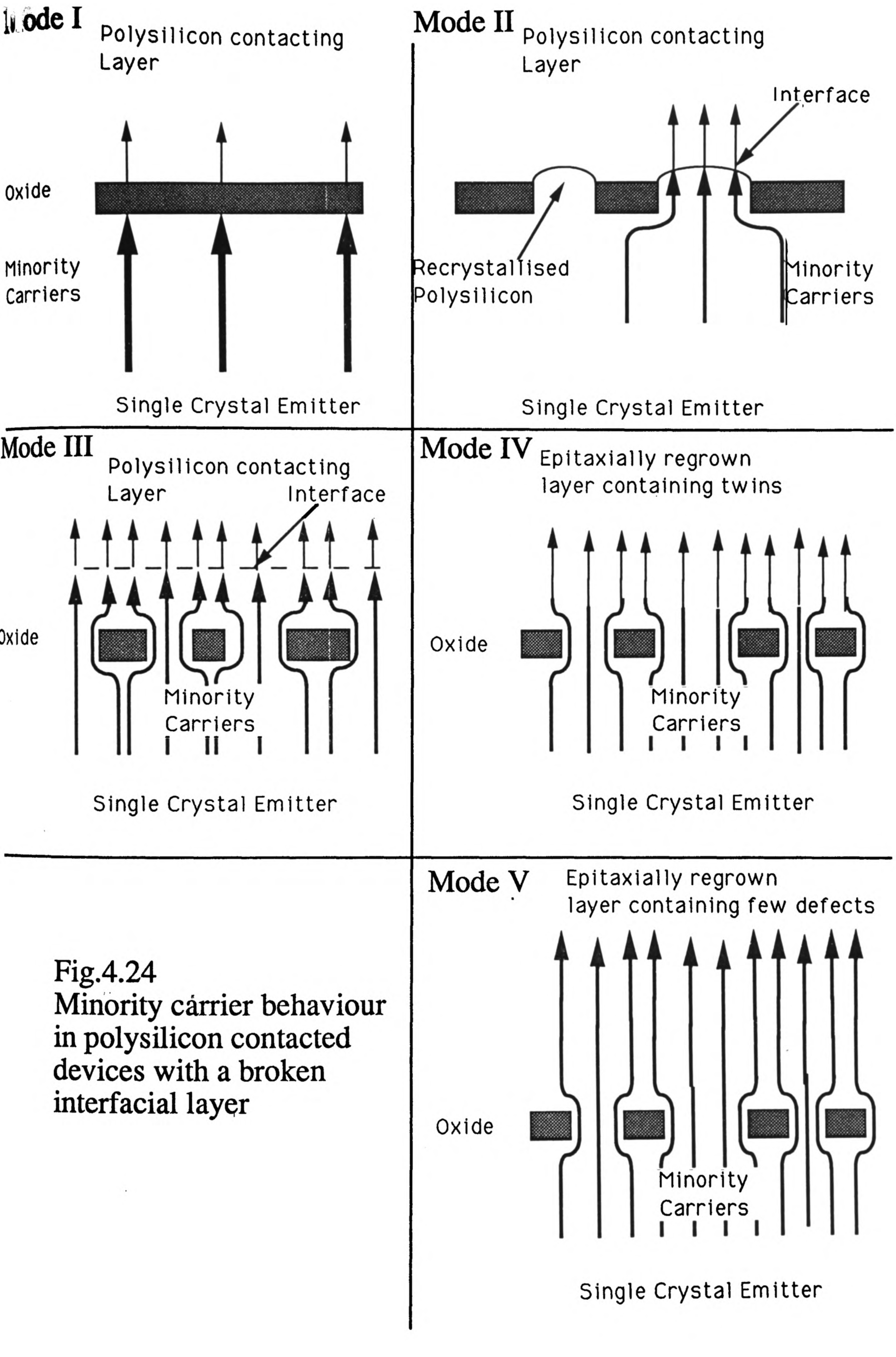


Fig.4.24
Minority carrier behaviour in polysilicon contacted devices with a broken interfacial layer

are spatially localised below the oxide. Fig.4.23(c) illustrates the effect of this model in the polysilicon contact, where the number of recombination centres is scaled with the oxide free area. This has the effect of increasing the effective minority carrier lifetime. Fig.4.24 illustrates the actual spatial components for recombination in the emitter and its polysilicon contact, and these are discussed below.

In the case of a real device with a broken oxide layer, the accumulated holes are not spatially localised. Thus holes arriving at the oxide will diffuse in the plane of the interface, below the oxide. If these holes intersect a gap in the oxide, which presents a significantly lower barrier than the oxide, they will pass through this gap, assuming that the separation of the gaps is a small fraction of the minority carrier diffusion length (approximately 0.7μm in this case), and that the density of recombination sites on the oxide surface does not increase as it breaks up. The hole current component passing through these gaps is not preferentially suppressed with respect to the electron current. This lowers the hole concentration under the oxide and reduces tunnelling. This does not include the effect of the applied field which would exist around the gaps as a result of the emitter (electron) current flow. This would provide a drift component to the hole transport through the gaps and thus would increase hole transport through these gaps. The concentration gradient of holes between the emitter side of the oxide and the gaps would of course result in a diffusion current as well. Furthermore, in the previous section, the majority carrier current flowing through the oxide was shown to fall off rapidly as break-up of the oxide layer progressed. The equation of Eltoukhy and Roulston for hole tunnelling current given above, shows that the tunnelling current increases exponentially with V_{ox}. Thus, as for the majority carrier current, when the voltage across the oxide decreases, controlled largely by majority carrier transport, the minority carrier tunnelling current also decreases exponentially.

The surface recombination velocity is commonly used as a lumped term when describing minority carrier behaviour as discussed in chapter 1. The use of a two single terms to describe the values of S_{Pil} and S_{Pnil} (the surface recombination velocity of an interfacial layer and of no interfacial layer respectively) throughout the break up of the layer, is however an unjustified simplification. Reference to fig.4.24 shows the contributions at various stages of layer break-up to the hole current. Thus it can be shown that the hole current in a device with a discontinuous interfacial layer contains components both dependent and independent of the oxide free area. Once the grain boundary has moved entirely away from the oxide as in mode III, fig.4.24(c), the hole current is largely independent of the oxide free area. These five discrete situations are represented by the five modes illustrated.

4.3.5 COMPUTER MODELLING

If the oxide presents a higher barrier to holes than electrons, then a greater proportion of the hole current would tend to flow through the gaps than for the electron current. In this work, the BIPOLE computer program[24] is used to simulate various combinations of polysilicon, monosilicon and oxide parameters in polysilicon bipolar devices. The author is deeply indebted to the program's author, Professor Roulston for his help and advice in creating device models.

The details of the parameters used to establish the models are given in appendix.1. The simulation model for the emitter includes doping profile in the monocrystalline region, polysilicon doping and layer thickness, hole mobility reduction, Auger recombination and an interfacial oxide with variable thickness and independent barrier heights for holes and electrons. An oxide of uniform thickness is asssumed, although it has been observed that the oxide is non-uniform in thickness. The grain structure is modelled by a uniform reduction

Device	Gain
14A oxide, polysilicon contact	969
8A oxide, polysilicon contact	365
no oxide, polysilicon contact	353
no oxide, single crystal contact	85

Computer modelling results

Table 4.4

device	Gain
shallow junction, 14A oxide, polysilicon contact	1260
shallow junction, 14A oxide, single crystal contact	1060

Computer modelling results

Table 4.5

in mobility in the polysilicon layer to 0.07 of the value for the single crystal silicon. This ignores the integer nature of the grain boundaries especially as it has been observed that only the first grain and its boundaries are associated with the minority carrier transport. However it provides a working simplification to utilise the package. The model of grain boundary mobility reduction is used as a working approximation, however it must be stressed that the minority current supression effect appears more likely to result from the nature of the polysilicon/monosilicon boundary, and not from the presence of the polysilicon itself.

Five models were created, based on the parameters of the fabricated devices, i.e. emitter depth, base width etc. (i) a polysilicon contacted device with a 14Å oxide (RCA), (ii) a polysilicon contacted device with an 8Å oxide (HF-maximum thickness), (iii) a polysilcon contacted device with no oxide, (iv) a single crystal silicon contacted device (the regrown case) and (v) a single crystal contacted device with a 14Å oxide (discussed later). Simple runs were executed for a V_{CB} of 5 volts. The results for the five different models are given in table 4.4. All measurements were taken at saturation.

In previous studies, it has been implied that the two proposed mechanisms for hole current suppression, i.e. oxide tunnelling and mobility reduction in the polysilicon are competing mechanisms. However, as described above, the holes leaving the oxide layer are then subject to the transport properties of the polysilicon. This results in the accumulation of holes which raises Sp_{IL} (the surface recombination velocity of the interface) and hence reduces the tunnelling current below that for a single crystal contact. This can be demonstrated using BIPOLE. Table.4.5 gives the gains for two such devices with 14Å thick layers. These results were obtained from a model incorporating a shallow emitter (100Å) to reduce the effects of recombination in the single crystal emitter. The presence of the polysilicon layer results in a significant increase in gain. Thus we can see, as expected, that the two processes are co-operative and not competitive.

4.4 CHARACTERISATION OF DEVICE BEHAVIOUR

A position has now been reached from which the electrical behaviour of the devices may be described. Consider first the HF case. With no pre-anneal, as 8Å is the maximum thickness of the oxide, and it is perforated, in parts, the effect of the oxide on gain supression should be negligible. This is equivalent to mode II in fig.4.24. Thus the gain suppression should be almost entirely due to the polysilicon layer. The experimentally determined gain enhancement in this device is approximately 2.6. This compares with the computer model value of 3.5 and with the value obtained by Ning and Isaac of approximately 3. Thus a reasonable fit is obtained for this sample. Considering now the 900°C pre-anneal, this appears little changed from the the previous sample structurally, and indeed little change is measured electrically.

The 950°C sample is particularly interesting. The polysilicon layer in this sample has completely regrown, but significant gain enhancement remains of approximately 1.5. The oxide layer is no longer contributing to this minority carrier supression, and the polysilicon layer has regrown. However, the layer contains a high density of growth twins and other defects, equivalent to mode IV, and the conclusion that might be drawn is that these twins are responsible for modifying the minority carrier transport in some way, such as mobility reduction. Effects of this type have been previously observed in twinned silicon in silicon on sapphire (SOS) material . This does not mean however that all twinned material would have such an effect. Since twin boundaries have high symmetry and low energy, the most likely reason for any modification of the conduction mechanism is the presence of steps in the twin plane or some extrinsic effect such as impurity segregation.

However all measurements of the effect of these twins on carrier transport have been carried out with the current flow parallel to the surface and under an applied field (drift). In the situation investigated in this work however,

the diffusion current is flowing approximately parallel to the twin boundaries and therefore should not need to cross them. An alternative possibility to explain the observed supression of hole current is the presence of a peak in the doping profile at the position of the original interface. As this sample was polycrystalline when implanted, it should have the characteristic flat initial profile falling rapidly in the original single crystal region. It has been demonstrated that when the polysilicon regrows, the active dopant concentration rises. If the segregated dopant peak at the interface were retained after this pre-anneal, it would provide the high-low thermionic barrier to carrier transport postulated by several authors but not observed in samples which are still polycrystalline.

In the 1000°C pre-anneal the re-grown layer no longer contains these crystalline defects and the gain enhancement is now 1 (a gain of 100) which is very close to the computer modelled value for a single crystal contact with no oxide, of 0.85. This is mode V behaviour. What is perhaps fortuitous is that the 1100°C sample shows a similar gain, given that it was implanted into a completely re-grown layer. However, this device may have unpredictable HF characteristics due to the great difference in the doping profile from the other devices.

Consider now the RCA case. The higher gain of the non pre-annealed sample results from the greater contribution of tunnelling in the oxide caused by the thicker interfacial layer, the gain enhancement is approximately 10, and again this figure is close to the value produced by the computer model of 9.69. This is mode I behaviour. The 900°C sample shows rapid loss of gain from a gain enhancement of 10 to a gain enhancement of 6. The oxide has just begun to break-up (mode II) with a maximum value of 5%. It might be expected that all oxide controlled gain enhancement would have disappeared as there are a large number of gaps in the oxide and this should be equivalent to mode II, behaviour (3.5), but there is apparently still a small contribution. This is believed by the author to result from the nature of the polysilicon/monosilicon interface in the gaps which introduces an area component to the S_p term.

In the 950°C pre-annealed sample, the gain enhancement has decreased to approximately 3.75 and at this stage there should be no contribution remaining from the oxide (mode II). Although the gain is high compared with the value of 2-2.6 for the HF case, it is close to the modelled figure of 3.5. However, the enhancement for the 1000°C is lower still at approximately 1.9. This is close to the value for the HF 900°C case in which epitaxial regrowth has just begun. However the model does not fully explain the behaviour of the RCA 950°C sample in which the observed gain is slightly higher than expected from the HF control case. The oxide in this sample is slightly thicker and more uniform than that of the HF control sample, and the difference is believed to be due to the difficulty of constructing an accurate representation of the non-uniform oxide for the computer model. Alternatively the effect could result from a reduction in the depth of the single crystal emitter caused by the thicker oxide acting as a diffusional barrier, SIMS results for the RCA cleaned devices are not available to confirm this. However other workers have reported the existence of a diffusional barrier in the presence of thicker oxides (for example Josquin et al [31] et al). In the RCA 1000°C sample, the behaviour is still mode II, but the gain is now marginally lower than for the completely regrown HF sample. This is probably due to the emitter being somewhat deeper, resulting from the larger number of gaps in the oxide allowing more diffusion of dopant.

4.5 CONCLUSIONS

Based on the observations of grain size and orientation in typical polysilicon device type layers in both this and the previous chapter, and in conjunction with the published data, it can be concluded that the minority carrier diffusion length in the polysilicon layer is comparable with or less than the grain dimensions in

the direction of current flow. Thus the hole mobility reduction model proposed by other authors does not satisfactorily explain the relevant gain enhancement measured in devices with continuous interfaces.

An examination of the data on dopant segregation to interfaces suggests that excess ionised impurities do not exist at the polysilicon/monosilicon interface. However the evidence of a relationship between a reduction in hole current and the segregation of dopant has been convincingly demonstrated by other authors. Thus it appears that the segregated dopant acts as a barrier to minority carriers in some as yet unknown fashion. It appears more likely that this boundary is responsible in some way for the portion of the gain enhancement often attributed to the electrical characteristics of the polysilicon. The additional and entirely separate model of gain enhancement resulting from carrier tunnelling through an interfacial oxide is confirmed by the observations made in this chapter.

Based on the TEM observations in both this and the previous chapter, a clear picture of interfacial oxide break-up and polysilicon layer behaviour has been formed for a number of pre-annealing conditions. This data has been combined with relevant electrical data to formulate a simple model for majority and minority carrier current flow in the emitter. This has led to some refinement over previously published work. The models general validity has been tested using a sophisticated CAD device modelling program, which produces results consistent with experimental observation.

Significant differences have been observed in this chapter between the phosphorus doped samples reported here and the similar As samples previously investigated by Norman Jorgensen. This is consistent with the observations made in chapter three.

An ideal device should have a high gain, which may be traded for an increase in base doping level, reproducible gain and a low series resistance. It has been demonstrated that a device of this type, as proposed in chapter three may

be constructed with useful DC characteristics. Devices were fabricated with a consistent gain of approximately 100 and an epitaxially regrown polysilicon emitter layer resulting in a lower emitter series resistance. Alternatively, it has been demonstrated that in exchange for some potential loss of gain reproducibility, a device may be fabricated with an epitaxially regrown layer for low series resistance, but still retaining some enhanced gain.

As an alternative to arsenic in polysilicon contacted emitters, phosphorous has the advantage of accelerating regrowth and reducing emitter series resistance. However, this must be balanced against the deeper emitter/base junction and hence increased depletion region area which result from the more rapid rate of diffusion of phosphorus.

References

[1] . M. Takagi, N. Nakayama, C. Terada and H. Kamioka. J. Jap. Soc. Appl. Phys.(suppl.), 42 pp.101-109, 1972.

[2] . H.C. de Graaff and J.G. de Groot, IEEE, Trans. Electron dev. ED-26(11), pp.1771-1776, 1979.

[3] . J. Graul, A. Glasl, H. Murrman, IEEE, J. Sol. Stat. Circ., SC-11(4), pp.491-494, 1976.

[4] . T.H. Ning and R.D Isaac. IEEE Trans. Electron Dev. ED-27(11), pp.2051-2055, 1980.

[5] . G.L. Patton, J.C. Bravman and J.D. Plummer, IEEE Trans. Electron Dev., ED-33(11), pp.1754-1768, 1986.

[6] . C.Y. Wong, C.R.M. Grovenor, P.E. Batson and D.A. Smith, J. Appl. Phys., 57, pp.438-442, 1985.

[7] . C.R.M. Grovenor, P.E. Batson, D.A. Smith and C.Y. Wong, Phil. Mag. A, 50 pp.438-423, 1984.

[8] . A. Neugroschel, M. Arienzo, Y. Komen and R.D. Isaac, IEEE Trans. Electron Dev., ED-32, pp.807-816, 1985.

[9] . A.A. Eltoukhy and D.J. Roulston, IEEE, Trans. Electron Dev., ED-29, pp.1862-1869, 1982.

[10] . Z. Yu, B. Ricco and R.W. Dutton, IEEE Trans. Electron Dev., ED-31(6), pp.773-784, 1984.

[11] . P. Ashburn and B. Soerwirdjo, IEEE Trans. Electron Dev., ED-31, pp.853-860 1984.

[12] . K. Sagara, T. Nakamara, Y. Tamaki and T. Shiba, Sol. Stat. Electronics, 31, pp.1139-1150, 1988.

[13] . D.W. Greve, D.L. Chen, A.M. Gruzman, IEEE BCTM, Minneapolis, 1987.

[14] . P. Ashburn to be published, 1989.

[15] . E.F. Chor, IEEE Electron Dev. Let, EDL-6, pp.516-518 1985.

[16] . G.R. Wolstenholme, N. Jorgensen, P. Ashburn and G.R. Booker, J. Appl. Phys., 61(1), pp.225-233, 1987.

[17] . F.S. Becker, H. Oppolzer, I. Weitzel, H. Eichermuller and H. Schaber, J. Appl. Phys., 56(4), pp.1233-1236, 1984.

[18] . A.H Reader, F.W. Schapink and S. Radelaar, M Poly-micro-crystalline and amorphous semiconductors.

[19] . P. Ahburn, Design and Realisation of Bipolar Transistors, Wiley, 1988.

[20] . N. Jorgensen, Private Communication.

[21] . N. Jorgensen, J.C. Barry, G.R. Booker, P.Ashburn, G.R. Wolstenholme, M.C. Wilson and P.C. Hunt, Inst. Phys. Conf. Ser.,76, 471, 1985.

[22] . F.M. Ross and W.M. Stobbs, Surface and Interface Analysis, 12, pp.3-4, Wiley, 1988.

[23] . A. Ourmazd, D.W. Taylor and J.A. Reutschler, Phys. Rev. Let., 59(2), pp.213-216, 1987.

[24] . D.J. Roulston, Bipolar Semiconductor Devices (to be published), 1989.

[25] . K.K. Ng and H.C. Card, J. Appl. Phys., 51(4), pp.2153-2157, 1980.

[26] . A. Cuthbertson and P. Ashburn, IEEE Trans. Electron Dev., ED-32(11), pp.2399-2407, 1985.

[27] . M.M. Mandurah, K.C. Sarawasat and T.I. Kamins, Trans. Electron Dev., ED-28(10), pp.1171-1175, 1981.

[28] . A.T. Paxton and A.P. Sutton, Acta Met.(to be published), 1989.

[29] . A. Bourret and J.L. Romierre, Polycrystalline Semiconductors,M vol.35, Springer Proc. in Physics, Springer-Verlag, 1989.

[30] . G.R. Wolstenholme, PhD Thesis, University of Southampton, 1988.

[31] . W.J.M. Josquin, P.R. Boudanjin and Y. Tamminga, Appl. Phys. Lett, 43(10), pp.960-962, (1983).

Chapter 5

DOPANT DELINEATION

5.1 INTRODUCTION

In the design of all high speed, high performance devices, accurate assessment of the geometry of the junctions and of the dopant distribution has become increasingly critical. When modelling the electrical behaviour of a high speed bipolar device the most important junction to consider is the emitter/base and the associated emitter region as these have the greatest effect on gain through V_{BE} and also through their effect on minority carrier lifetime.

In analysing dopant distributions in semiconductors, it is desirable to know the two dimensional concentration profile of a device section. There are two obvious available routes to follow, either modelling the results using a mathematical construction, based essentially on diffusional mathematics, or some direct measurement technique. Commercial computer modelling packages are widely available, for example SUPREM and TITAN. It is not, however, clear as to the accuracy of such models when applied to junctions of small dimensions and it has become increasingly important to assess the effects of micron and sub-micron features on the real dopant distributions in these devices.

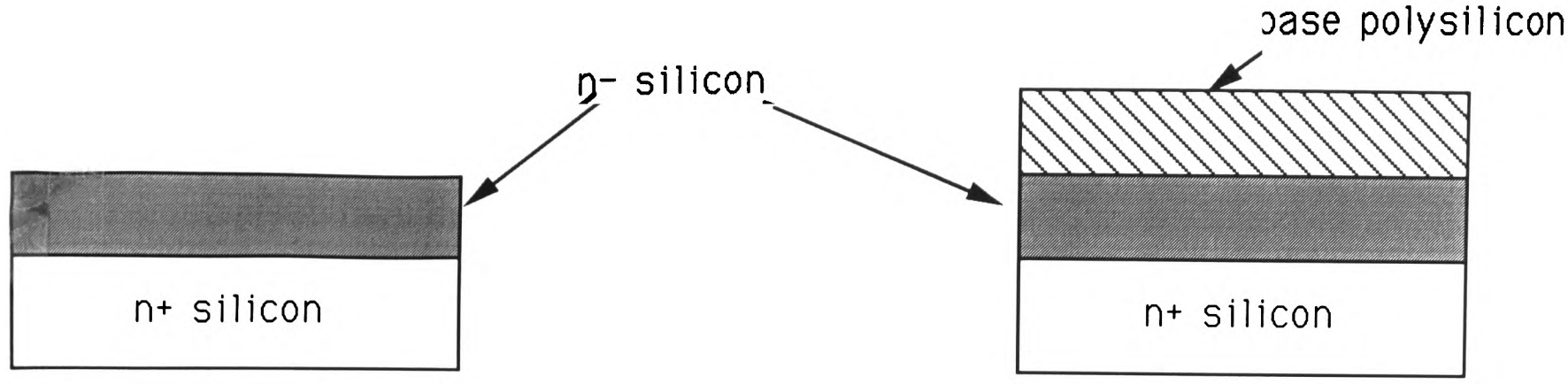
It is important firstly, to consider the dimensions of the structures which it is required to measure. The construction of all modern devices is based essentially upon planar fabrication techniques. These involve preferential exposure

of various areas on a wafer to different process stages (such as implantation), or the preferential removal of deliberately grown surface layers (such as oxides and nitrides), or indeed a combination of the two. Silicon lends itself particularly well to such treatments due to the ease with which oxides and nitrides may be grown from it. The definition of the preferential areas to be preserved/removed is carried out by the use of photo-lithographic techniques. This involves the exposure, through a carefully designed mask, of a photosensitive coating which has been applied to the wafer surface.

Improved performance in terms of cost and speed has revolved around a reduction in the minimum defined feature size and hence a reduction in discrete device dimensions. The radiation used to expose the photosensitive film has generally been light and this essentially limits the minimum definable feature size for this technology.

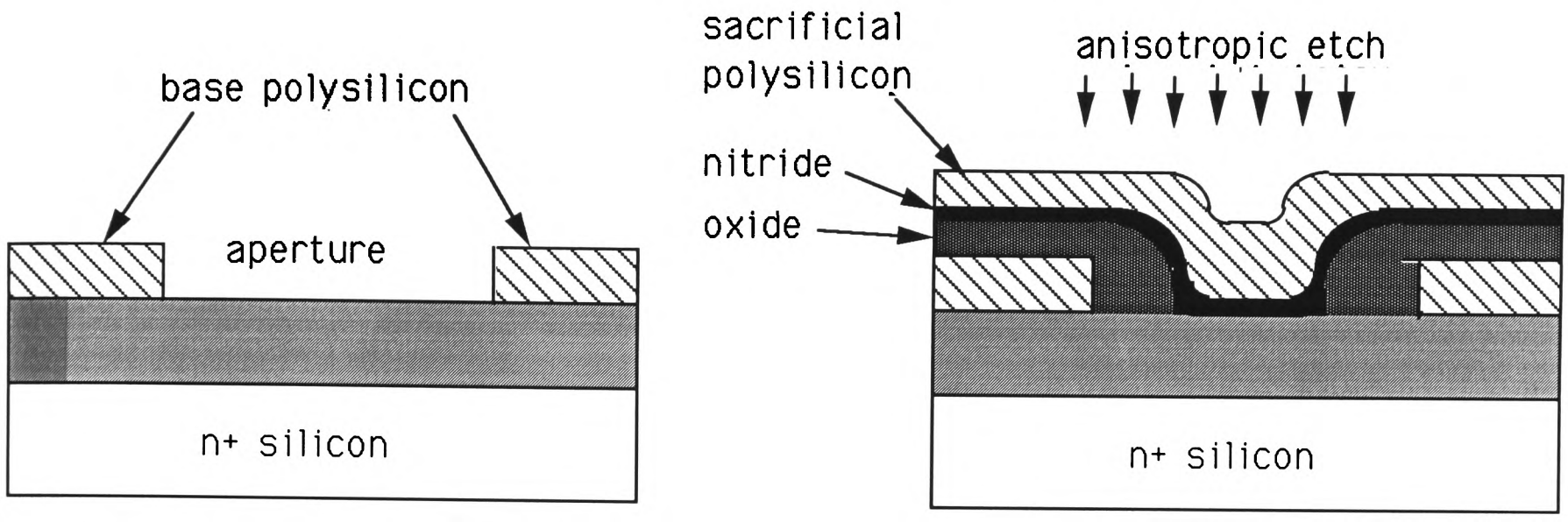
During the early period of planar device technology, white light was shone through a mask series which exposed the whole wafer at one time. As the requirements of manufacture have become more demanding the procedure has been refined so that in the state-of-the-art process discussed here, exposure is by ultra violet (UV) light (short wavelength), and the mask contains the information for one integrated region only. This mask is stepped onto each region in turn to reduce the effects of lens aberrations on the lithographic process. This results in a minimum feature size, which is fundamentally limited by the wavelength of light used. The minimum feature size realistically definable by advanced, conventional lithographic techniques is $\approx 1\mu$m and this represents the present state of the art in production technology [1]. Overall minimum device dimensions are limited by the repeatability of the translation mechanism used to align successive photolithographic stages. This makes self-aligning processes particularly attractive (see chapter 1).

The work in this chapter was carried out on high speed bipolar devices having a maximum measured frequency of oscillation (f_{mosc}) of up to 20GHz [2]



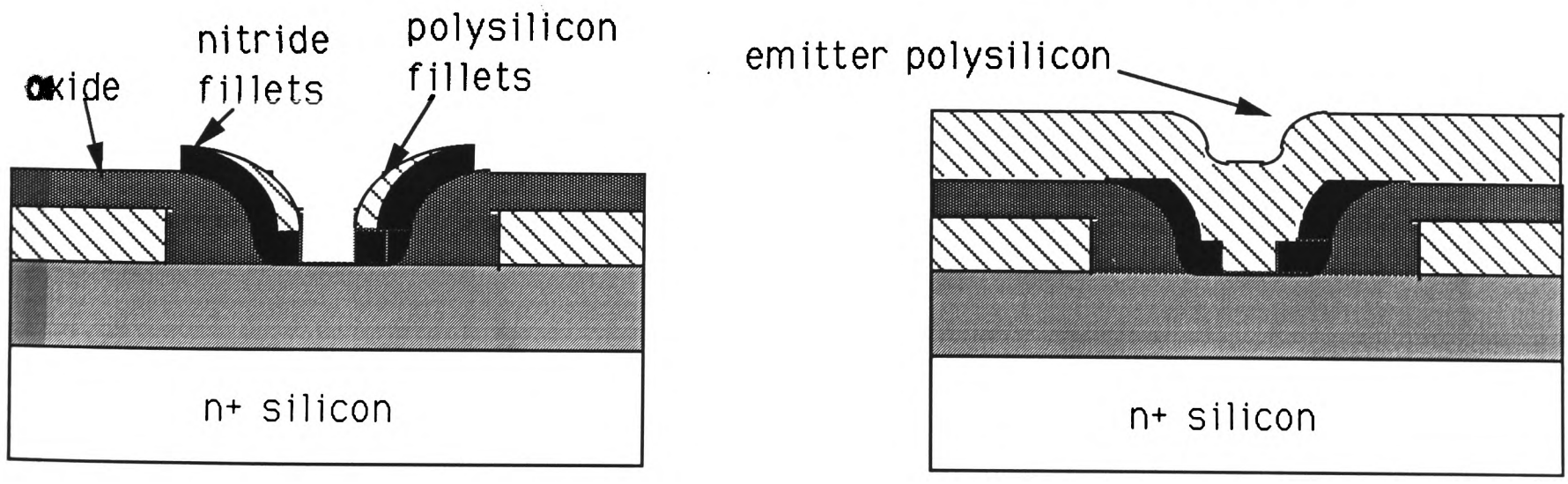
(i) collector epitaxial layer deposition

(ii) base polysilicon deposition + base implant



(iii) Emitter aperture formation +base drive-in

(iv) oxide deposition,nitride deposition sacrificial oxide deposition followed by anisotropic polysilicon etch and then nitride etch

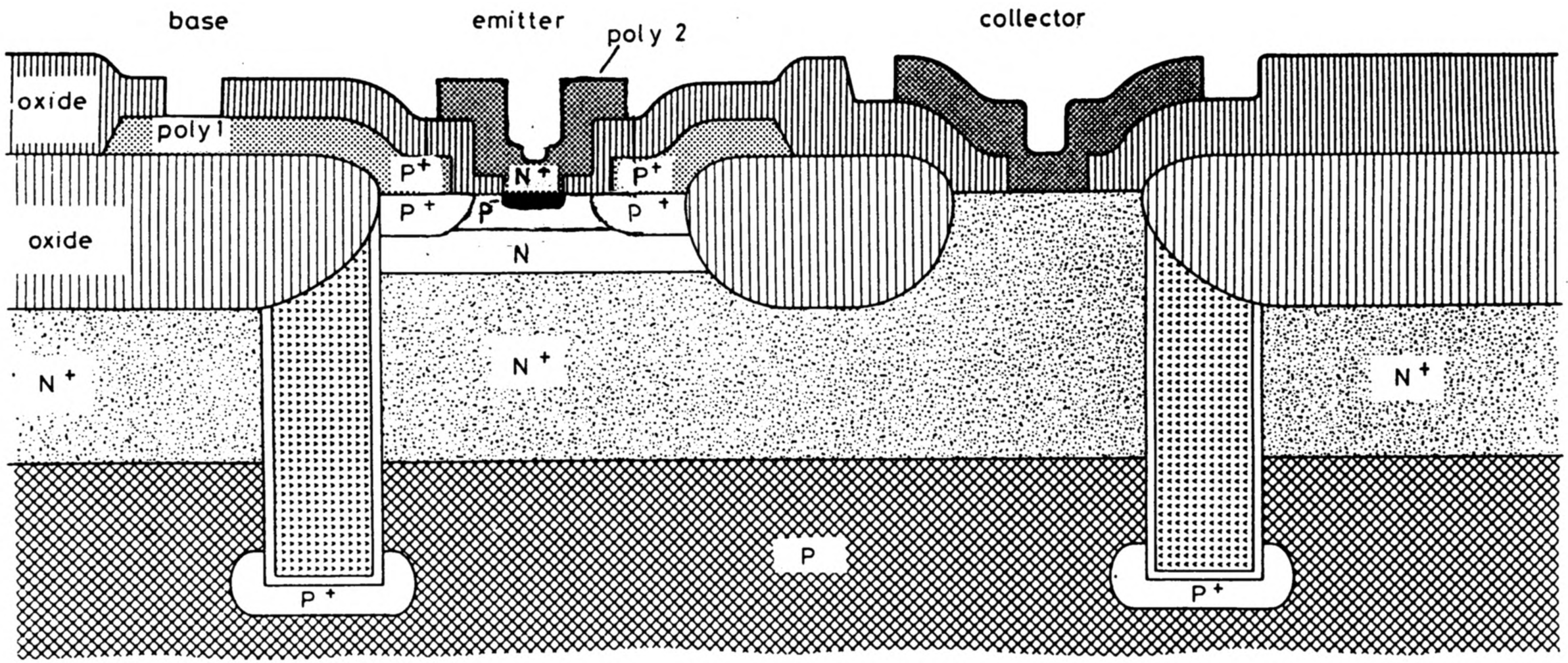


(v) emitter region after anisotropic etch followed by nitride etch

(vi)polysilicon contacted emitter

Fabrication procedure for nitride isolated, polysilicon contacted emitter

Fig.5.1



Schematic cross-section of complete polysilicon contacted device

Fig.5.2

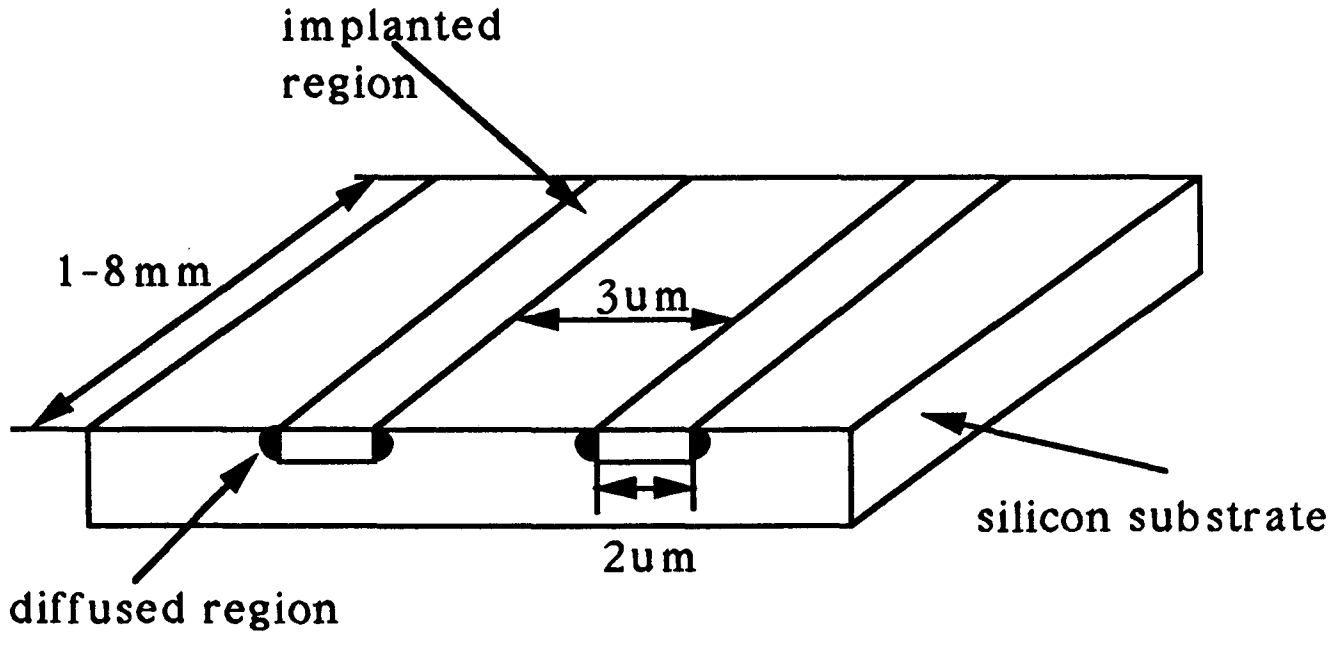
[3]. These are amongst the very fastest silicon bipolar transistors in the world today. The critical features of these devices in achieving this performance are the reduction in the area of the e/b junction by use of a polysilicon diffusional source as described in chapter 1, and the introduction of additional isolation between the emitter and base contacts. Thus both are effectively reducing the capacitance contributing to the time constants of the device. Because of the presence of two independently grown polysilicon layers (for the base and for the emitter) this type of process is generally known as a double polysilicon contacted structure (sometimes referred to as the 'double poly process').

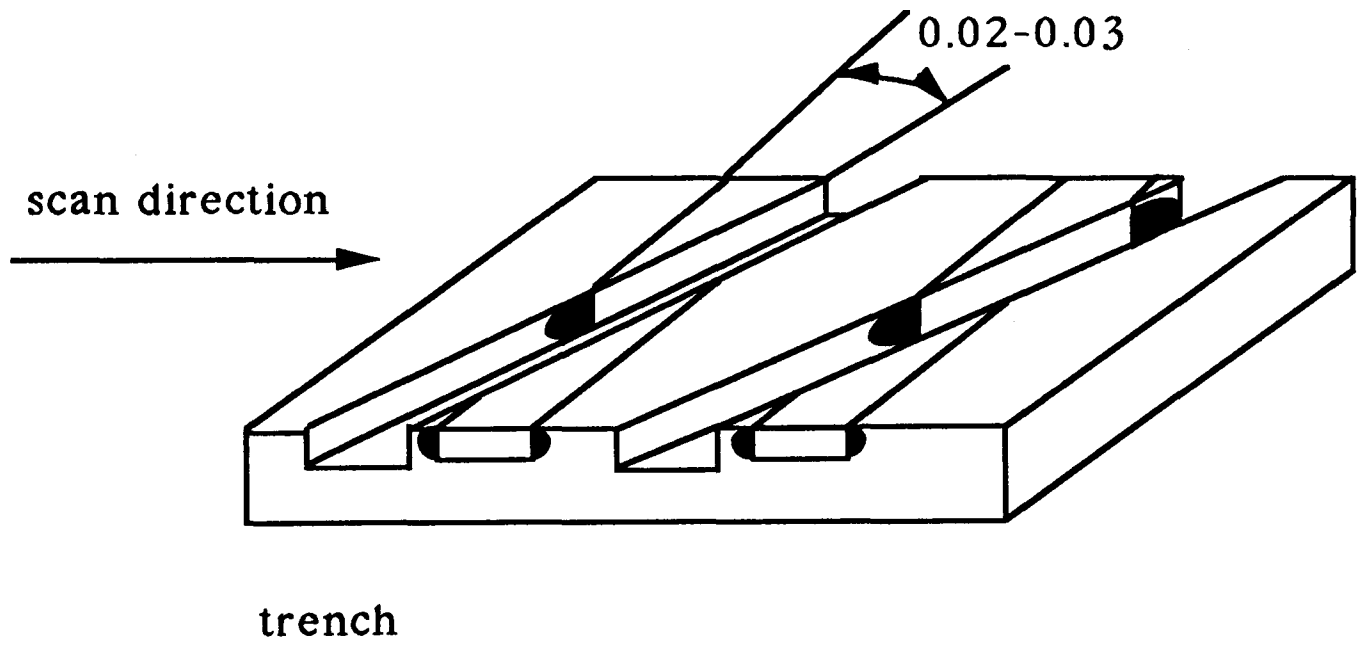
Additional emitter/base contact isolation, beyond the more usual thermally deposited oxide, is achieved by the introduction of nitride 'sidewalls' during fabrication [4]. The fabrication process is illustrated schematically for the emitter and base of a single bipolar device in fig.5.1. Fig.5.2 shows a schematic for a whole device including trench isolation and the collector.

By growth of a conformal polysilicon layer into this feature during the definition of the nitride sidewalls, followed by an anisotropic plasma etch, the effective emitter 'window' (i.e the area of contact between the eventual emitter polysilicon and the single crystal material) can be reduced to as little as 0.4μm wide. Reduction below these dimensions is restricted by the difficulty of growing conformal layers of polysilicon into narrower gaps.

5.1.1 1D MEASUREMENT OF DOPANT CONCENTRATION

Measurement of dopant profiles in real devices has been investigated by a number of workers, and a variety of different techniques has evolved, most of them one dimensional. These have extremely good depth resolution but very poor lateral resolution. They may be divided into two groups, chemically sensitive and electrically sensitive techniques.





2-D SIMS technique

Fig.5.3

The most commonly used chemically sensitive technique is Secondary Ion Mass Spectroscopy (SIMS), described in chapter 2. The eroded hole in the sample limits the lateral resolution to approximately 2μm. Depth resolution can be of the order of 10nm. Sensitivity to impurity concentration for typical dopant atoms in silicon (As,P,B) is of the order of 10^{16}cm^{-3}.

Spreading resistance measurement is the most commonly used technique for determination of the electrically active dopant concentration, i.e the density of ionized dopant atoms. In this technique, the sample is polished to produce a bevel (approximately 17') through the doped region of interest. The spreading resistance of this region is then carefully mapped by applying a low voltage (approximately 10mV) across a pair of probes, and taking measurements from a series of points down the bevel. A plot of spreading resistance against depth is thus obtained. It is possible, fairly simply, to convert this to a measurement of bulk resistivity and from this, to extract a one dimensional dopant profile. This technique can measure dopant levels of approximately $1\times10^{16}\text{cm}^{-3}$. Lateral resolution is limited by the separation of the electrical contacts, which is approximately 30μm or greater. Depth resolution is controlled largely by the angle of the bevel and is typically 10nm.

5.1.2 2D MEASUREMENT OF DOPANT CONCENTRATION

Although SIMS is generally used as a one dimensional tool, work has been conducted into two-dimensional SIMS analysis of specially prepared device-like samples as shown in fig.5.3 [5]. The structure to be investigated is produced in the form of a long strip with the extended dimension perpendicular to the profile to be measured. A masking layer of oxide is then grown on the wafer surface and patterned into a diagonal stripe crossing the region of interest at a shallow angle. This structure then receives an anisotropic etch which results

in trenches crossing the regions at a shallow angle angle as shown in fig.5.3. A SIMS depth profile can now be obtained by carrying out a series of scans across the selected regions. Each scan removes a known depth of material. The first scan is taken from the region at which the trench first begins to cross the profile. By repeating this technique at regular intervals along the trench, as illustrated, a series of profiles can be obtained. Each profile contains the integral of all the dopant present up to the trench edge for any given depth. Knowing the angle of the trench to the doped region and the separation of each series of scans, the two dimensional dopant distribution may be extracted. The lateral resolution is limited primarily by the sharpness of the trench edge, this edge is eroded during the profiling, and this is the primary difficulty with the execution of this technique. A lateral resolution of approximately 30nm is claimed for the technique, with an eventual lateral resolution of as little as 5nm.

This technique, whilst powerful, suffers from a number of disadvantages. SIMS is a chemically sensitive technique and thus does not reveal the electrically active profile. It has been shown in chapter 4 that the total concentration of dopant atoms, may be significantly larger than the concentration of ionized carriers particularly in regions of high dopant segregation. The problem of lateral erosion of the step by the analytical beam must also be resolved. Additionally, the structures neccessary to utilise this technique are not real devices and require the design of specialised masks. Finally, it requires the removal of all surface structure on the device. However, this technique shows great promise.

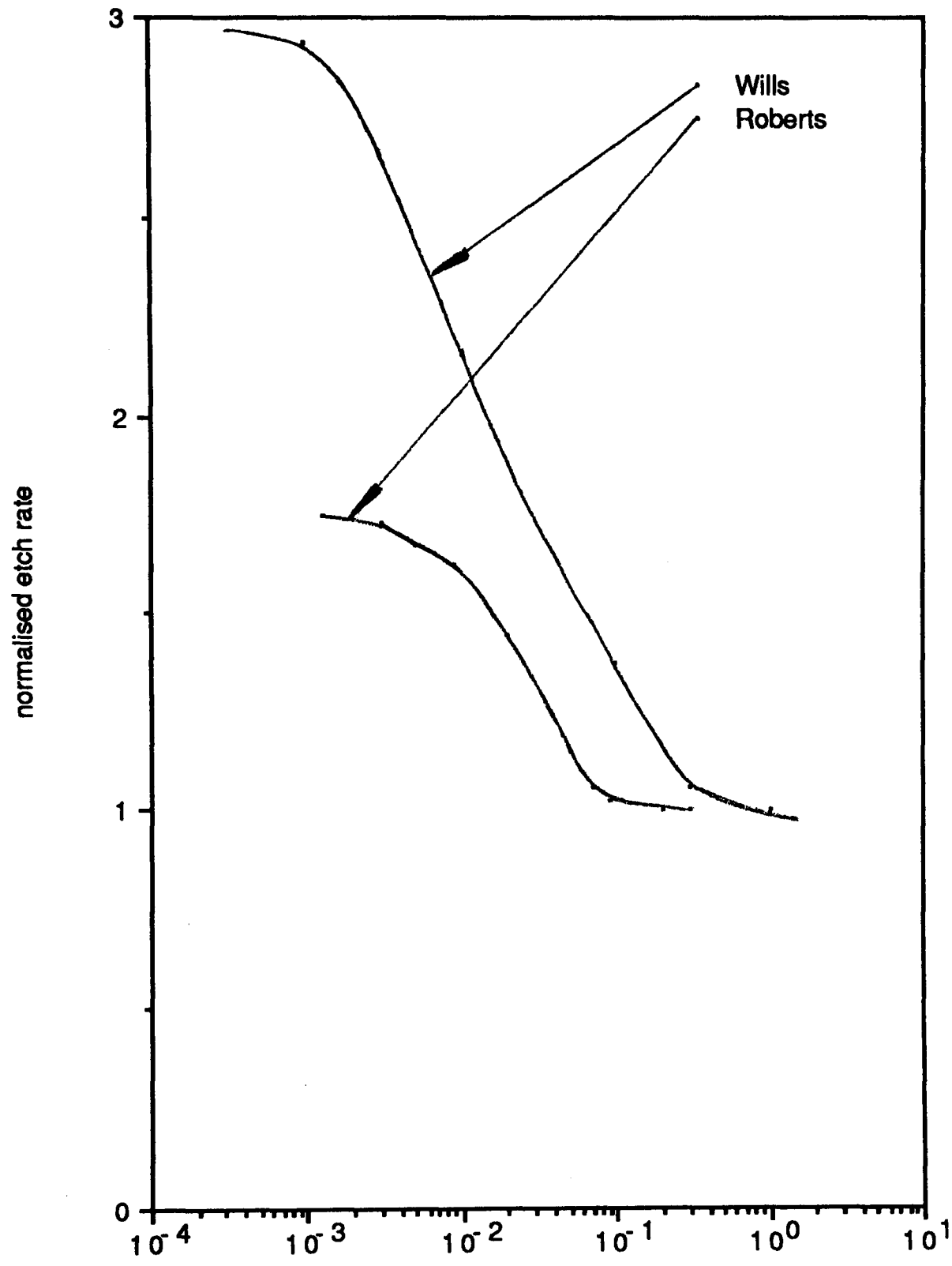
An alternative technique [6] relies on the sensitivity of thermal oxide growth in silicon to dopant concentration in the silicon. An oxide layer is grown on the surface of the area of interest. The thickness of this layer is related to the dopant concentration for high concentrations (approximately 10^{20}cm^{-3}). A layer of this oxide approximately 50Å thick is grown on the surface of the sample, and the bulk of the wafer is then removed from the back surface by jet thinning. This leaves an oxide replica whose thickness is related to the dopant concentration.

By use of a dedicated STEM system, a measurement of the thickness of the oxide may be obtained. It suffers however from poor sensitivity and is typically unable to detect below $10^{20}cm^{-3}$. Lateral resolution is of the order of 10Å. Again this technique is believed to be chemically rather than electrically sensitive.

The final technique considered here is also the work of Hill et al [7]. In this technique, simultaneous anodic stripping is performed on a series of test structures of different geometries. The conductivity in the test structures is monitored as the material is removed. From this the 2-D dopant profile may be extracted using a computer fitting program. This technique has been successfully applied to the measurement of doping levels as low as $1\times10^{17}cm^{-3}$, with a resolution of 20nm×20nm. Asymmetry in profiles obtained from implants at an angle to the wafer surface normal have been obtained under different annealing conditions. The asymmetryin the resulting 2-D dopant profiles convincingly demonstrates this technique [8]. However this technique again requires the fabrication of special device structures. This technique is suitable for investigating variations in general processing conditions such as implantation angles, whilst the 2-D SIMS profile is probably more suitable for investigating actual device type structures. The oxide replica technique is suitable for assessing dopant geometry but not dopant profiles and requires the use of a STEM for the best resolution.

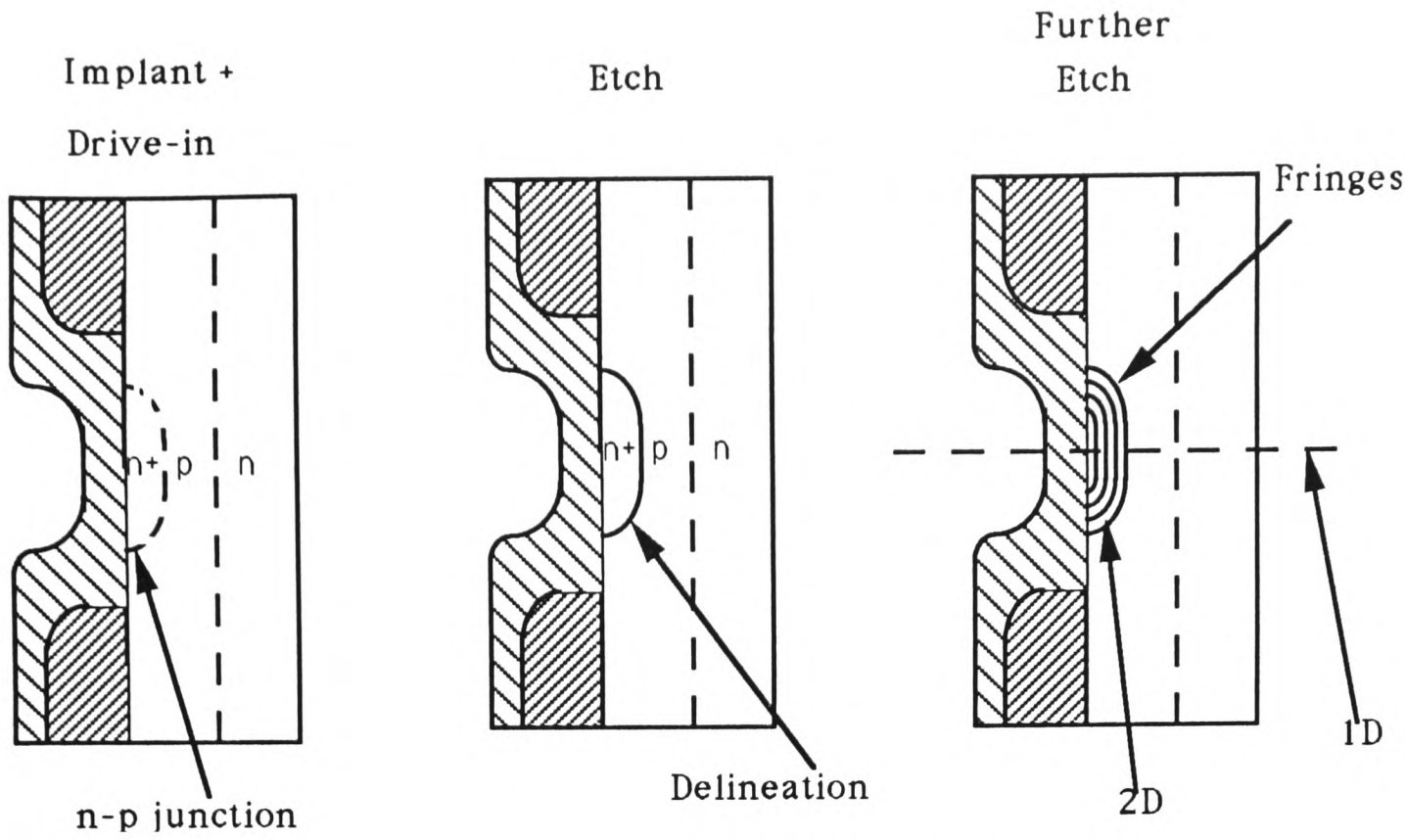
5.2 2D ETCH PROFILING

Marcus and Sheng [9] produced an alternative process for dopant delineation. They showed that an aqueous etch of HNO_3 containing 0.5% HF (by volume) at room temperature, was sensitive to the distribution of n-type dopant in silicon. In their work, a cross-sectional TEM sample was produced from a device-like n-type doped region of a silicon wafer. This sample was then examined using diffraction contrast in the TEM which produced an image of the

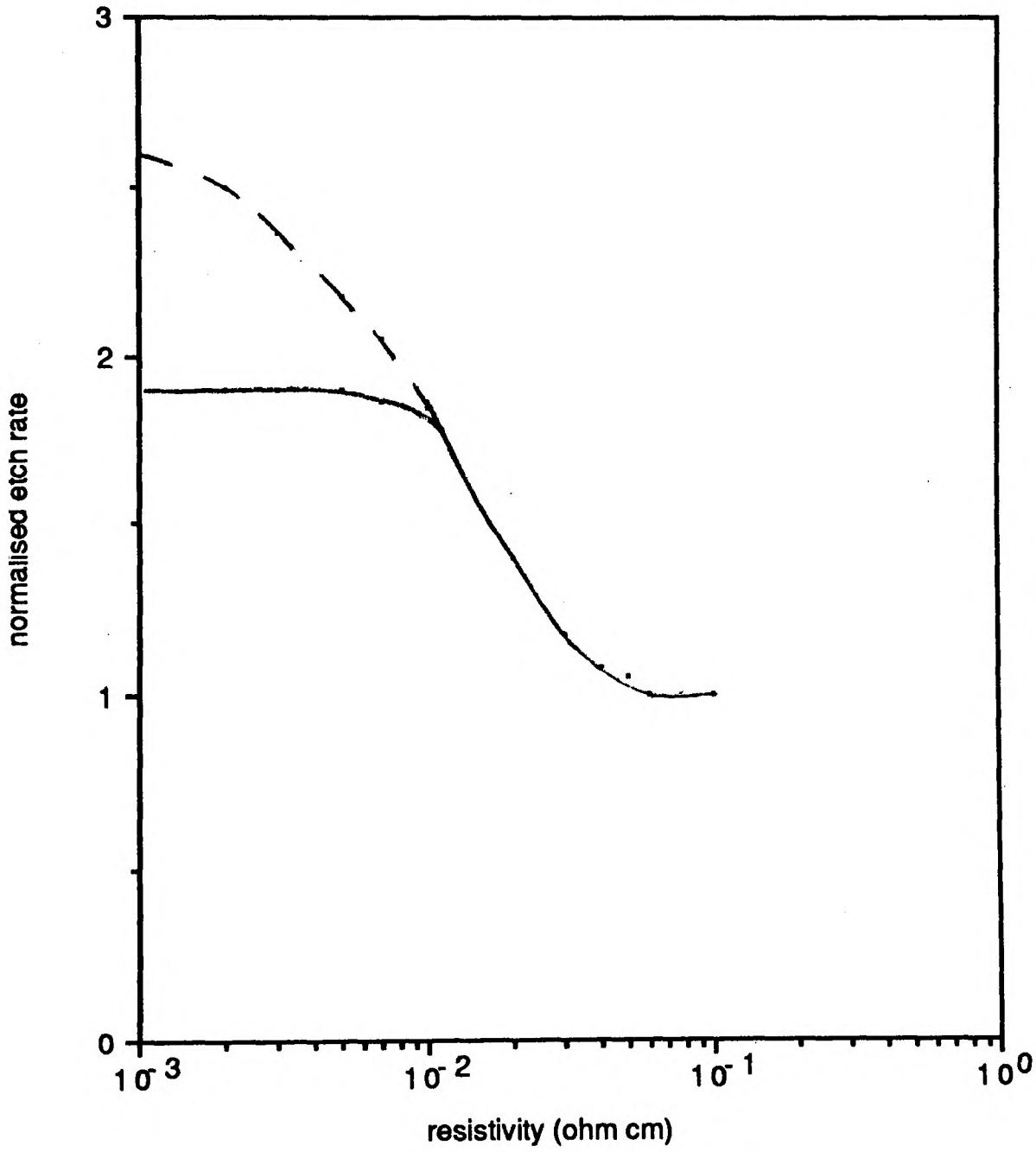


Calibration curves of Roberts and Wills

Fig.5.4



(a)TEM etch technique



(after Roberts)

(b)0.3% etch rate calibration curve

Fig.5.5

etched region, revealing a series of thickness fringes above a delineating line, below which the sample appeared to be of uniform thickness. By comparison with a computed model of dopant distribution, they deduced that this delineation occurred at approximately 10^{19} cm^{-3}. In their sample this delineation occurred at a depth of approximately $0.15\mu m$.

It is important to understand that this does not imply that the background (i.e.undelineated) section of the sample does not etch, but that it has a uniform or near uniform etch rate. Wills applied essentially this technique to the examination of n-p junctions in the scanning electron microscope (SEM), by etching bulk specimens [10]. He examined junctions approximately $1\mu m$ deep and used stereoscopic techniques to calculate the quantity of material removed at any point within the delineation. Roberts [11] [12] carried the work of Sheng and Marcus on TEM foils further, and measured the bulk etch rate for various, well characterised dopant concentrations. He produced a plot of etch rate against dopant concentration for the same etch as Marcus and Sheng and showed that the etch rate tended to a uniform value for low doping concentrations. This background etch rate was taken as the base line for the calculation of a normalised etch rate. The normalised etch rate (n.e.r) was calculated by dividing the etch rate for a given concentration by this background etch rate. The curve produced from this for normalised etch rate against resistivity, for an 0.5%HF/99.5%HNO_3 etch is shown in fig.5.4. Also included in fig.5.4 is the calibration curve of Wills for the same etch.

Roberts first proposed using these fringes to produce a 2-D dopant profile, as shown in fig.5.5a. He found[16] that an etch containing 0.5% HF was too rapid for accurate use with TEM foils. In these cases, considerably less material was to be removed than in the bulk samples, and the termination time between removal from the etch and rinsing in distilled water was a significant fraction of the total etch period. To minimise this source of error, Roberts cooled the etch to 5°C using an ice filled, thermostatically controlled water bath. Cooling the

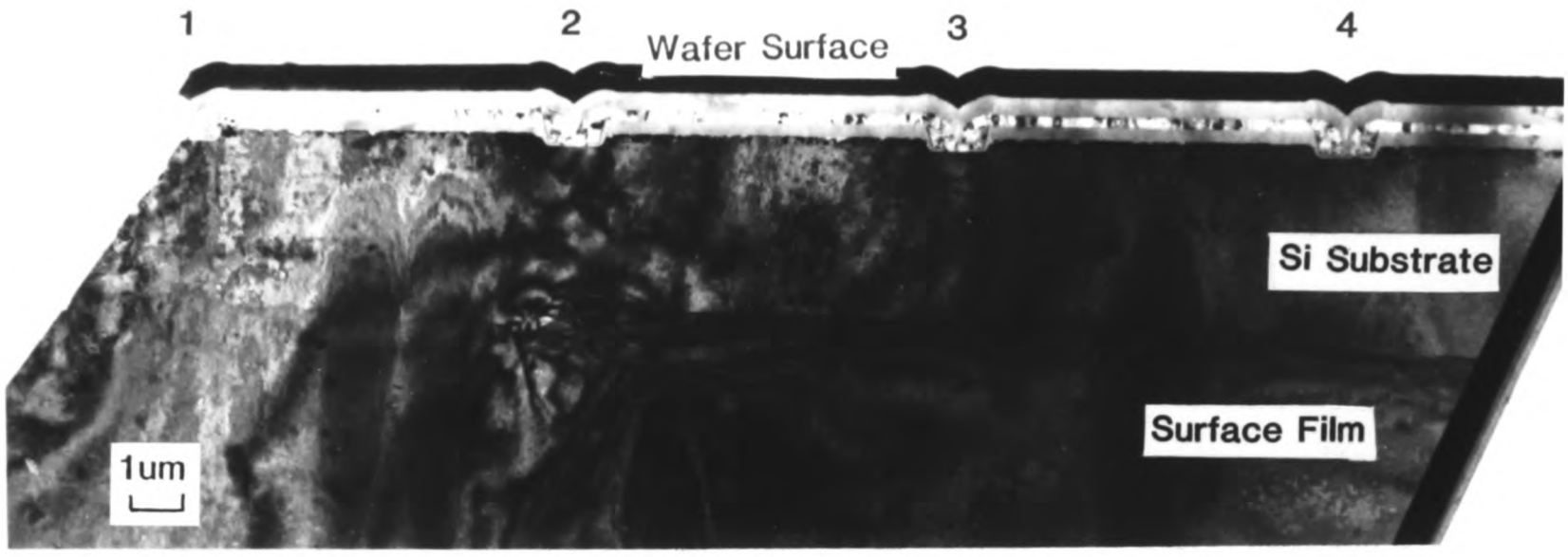


Fig5.6a XTEM of Foil with Emitters

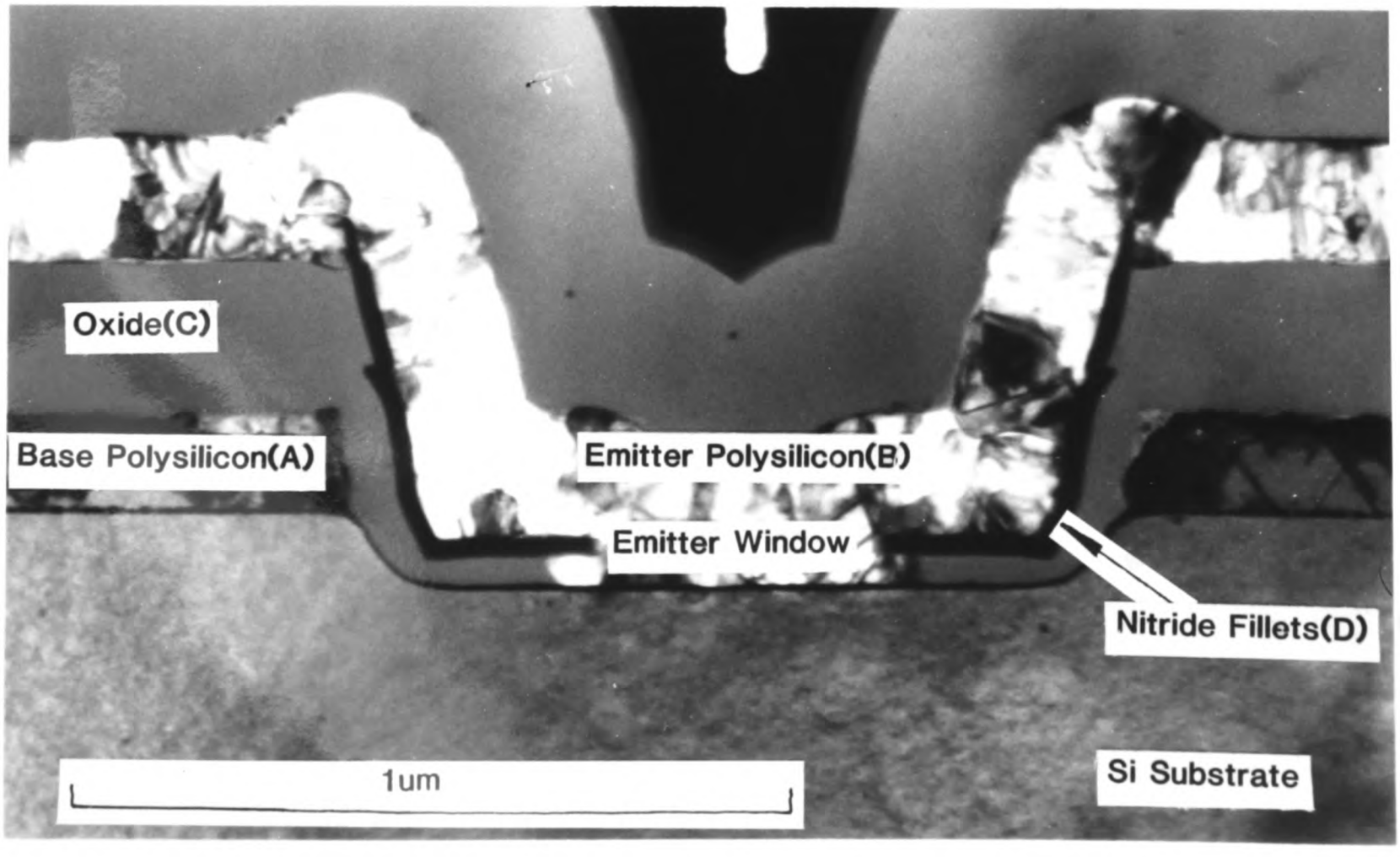
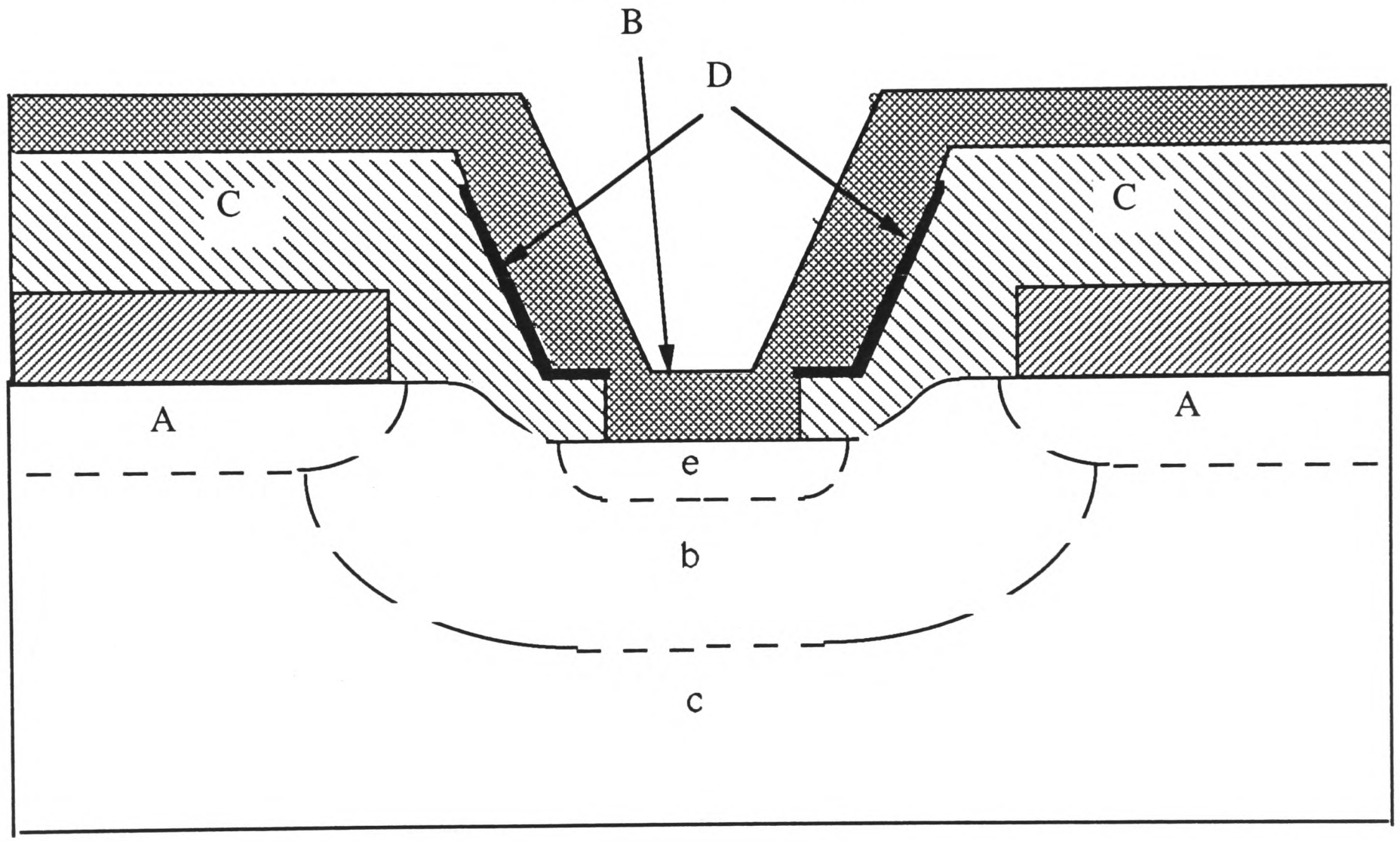


Fig.5.6b XTEM of Emitter

A=Base polysilicon
B=Emitter polysilicon
C=Oxide isolation
D=Nitride isolation



Schematic of a polysilicon contacted bipolar transistor

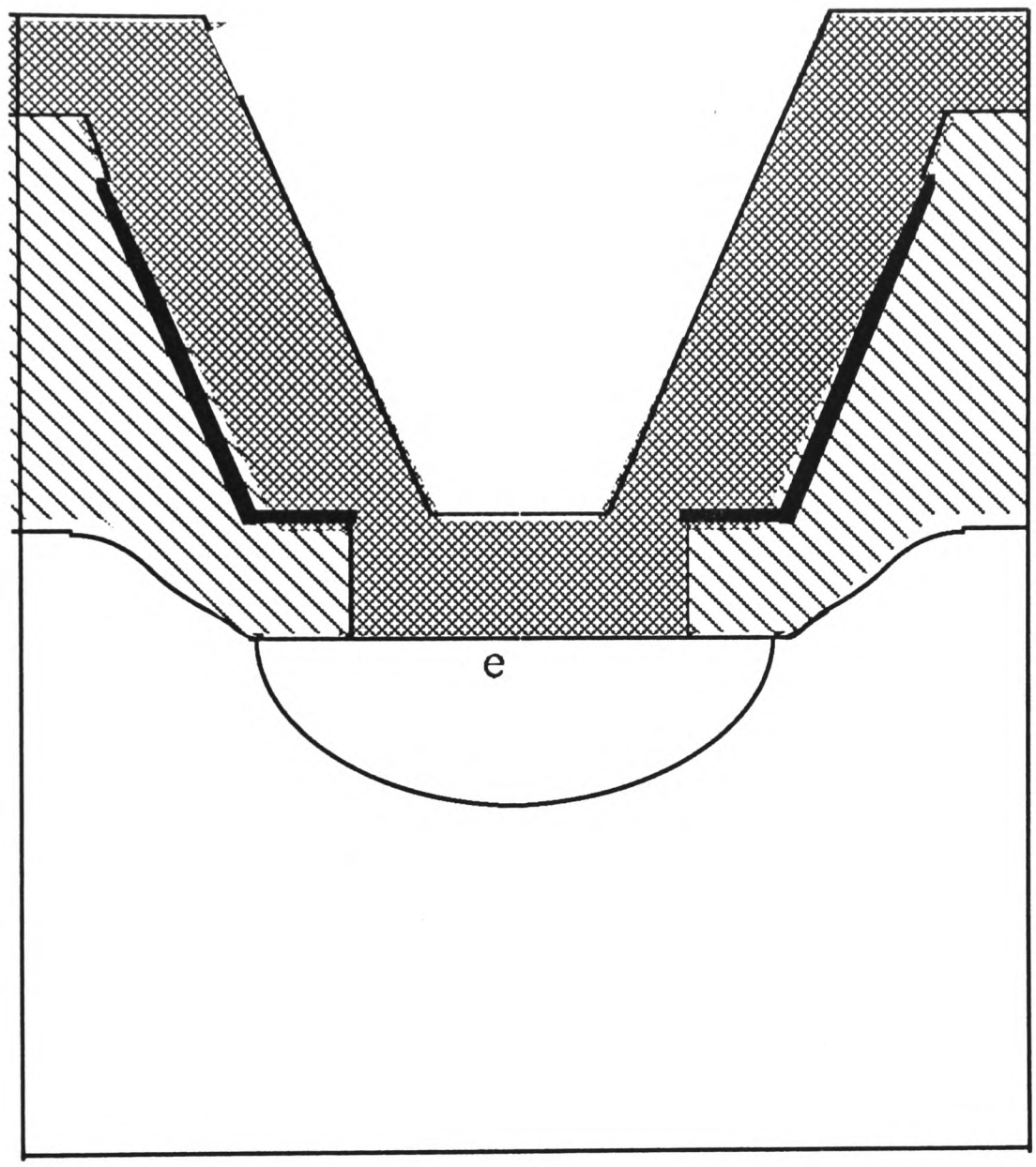
Fig.5.7

solution however, not only slows the etch down but also reduces the sensitivity (i.e. rate of change of etch rate with dopant concentration). However, Roberts had shown that reducing the concentration of the HF also slowed the etch down but in addition resulted in an increase in sensitivity to bulk resistivity. Therefore the final etch chosen by Roberts for TEM etching combined these two effects to produce an etch of only marginally different sensitivity to that of Sheng and Marcus, but with an etch rate of approximately half the original rate. The etch chosen was 0.3%HF and 99.7%HNO_3 cooled to 5°C The plot of n.e.r against dopant concentration produced by Roberts for this etch is shown in fig.5.5b and this calibration is used throughout the present work to deduce doping profiles from etch measurements.

5.3 RESULTS AND DISCUSSION

5.3.1 DEVICE STRUCTURE AND EMITTER GEOMETRY

Fig.5.6a is a (110) bright field, low magnification micrograph showing a typical array of emitters on an XTEM thin foil as used throughout this chapter. A surface film of polymerised lacomit can also be seen in the lower right hand of the foil Fig.5.6b shows a (110) bright-field (bf), many-beam (mb) cross-sectional TEM micrograph of an emitter structure. This is a suitable device for etch delineation, as the foil surface is reasonably flat. This is important, as the distribution of thickness fringes in the delineated region would otherwise vary. The techniques of device sample preparation are fully discussed in chapter 2. Comparison with the schematic of fig.5.7 clearly reveals the reproduction of the proposed design in a real device. A shows the polysilicon base



Emitter shape proposed by Hwang et al

Fig.5.8

SP/HF pre-clean
polysilicon deposition, 2000A
Arsenic implant 40kV 1E16cm-2
Emitter drive-in RTA 27s. 1100C

Table 5.1 emitter processing, 0.133um emitter

SP/HF pre-clean
polysilicon deposition, 2000A
Arsenic implant 40kV 1E16cm-2
Emitter drive-in RTA 26s, 1075C

Table 5.2 emitter processing 800A emitter

diffusion sources/contacts, B shows the polysilicon emitter source/contact, C shows the oxide isolation and D shows the nitride isolation sidewalls. It can be seen from fig.5.6b that the minimum lithographically defined feature size is the emitter aperture through which the isolating nitride sidewalls are grown. This aperture is approximately 1μm wide. The emitter window width can be measured from this micrograph and is approximately 0.4μm. The base contact polysilicon can be seen to have regrown leaving a monocrystalline layer containing regrowth twins. In the corners at the ends of the base polysilicon regions, the monocrystalline regrowth front can be seen. This illustrates that the base contact regrowth is only just complete at the end of normal device processing.

The shape of the dopant distribution in polysilicon contacted high speed emitters has been discussed recently by Hwang et al [13]. They suggested that small scale emitters may not have a planar e/b junction, but a dish shaped e/b junction as shown in figure 5.8. The work is a theoretical prediction based upon the concept that the polysilicon/monocrystalline silicon (monosilicon) interface has a diffusion rate similar to bulk monosilicon. Previously it had been assumed that this interface was similar to a grain boundary and therefore that diffusion along it was rapid compared with bulk diffusion rates, thus the interface acts as a nearly infinite dopant source. If this were not the case, it would have a significant effect on device performance as the emitter became shallower.

Samples of high-speed bipolar emitter structures were fabricated as shown previously. The author is indebted to Dr.Peter Hunt and Dr.Martin Wilson at Plessey Research Ltd for the supply of these device structures. The samples were taken from a test wafer specially produced for TEM foil preparation. The emitter processing details are given in table 5.1. The sample contained a series of emitters 2mm long in one dimension, and of normal device size in the other direction. Whilst this simplifies sample preparation, it is possible to produce a TEM cross-sectional foil from a real device using modern techniques such

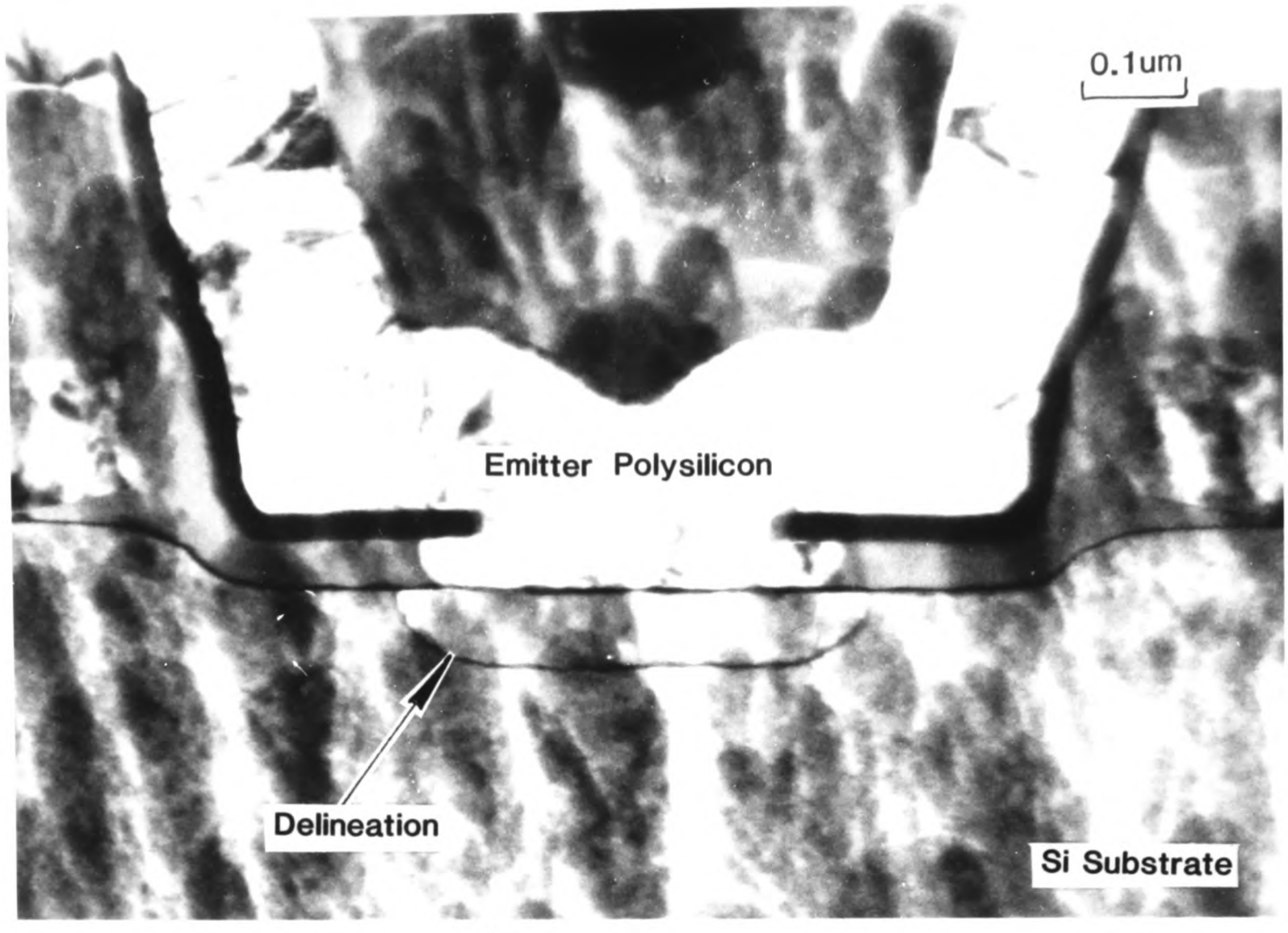


Fig.5.9 XTEM of Emitter after 45s. Etch

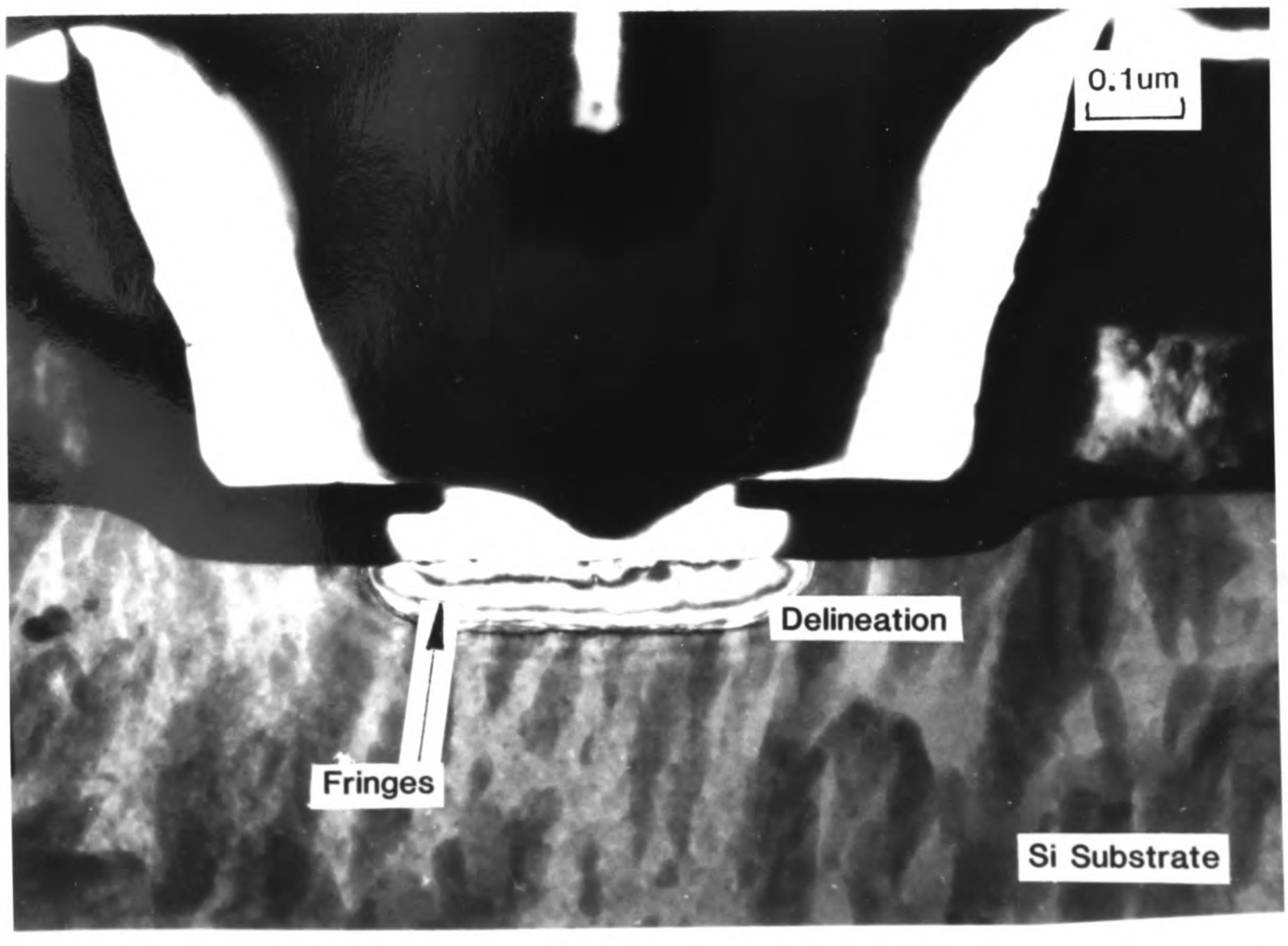


Fig.5.10 XTEM of Emitter after 105s.

as scanning ion milling (often referred to as Precision Ion Milling). TEM cross-sections of emitter regions were produced as described in chapter 2. This sample was examined in a Philips CM12 S-TEM operating at 120kV. Samples were left in the microscope for at least half an hour, with the cold trap full, before turning the filament on, to ensure that the contamination levels were as low as possible. This was carried out in an attempt to reduce the polymerisation of hydrocarbon contaminants on the surface of the sample by the electron beam. A polymerised film was considered to be a possible barrier to the etch.

The first sample investigated has been shown in fig.5.6b Unlike the approach taken with later samples, this sample was glued to the holder with Lacomit so that one side of the sample was protected from the etch by the Lacomit. The sample was first etched for 45 seconds, and then examined again in the CM12. This is illustrated in fig.5.9, which is a (110) mb bf image of the sample. The mottled contrast on the surface is caused by part of the polymerised layer of Lacomit formed during ion milling, which was still adhering to one face of the XTEM sample. It can be seen that the device structure is still intact. Below the emitter window, an area of lighter contrast is visible. This is the doped region of the sample which has been differentially etched. The depth of the delineated front is 0.072μm from the interface.

The shape of the emitter region is clearly defined and it can be seen already that the shape of the emitter base junction (sometimes referred to as the emitter front) is flat bottomed with nearly circular section ends. This is the shape expected based upon the assumption that the polysilicon/monosilicon interface has diffusion rates similar to grain boundaries. Thus it is clear already that the assumption of Hwang et al does not match experimental observation[4].

Following these observations, the sample was remounted as before and etched for a further one minute. The TEM image of the same emitter, imaged under the same conditions as detailed above, is show in fig.5.10. The fringes now extend approximately 0.13μm from the interface, with a lateral spread of

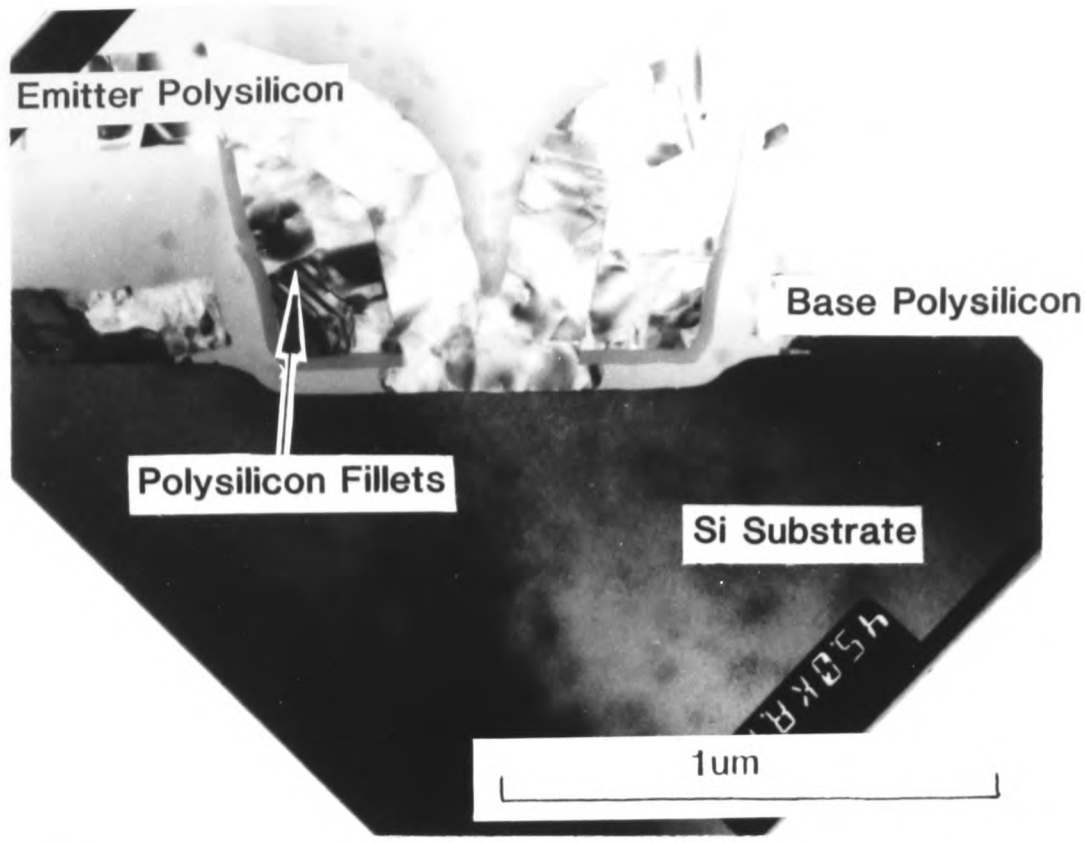


Fig.5.11 XTEM of Emitter with Polysilicon Fillets

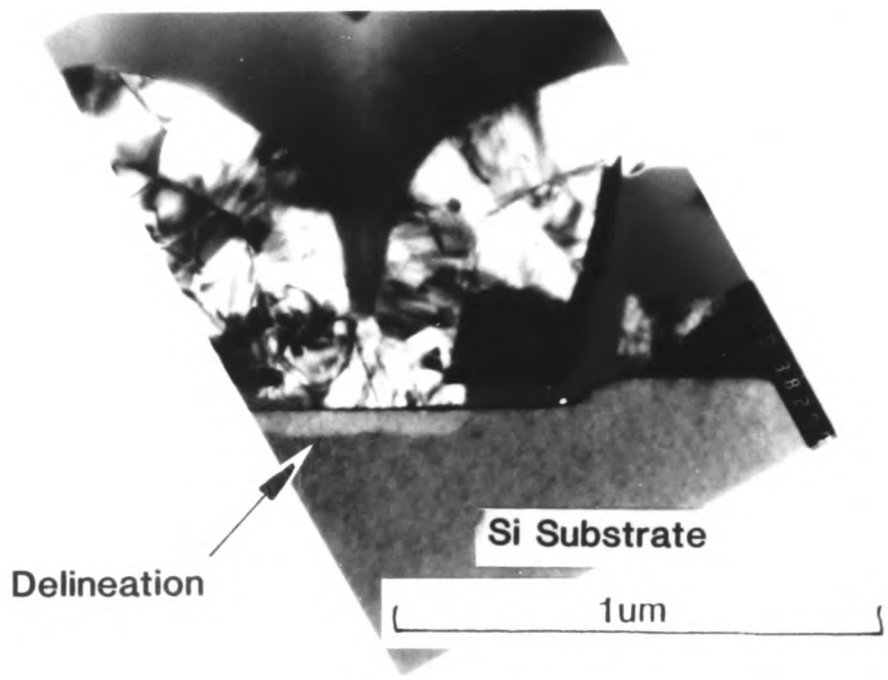


Fig.5.12 XTEM of of Emitter after 45s. Etch

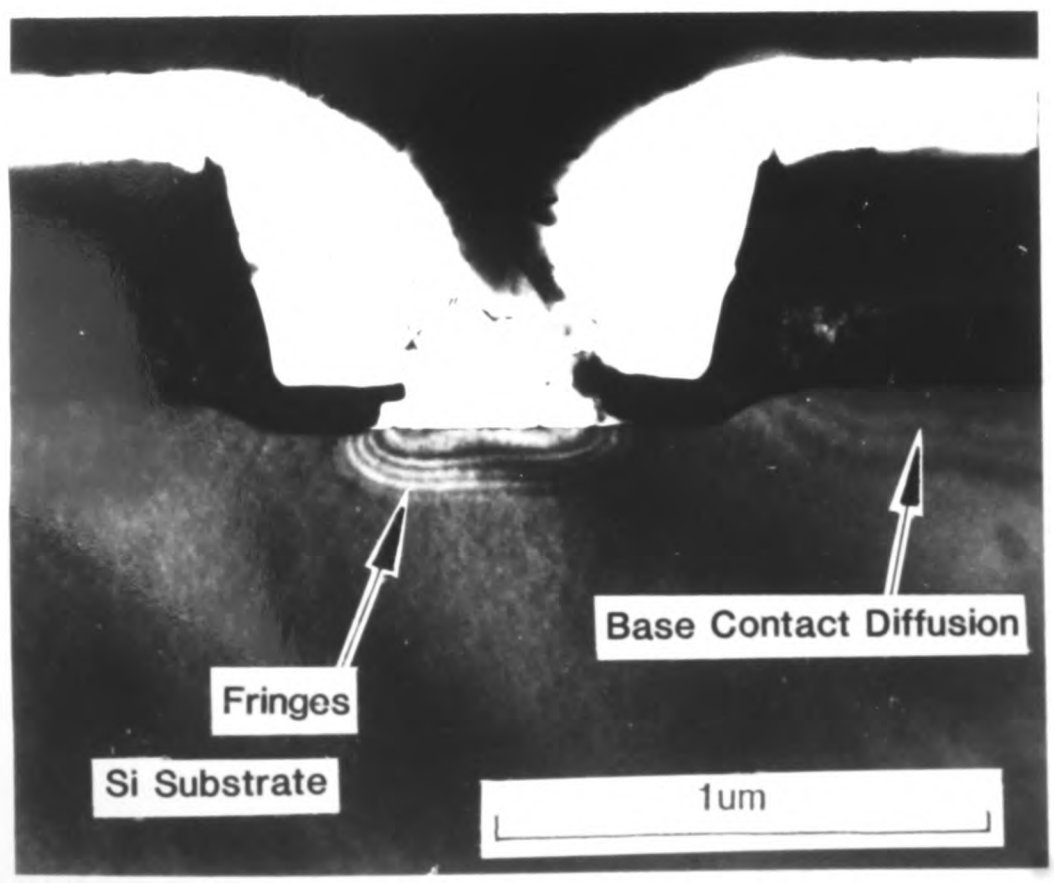


Fig.5.13 XTEM of Emitter after 105s. Etch

approximately 0.1μm beyond either side of the polysilicon contact pad. The sample can now be seen to have a series of fringes within the delineated region. These are thickness fringes produced by the interaction of the diffracted beam Bloch waves with each other and with the main beam Bloch wave. For 120kV electrons in silicon the depth separation of these fringes is calculated to be 24.5nm for the (110) BF condition. The shape of the emitter is planar, as before. Examination of the base-contacting polysilicon region shows a series of fringes here also. These are due to the etching of the diffused base contact (p^+ in the monosilicon). Close to the edge of the sharply defined region in the emitter, the fringes are very close together. Because of this, the exact number could not be counted, and therefore measurement of the etch rate and hence the doping profile was not carried out on this device.

The work of Hwang et al, highlights the potential hazards which may result from the scaling of device dimensions. To investigate the effects on the shape of the e/b junction resulting from changes in the structure of the device, a second set of samples was fabricated. This was identical to the set previously examined except that the polysilicon fillets which protect the sidewalls during fabrication, as shown in fig.5.1, were not removed. A cross-sectional TEM sample was produced as before and a device from this is illustrated in fig.5.11. The polysilicon fillets can be easily identified. This sample was etched on the same Teflon disc at the same time as the previously investigated sample. This allows a direct comparison of the two samples to be undertaken.

The (110) mb bf image of the sample after a 45 second etch is shown in fig.5.12. Although the general shape of the delineated region is similar to that previously observed, it can be seen that the emitter front is not entirely planar. An inverted dish shape can be seen in the centre of the planar section of the front. During the subsequent one minute etch, the emitter examined above broke away from the foil and was lost. Fig.5.13 shows a similar device on the same foil after a total of 1 minute 45 seconds in the etch. This emitter which was imaged

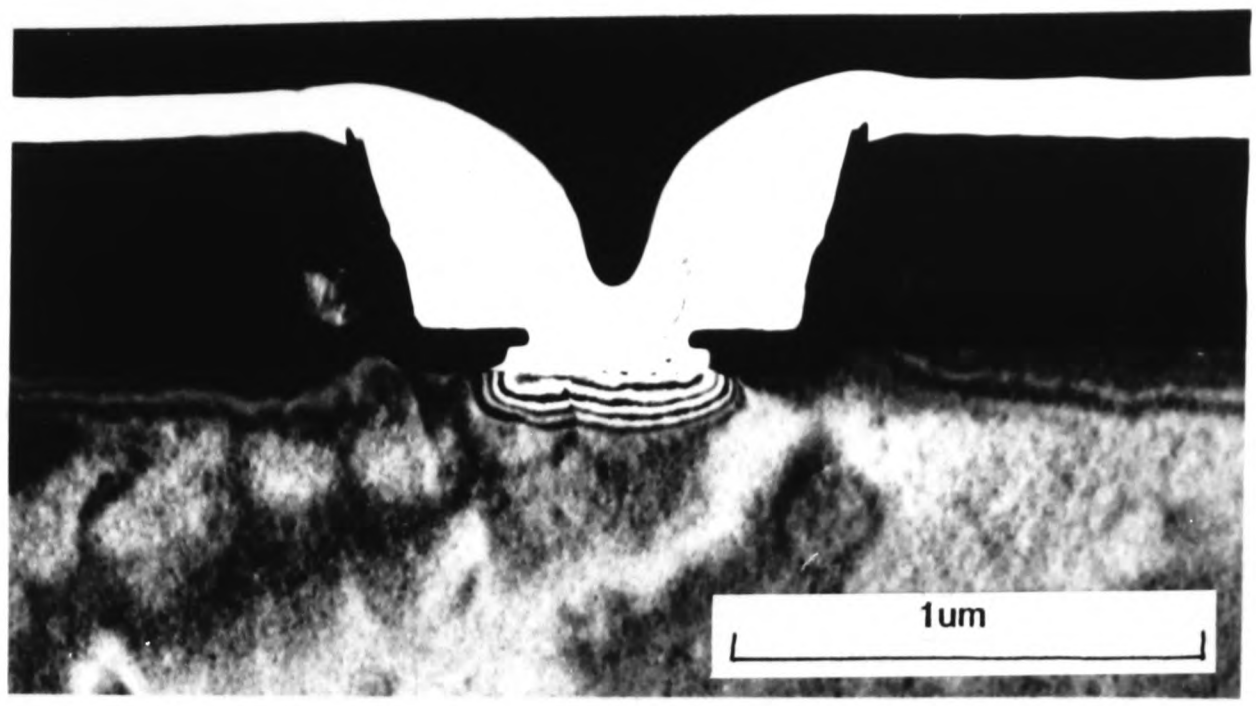


Fig.5.14a XTEM Emitter 105s.Etch (110) BF

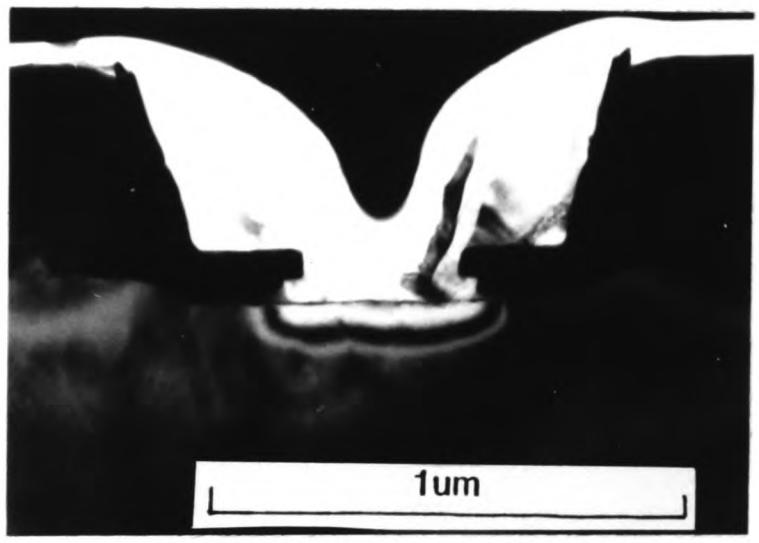


Fig.5.14b (220)

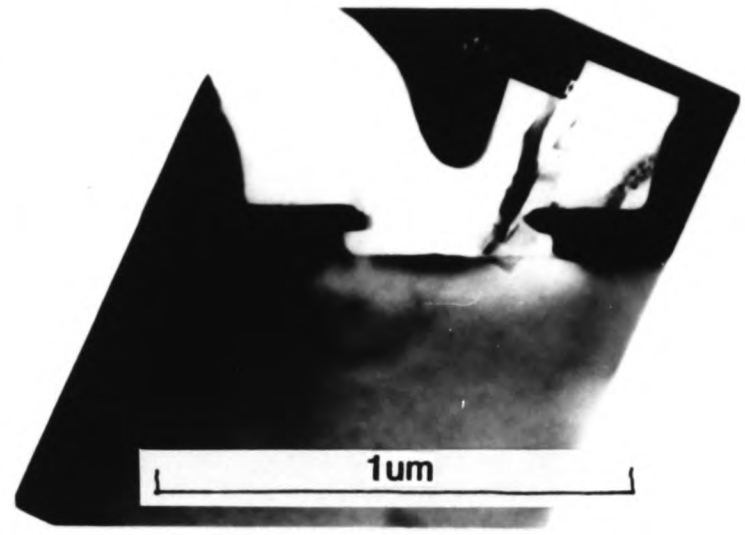


Fig.5.14c (400)

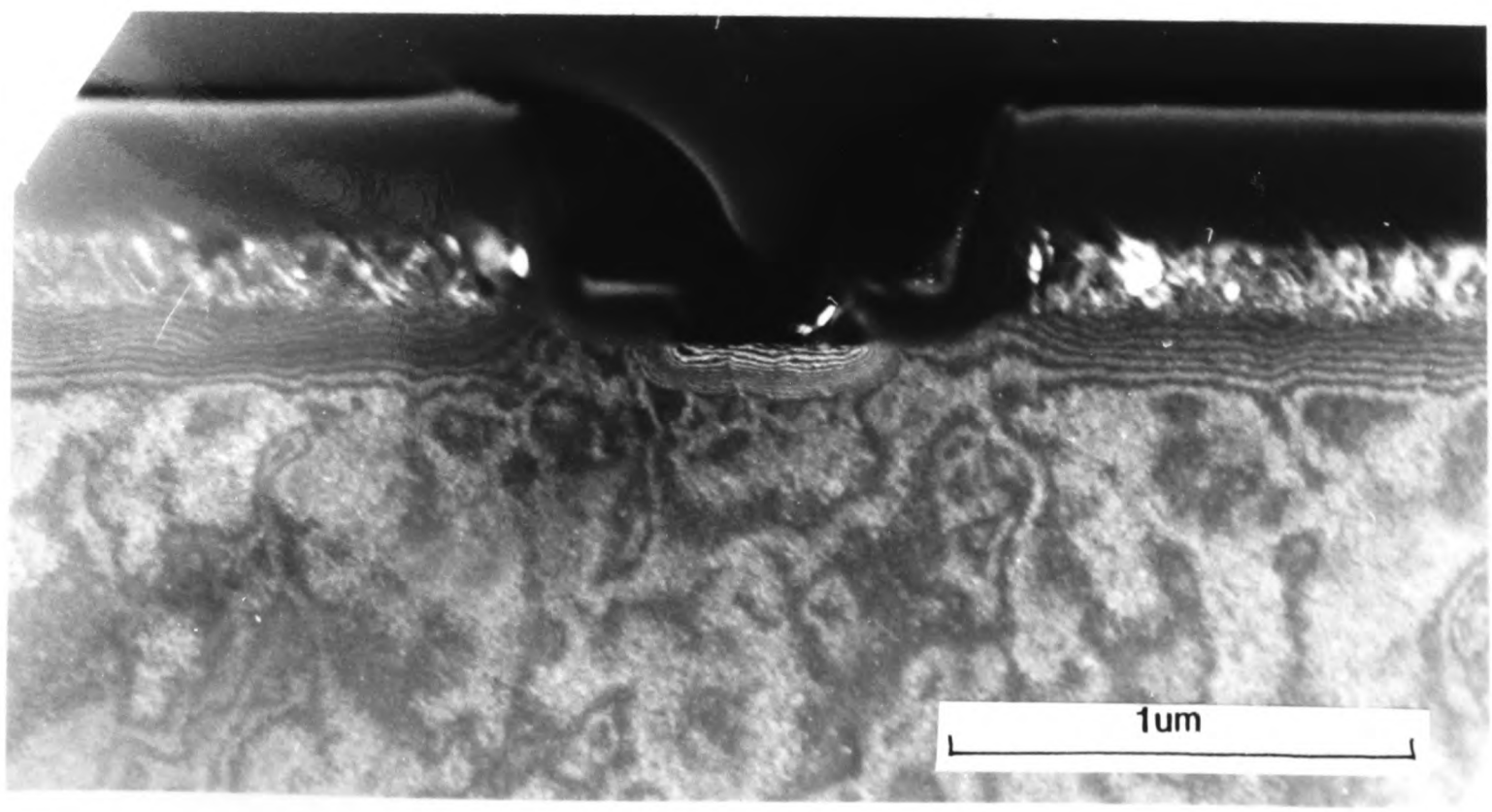


Fig.5.14d g3g (220) Weak Beam

in a relatively thick area of the foil also shows a variation in planarity of the emitter front, with a depth of delineation of approximately 0.15μm in the centre of the emitter, and a lateral spread under each side of the polysilicon contact pad of approximately 0.1μm. The centre of the emitter region is 0.01μm closer to the contact than the outside edges. Once again the diffused base contact is clearly visible.

The change in thickness of material corresponding to adjacent fringes observed can be varied over a range of values by altering the imaging diffraction conditions. This effect can be used to measure the thickness profile of samples with differing wedge angles. Fig.5.14a illustrates an emitter fabricated with a polysilicon fillet, from the foil detailed above. This emitter in a relatively thin area of the foil allows good contrast from a variety of imaging conditions. However, it can be clearly seen that this emitter contains a defect. This is believed to be a dislocation or crack produced during specimen handling.

Fig 5.14(a-d) illustrates an emitter region from the same foil under a variety of imaging conditions. Clearly it is possible to go from a large extinction distance (a 400 reflection, fig.4.14c), to a small extinction distance (g3g weak beam fig.4.14d), however, it is important to bear in ming that tilting the specimen will alter the observed geometry. Suitable imaging conditions may therefore sometimes be achieved, even when the sample has not been etched for the optimal time for (110) mb bf imaging.

The difference in behaviour between the sample containing the polysilicon fillets and the 'normal' sample is interesting and illustrates the type of application in which high resolution delineation may be useful. However, it is important to consider carefully the potential causes of variation of etch rate. Obviously, the effect may be caused by a real variation in the doping profile. Alternatively some other effect, e.g. stress, may be important. During refinement of the etch technique, various surface cleans were applied to the foil to attempt to provide a strictly repeatable surface to the etch. Amongst these, an aqueous solution of

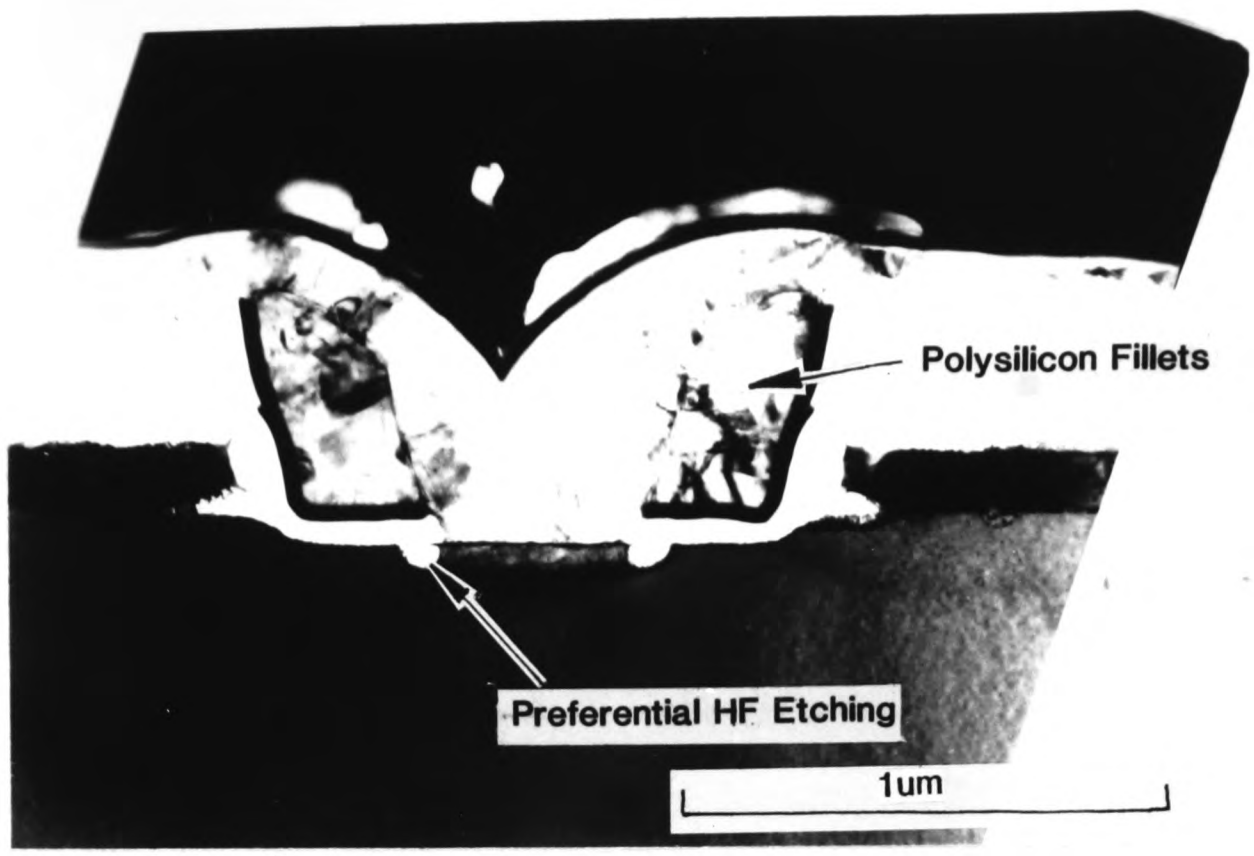


Fig.5.15 XTEM of Emitter after HF clean

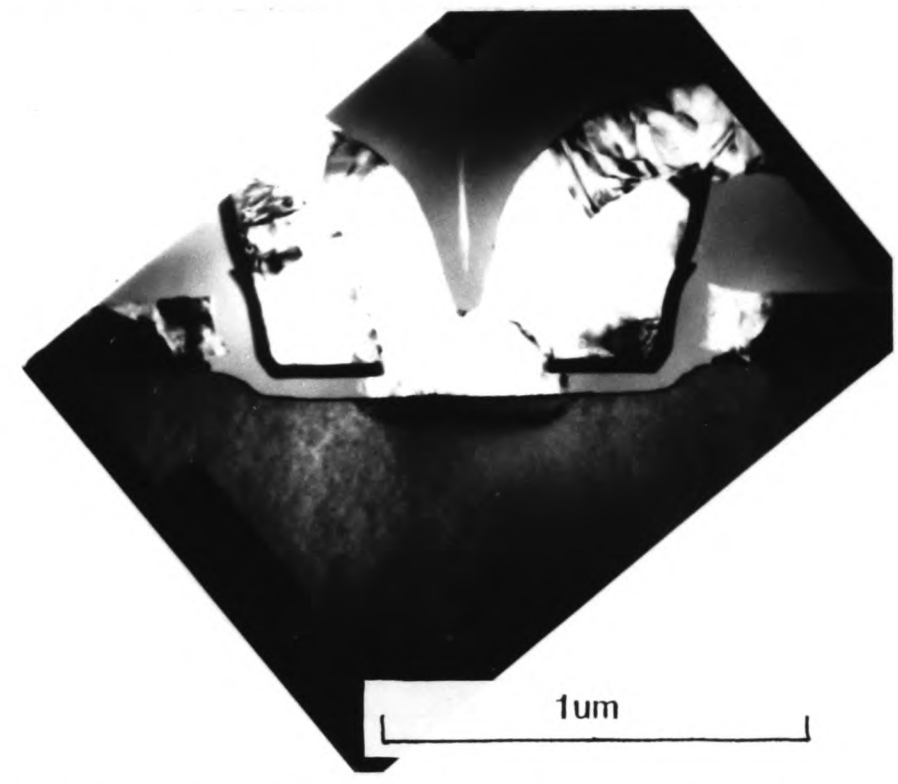


Fig.5.16a XTEM of Shallow Emitter

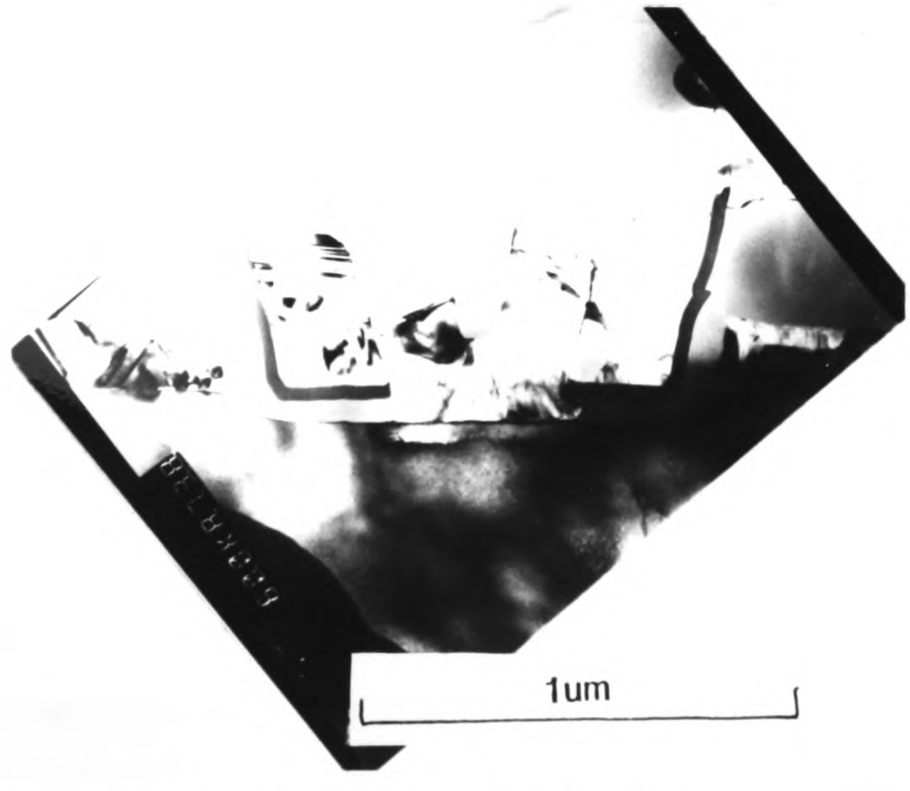


Fig.5.16b XTEM of Shallow Emitter

HF (40% by volume) was used, the samples being immersed in the solution for approximately 30 seconds to 1 minute.

Fig.5.15 illustrates a sample etched for 1 minute, followed by the above HF clean. The surface oxide has been removed during the clean, however the delineated region can be clearly seen. The micrograph clearly shows that the edges of this delineated region have been removed during the clean. This experiment was repeated on three other samples, and the same effect was observed in all cases. This suggests that other effects in addition to the dopant profile may influence the local etch rate. The most reasonable explanation seems to be that stress from the polysilicon fillets is responsible. This stress may result from differential thermal expansion during cooling down, or perhaps from the change in lattice constant which occurs in heavily doped silicon, or indeed a combination of the two. Care must therefore be exercised in the interpretation of etch results.

It can be shown therefore that differences may occur between similar samples of only slightly different structural configurations. If this difference is due to a non-planar doping distribution, then it will have a significant effect on the scaling of emitter depths in future generations of high speed bipolar devices. If the effect is due to local stresses set up in the surface structure of devices, then this may limit adhesion of small scale surface features as scaling of these device features occurs.

To investigate the effect of scaling the emitter depth, a device was fabricated using identical mask dimensions and including the polysilicon fillets as previously. The processing details are given in table 5.2. The processing was similar to that used previously, but the emitter drive-in was carried out using a rapid thermal anneal (RTA) carried out by lamp annealing at 1075°C for 26 seconds. Four foils were prepared from this wafer and and these were then etched using the technique described in chapter 2, so that both sides of the samples were etched. A micrograph from two of the foils is shown in fig.5.16. A total of 12 devices were examined in these foils. A delineation could be seen

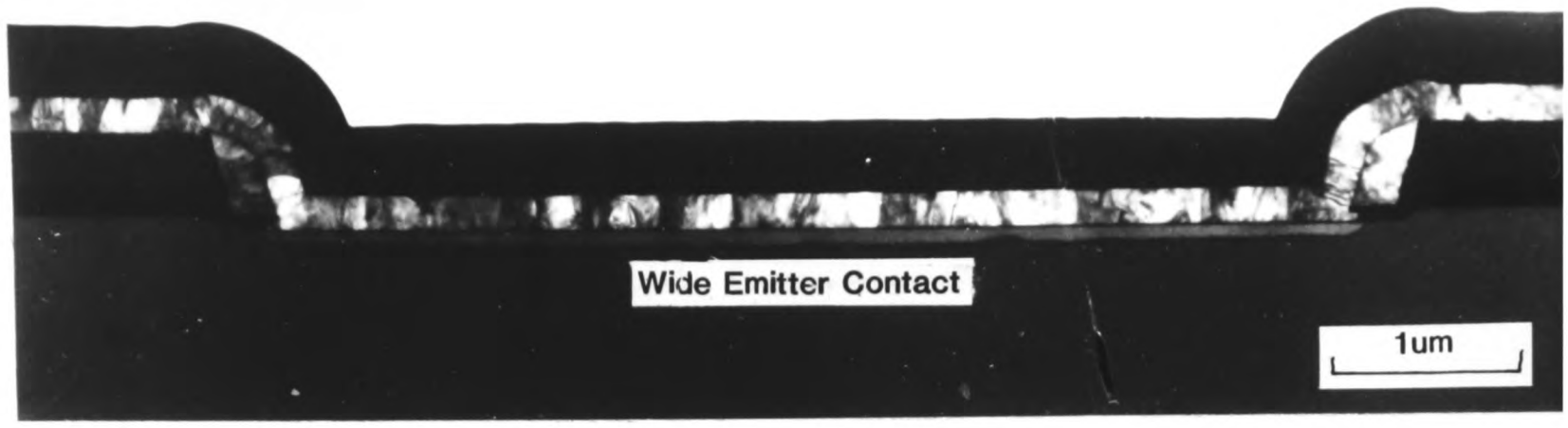


Fig.5.18a XTEM of Shallow Emitter

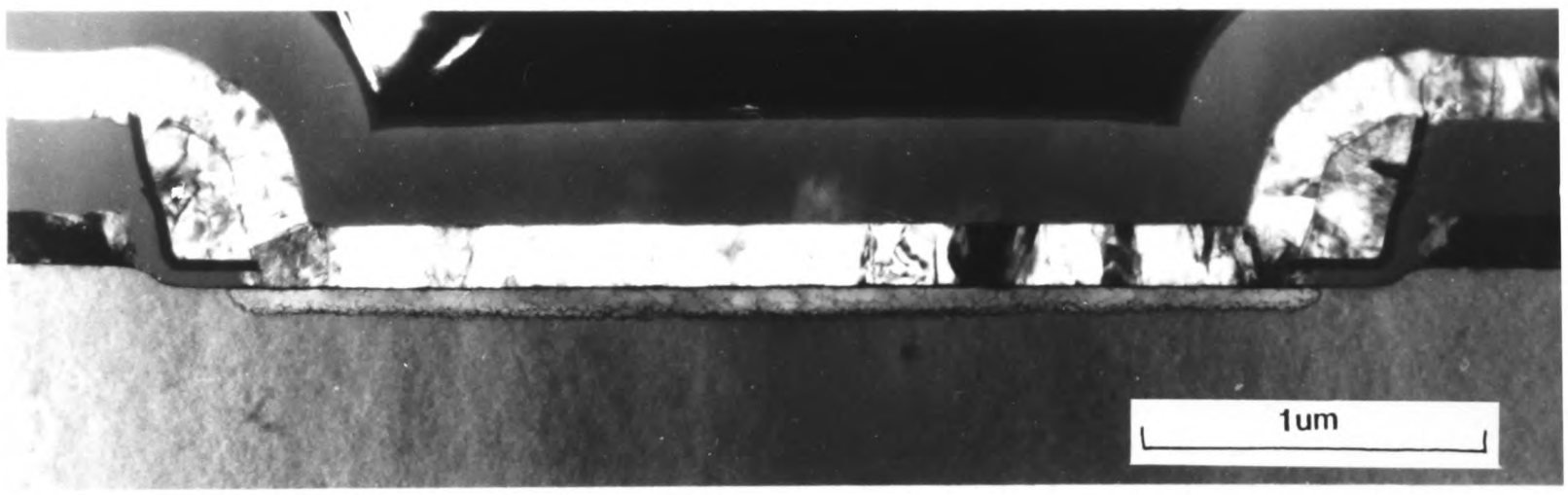


Fig.5.18b XTEM of Shallow Emitter

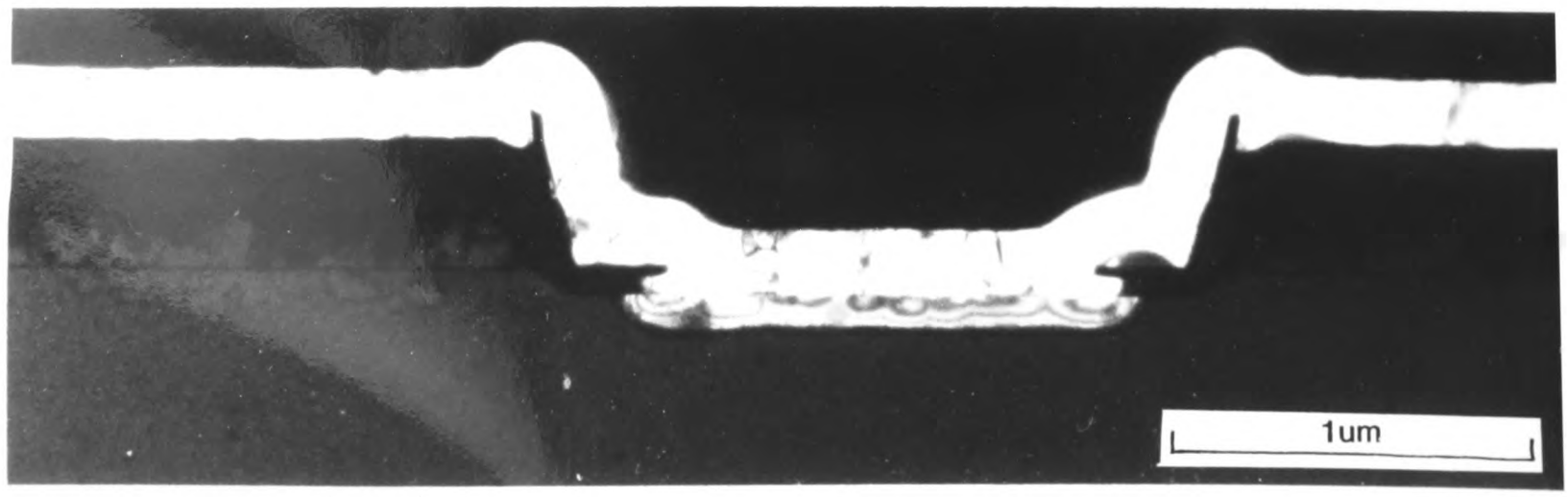


Fig.5.18c XTEM of Shallow Emitter

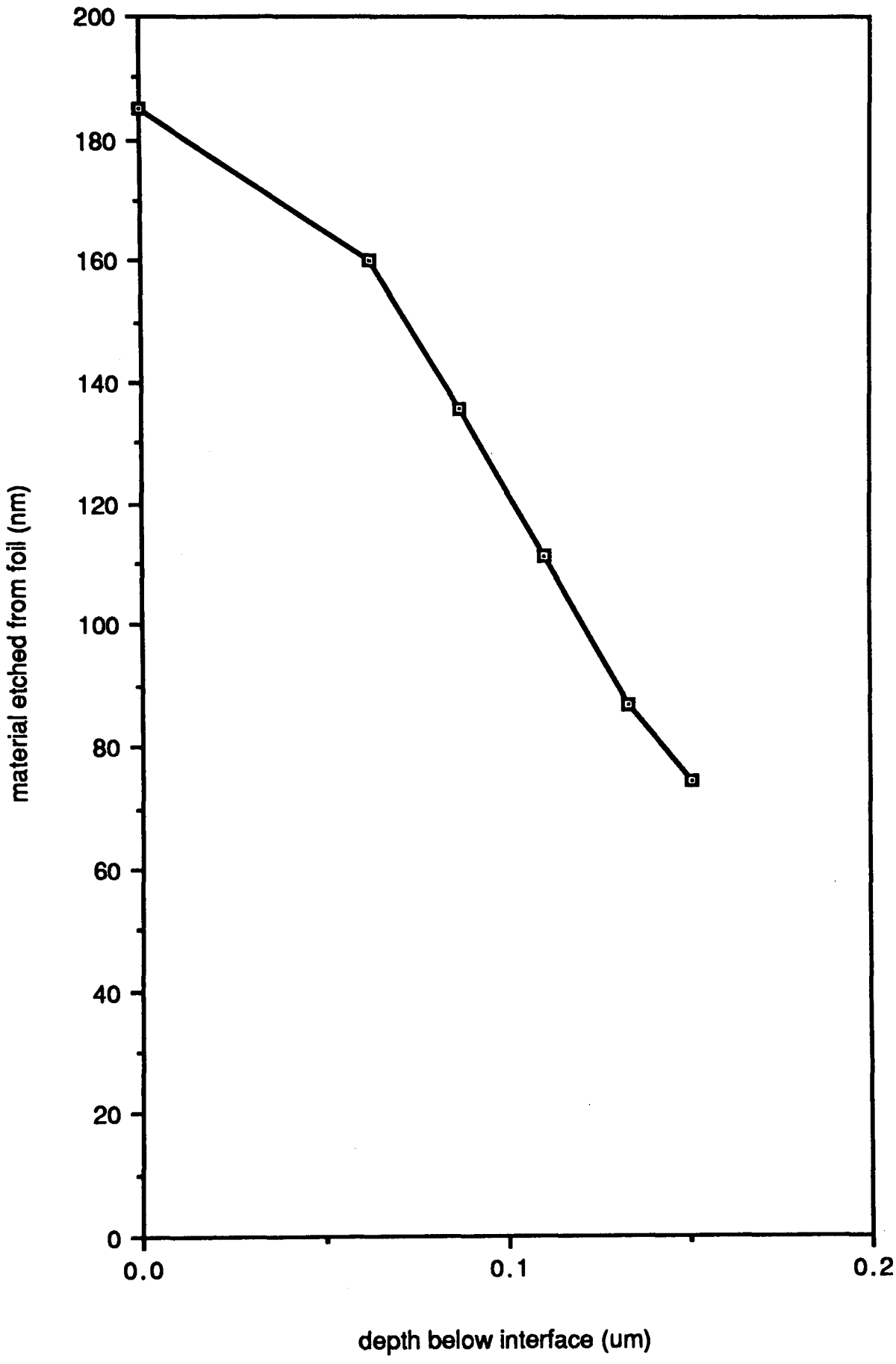
below the polysilicon contact in all samples. Fringes were not observed within the delineated region of any of these devices, this was observed in many etched samples, the reason is unclear. The delineated regions do not all have identical shapes, but all are closely similar, showing the same W shaped delineation as the deeper emitter sample with polysilicon fillets. This confirms the importance of these fillets in the appearance of this shape of delineation. The average depth of the delineation centre is 510Å with minimum and maximum values of 450Å and 580Årespectively. The delineation extended a maximum of 200Å under each side of the polysilicon contact pad. It is not clear from the present work whether the differences in observed emitter depth constitute real differences between the devices or are artefacts of the etching technique.

The junction depth measured by SIMS profiling was approximately 750-800Å. In all cases the experimentally measured value using the etch technique is significantly lower than this for normal emitters. To investigate the effects of device structure on this discrepancy in emitter depth, three foils were prepared of the shallow device structure from regions of the test chip containing wider devices. Such a device is shown in fig.5.18a. All these devices were 1.3μm or more in width measured at the emitter window. The devices were all etched, on both sides, for one minute and then examined in the TEM, a second device of this type is shown in fig.5.18b. The general shape of the delineation shows a curved bottom, deepest in the centre where the depth is approximately 800Å, whilst the edges below the fillets are approximately 500Å deep. 5.18c illustrates a wide emitter structure for an emitter without polysilicon fillets. The emitter delineation in this device is planar along its base, as for the narrower emitter with the same structure. The reduction in delineation depth observed at the edges of the emitter window in fig.5.18b therefore appears to result from the presence of the polysilicon fillets, whether by a direct effect on dopant distribution or by some other effect.

From this it can be shown that the non-planarity observed in the emitter front can be attributed to the presence of the initially undoped polysilicon fillets. In the devices of normal width the effect of these fillets in reducing the measured emitter depth would extend across the whole width of the emitter thus making them appear shallower. A full discussion of the possible causes of the offset measured between the etching results and the SIMS profiles is given on p.115.

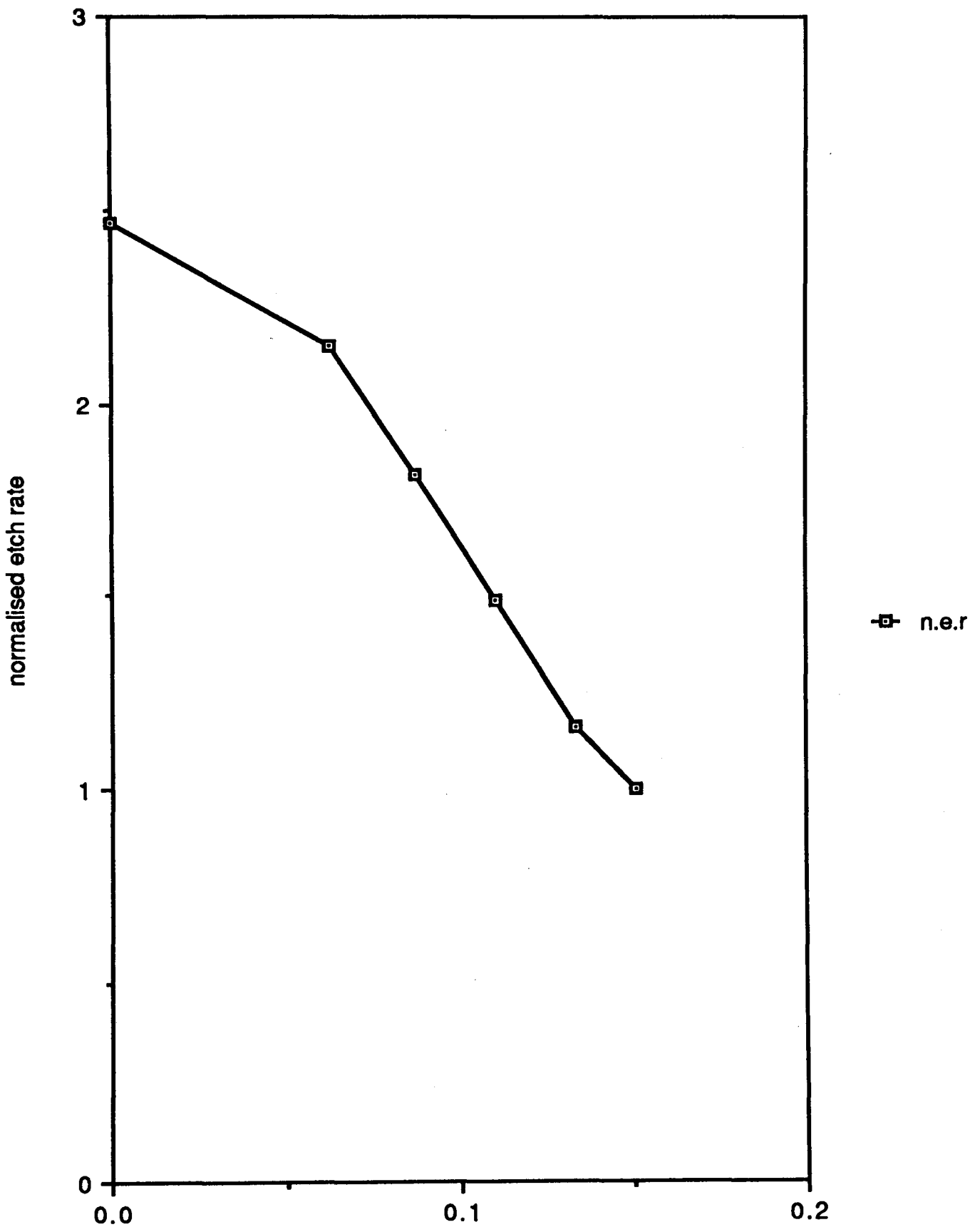
5.3.2 EXPERIMENTAL 1-D DOPANT PROFILING

Having obtained a suitably etched sample, the two dimensional (2D) dopant distribution can be determined. In the present work, this was carried out for both the emitter profile and the base contact profile. Measurement of the complete thickness profile requires certain simple assumptions to be made. Unless a fringe can be clearly seen to be coincident with the position of the polysilicon/monosilicon interface, an assumption must be made about the thickness difference between the first visible fringe and the position of this interface, in the present work, the interface was placed half an extinction distance ahead of the nearest fringe. A similar assumption must be made about the end point of the fringes. The sharpness with which the end point is defined, depends on the exact shape of the doping profile under investigation. If the last visible fringe (light or dark) is clearly defined, then the position of the background level is set one half extinction distance less than the nearest fringe (i.e. the first observed fringe is one half extinction shallower than the background thickness). Positioning of this end point depends upon the exact detail observed in the micrograph. If the last fringe is very weak, then the etch rate is positioned at this fringe. It must be borne in mind that imaging fringes in the thicker regions of a foil may make the last fringe appear weak, even when it is not very close to the background doping concentration (i.e. $2\times10^{17}\mathrm{cm}^{-3}$ or less).



material etched from foil vs. thickness

Fig.5.20(a)

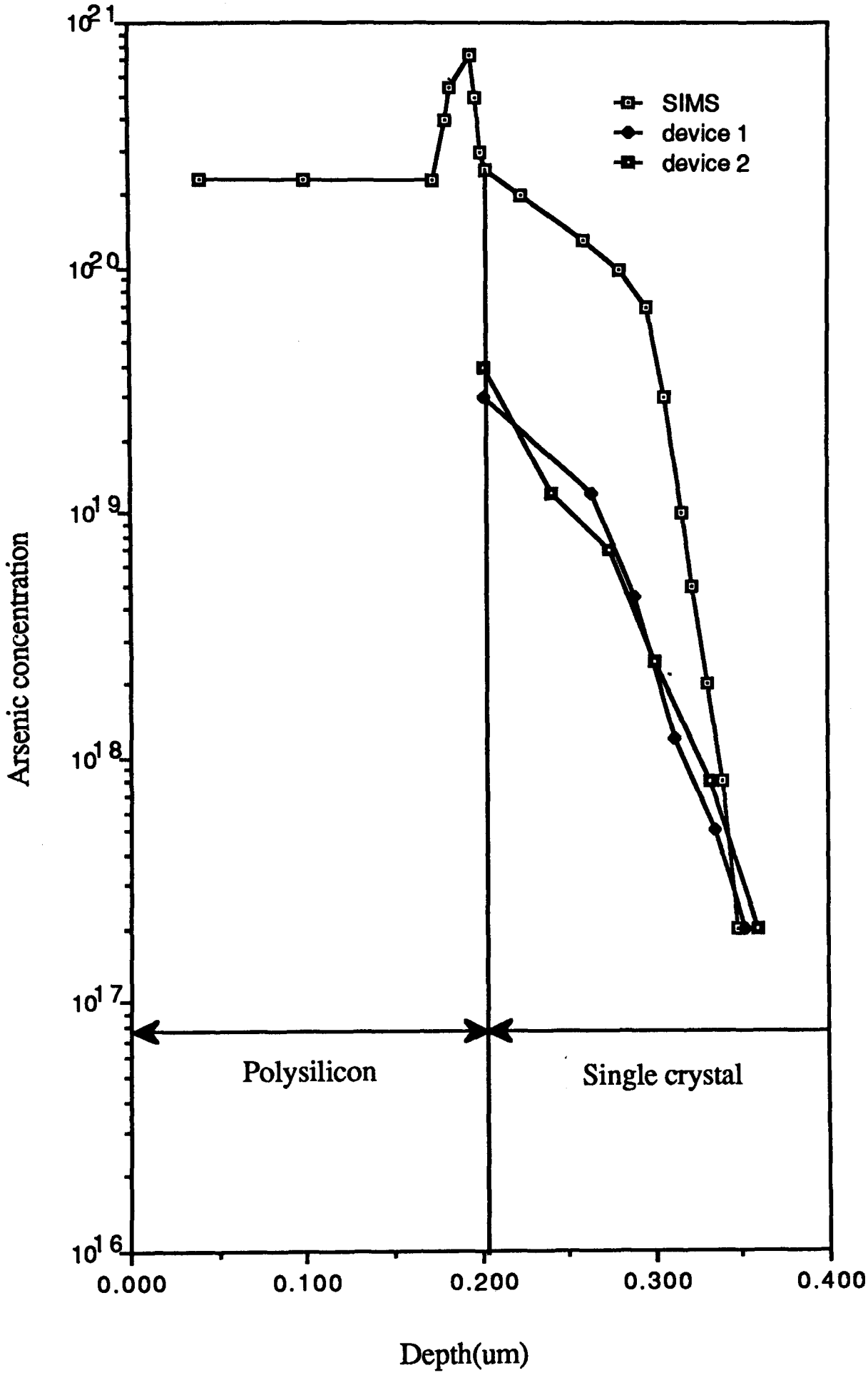


n.e.r. vs. depth

Fig.5.20(b)

Two 1D profiles of thickness were produced, one using the (110) mb bf image of fig.5.13 and the other the (110) mb bf image of Fig.5.14a. Fig.5.19 illustrates the thickness profile obtained from a section through the middle of the emitter in fig.5.13. This device was in a thick region of the foil, and so contrast from the fringes was low. To produce useful fringes, the micrographs were imaged processed using a Semper 6 image processing package to enhance the available detail. The enhanced image is shown in fig.5.24. The only available output device for this type of image was a postscript laser printer. Postscript has a very limited grey scale range and so some detail enhanced by the image processing system is not clearly visible on the reproductions. From this, thickness information could be obtained and processed. The background etch rate taken from Roberts work is assumed to be $0.71 nms^{-1}$. The etch period is 105 seconds, and the etching is assumed to have taken place on one side of the foil only, as the foil was mounted flat in this case, as detailed earlier in this chapter. This gives a total thickness of material removed from the low-doped background of 74.5nm. Using this value, the total thickness of material removed from each region of the 1D profile can be determined. If the etch calibration curve and the background etch rate are accurately known, then the technique is self consistent and requires no external fitting. The light fringes were used for measurement in this case. It was observed that at the edges of the delineation, a dark fringe was just visible, therefore the background etch rate was set at one half an extinction distance less than the first light fringe. The resulting thickness vs. depth profile is illustrated in fig.5.20a.A figure for the n.e.r at each fringe can now be calculated and this is illustrated in fig.5.20b.

From this it is obvious that the n.e.r.s of more than 2.1 have been measured. This is greater than any value on the calibration curve produced by Roberts whose etch curve flattens out at high doping concentrations. This effect is believed by the present author to have resulted from transport-limited etching. Work carried out by Wills[13] on the 0.5%HF etch illustrated that etch



Device etch profiles

Fig.5.21

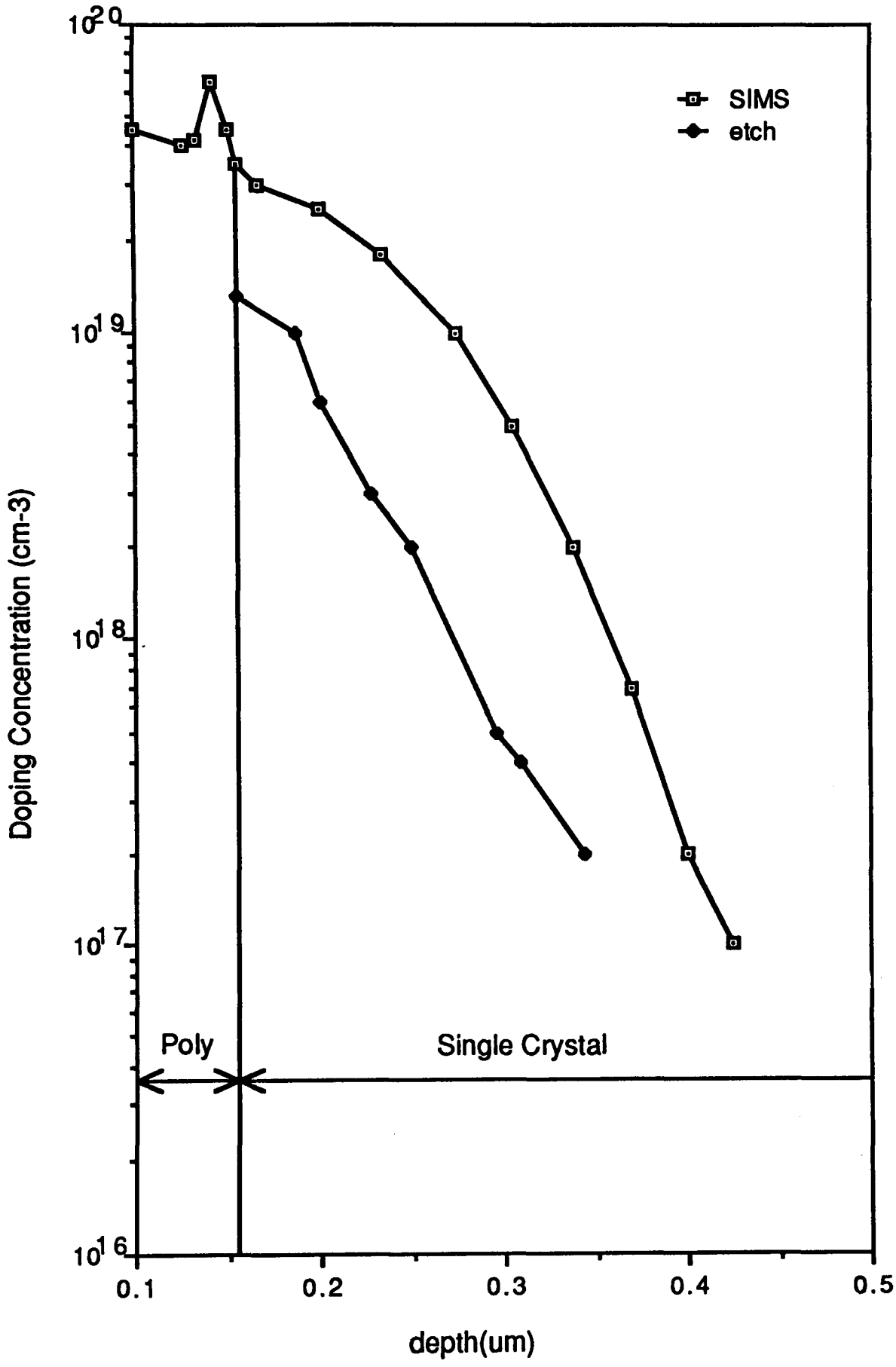
rates greater than those measured by Roberts were possible if a more rapid stirring rate was used, although the same general shape of curve was observed (see fig.5.4a). Wills also established the importance of using a constant position on the etching disc for all measurements of etch rate, and concluded that a large part of the scatter in the results of Roberts was caused by variations in the positioning of consecutive samples during the calibration etching. In this situation, the rate of consumption of etchant in regions of high doping concentration is more rapid than the rate of exchange of depleted etchant for fresh etchant caused by the stirring action of the teflon holder. Thus the apparent etch rate ceases to rise. Roberts did not always observe the flattening of the etch curve to be as severe as in the average profile used here, and there is significant scatter in the results. In the present work, the sample was always placed in the same pre-determined position. To produce values of resistivity for n.e.r values greater than 2.1 the graph was extrapolated from the straight centre section with a less severe curve than that produced by Roberts, based on a comparison of the etch curve of Wills with that of Roberts for the 0.5%/99.5% etch. This is illustrated by the dotted line on Roberts' calibration curve (fig.5.5b). Thus the tailing effect of the etch at high doping concentrations is reduced, for the reasons stated.

A 1D profile of resistivity below the emitter was thus produced using the modified etch rate curve of Roberts. The active doping concentration when n$\gg$p may be calculated from the resistivity (ρ) using the expression;

$$\rho \simeq \frac{1}{q\mu_n n}$$

Where μ_n is the electron mobility and n is the electron concentration. A plot of resistivity vs. active dopant concentration taken from Sze [14] was used to convert the resistivity vs. depth calculated for the sample into a profile of doping concentration vs. depth, this profile is shown in fig.5.21.

Plotted on the same curve is a SIMS dopant concentration vs. depth profile obtained by Loughborough University. The analysis was performed on a blanket



Diffused base (p-type) 1-D profile

Fig.5.23

wafer processed at the same time and in exactly the same way as the device wafers, but excluding the lithographic delineations. Comparison of the shapes of the dopant profiles shows that the delineation is slightly deeper than the depth measured from the SIMS profile. There is also an offset towards the polysilicon/monosilicon interface of the etched delineation compared with the SIMS profile, this offset becoming larger as the doping level rises.

One other device on the same foil was also examined. This device was in a thinner region of the foil than the first device examined, and is shown in fig.5.14a. It can be seen that there is a defect running through the delineated region which has affected the local distribution of fringes. This device was chosen despite the defect, as it occurred in a suitably thin part of the foil. The section chosen was taken at a point away from this defect and the profile in this region was assumed to be representative of a device without defects. The fringe data was extracted from the micrograph as described above for the previous specimen. to produce a 1D doping profile of the emitter regions. The 1D profile from the section A-A is given as device 2 in fig.5.21. The two 1-D profiles show excellent agreement with each other. However, both deviate from the SIMS profile at high doping concentrations. This self consistency within the same foil makes the examination of different structures fabricated close to each other on the same wafer a potentially useful application of this technique.

A profile of the doping concentration in the p^+ contacted base region was similarly obtained. Due to the small number of fringes visible in the (110) mb bf image, the profile measurement was performed on the weak beam g3g (220) image. In this case, the extinction distance was calculated to be 13.8nm. This was confirmed by producing a thickness profile from the same region using the (110) mb bf thickness fringes. This was found to correspond very closely to the calculated weak beam contours. The 1-D profile of dopant concentration vs. depth for this region, using the weak beam image was calculated in the manner described above, and this dopant profile is illustrated in fig.5.23, with

the corresponding SIMS profile. In this case however it was possible to obtain the SIMS profile from a blanket area on the same wafer as the devices whch were thinned for the TEM etching measurements. A similar offset in the doping profile was again observed, as for the emitter regions. The general shape of the SIMS profile is well reproduced.

It is now necessary to examine the possible causes of the offset between the etch measurements and the SIMS data. The offset must be due to one of four effects. Firstly, there may be a genuine difference in the doping profiles between the blanket areas from which the SIMS results are obtained and the small scale device regions, due to either process variations or scaling effects of the emitter. The emitter and base in the sample with the deeper emitter both show offsets, despite having significant geometrical differences, additionally, the base SIMS region is on the same slice as the device, thus process variations seem unlikely. Secondly, the two techniques may be measuring different effects. Thus while SIMS is a chemically sensitive technique, the etch sensitivity may be due to the electrically active dopant concentration (Wilshaw, see[13]). Cuthbertson and Ashburn [15] have shown that in polysilicon contacted devices of this sort, the chemical concentration of dopant atoms may exceed the concentration of ionised carriers in the single crystal emitter, although their data has only one point within the single crystal region. In this case, a displacement in the etch results downwards would be expected, as observed. The displacement observed by Cuthbertson and Ashburn in the single crystal region was approximately 20nm at a doping concentration of $4\times10^{18}cm^{-3}$. The displacement observed in the emitters in this work was slightly greater at 34nm. However the displacement between the diffused base region and the corresponding SIMS profile is significantly greater at 50-100nm.

Thirdly, the displacement may be an artefact of the experimental technique. For example, the presence of an oxide on the foil surface might result in a period of equilibriation being necessary at the beginning of the etch. The etch

mechanism of HF/HNO_3 can be considered as oxidation of the wafer surface by the HNO_3 followed by removal of the oxide by the HF. The fact that the etch is only chemically sensitive at low HF concentrations [9], suggests that the dopant sensitive property of the etch is related to the rate step controlled by the HNO_3 reaction. This would lead to the formation of an equilibrium thickness of oxide. If the native oxide on the surface of the foil were of a different thickness to this value, then there would be a period of equilibriation for the etch. This would be true even for the case of no oxide being present when the etch was at equilibrium. This effect would be insignificant for the calibration samples used by Roberts, as these were bulk samples etched for relatively long periods of time (compared with the foils).

A fourth possibility may be identified by comparing the results from the deep and the shallow emitter. The final fringe in the deep emitter occurs at approximately the depth predicted from the SIMS data but there is an offset in the shallower part of the emitter from the SIMS measurement. The shallow emitter exhibits an offset at the delineation, but this offset is not observed in the centre of wide emitters of the same type. It could be concluded from this that an offset between the SIMS data and the measured etch profile may be the result of an effect associated with the edge of the emitter region, and that this effect only operates within approximately 1200Å of the monosilicon/polysilicon surface. The wider devices appear to confirm this. It is not clear however whether this effect acts on the actual dopant distribution or on the mechanism of the etch.

However, the most reasonable explanation is that the etch calibration curve of Roberts is insufficiently close to the real etch rate curve even when the correction described has been applied.

In an attempt to measure the etch rate of the foil directly, direct measurement of the foil thickness was undertaken. The method used was convergent beam electron diffraction (CBED) using the method of Jotsons and Kelly [16]

as described by Tanaka and Terauchi [17]. The technique of CBED is described briefly in chapter 2. Measurements were carried out on a series of foils. In all cases the foils had been etched previously to the first measurements using the delineating etch. This was to ensure that the foils did not contain any surface films resulting from specimen preparation which might have affected the etch.

The technique was found to be accurate to approximately $\pm 6\%$, calculated by measuring the same region on the same foil on four separate occasions. However it was found that the foil thickness did not vary after etching when measured by CBED. Thus it is clear that the sample has ceased to etch. This has been attributed by the present author to the polymerisation of contaminant hydrocarbons in the TEM from one of the vacuum systems used. This polymerisation is believed to be due to the high beam currents present in the small electron probe diameters used for CBED as it does not generally occur in diffraction contrast imaging when a more diffuse electron beam is used. Small dark spots were observed on the wafer surface after measurement and coincident with the points of measurement, these were believed to be contaminants polymerised under the beam. A variety of organic solvents were used to clean the samples both before and after CBED, but the effect persisted. Thus it is attributed to contamination from the electron microscope used. An alternative might be to use a microscope with an ultra clean vacuum system.

5.3.3 EXPERIMENTAL 2-D DOPANT PROFILES

It is now possible to produce 2D profiles from the experimental 1D profiles. In this work, 2-D profiles were obtained by assuming that the observed thickness fringes represent dopant iso-concentration contours. The micrographs used in fig.5.13 and fig.5.14a were enlarged photographically to produce an approximately A4 image as shown in fig.5.24 and fig 5.25 respectively. The image

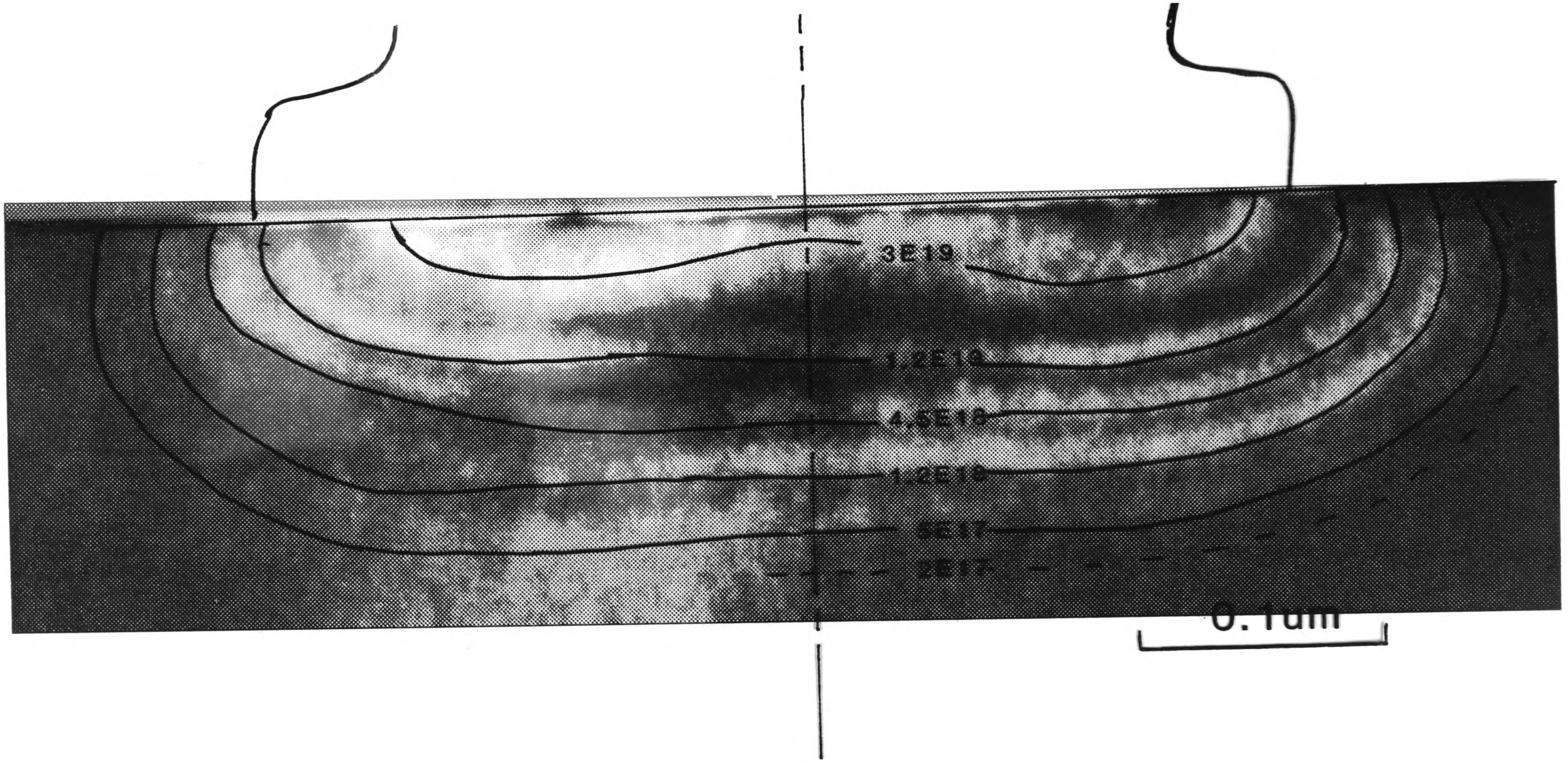
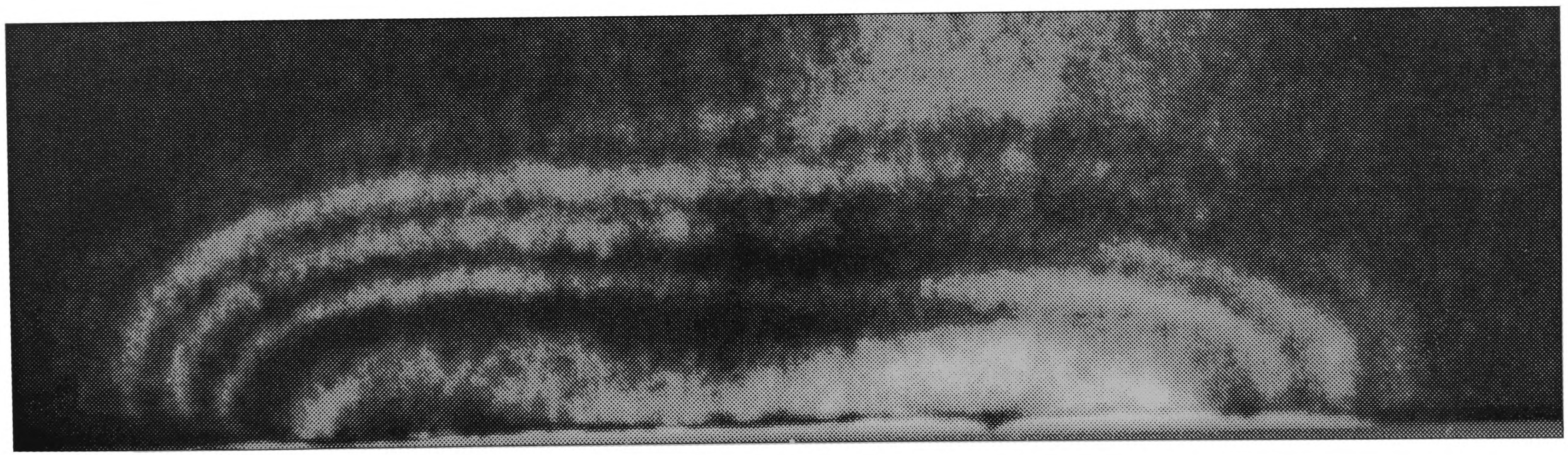


Fig. 5.24 2D profile device 1



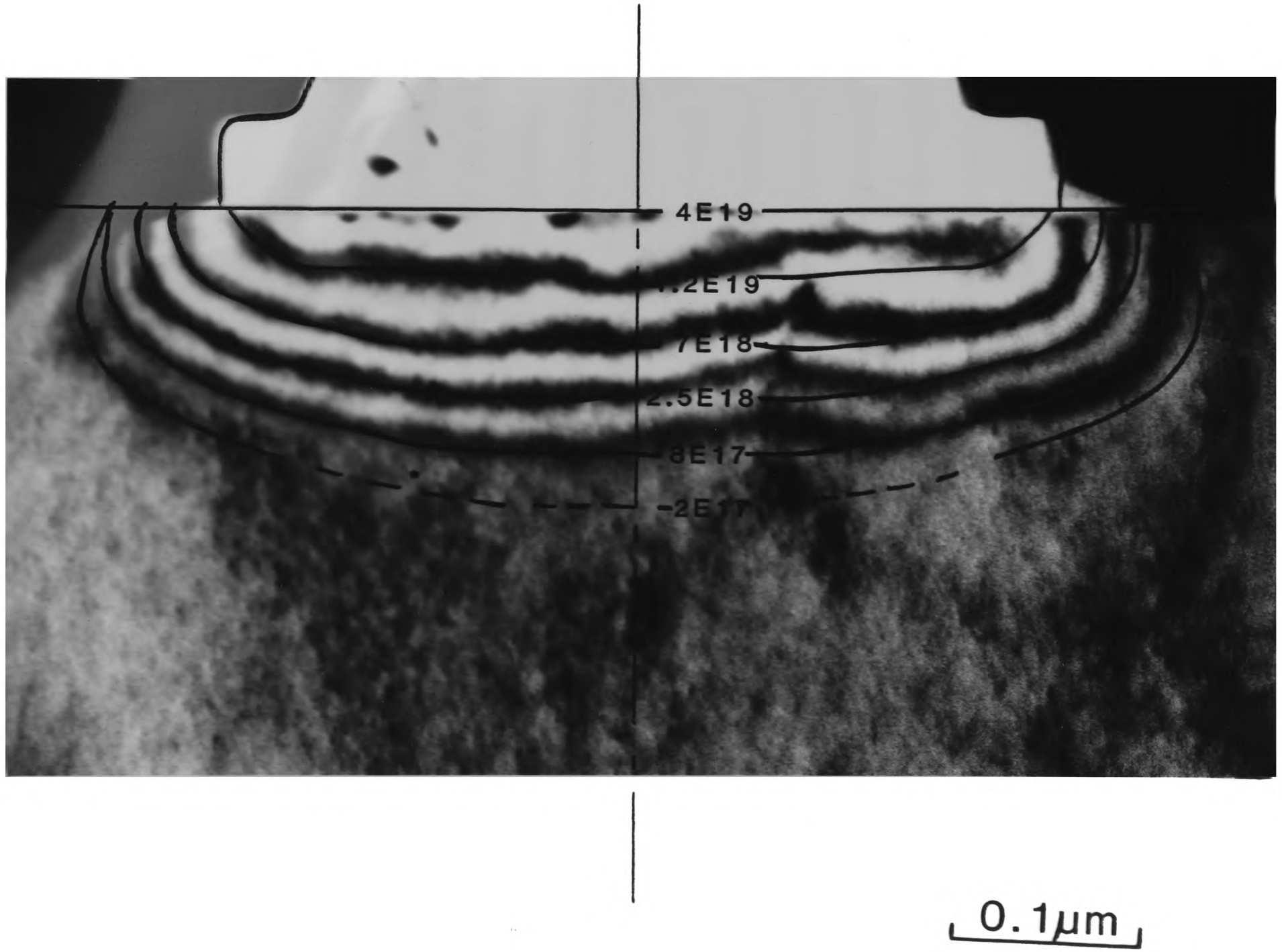
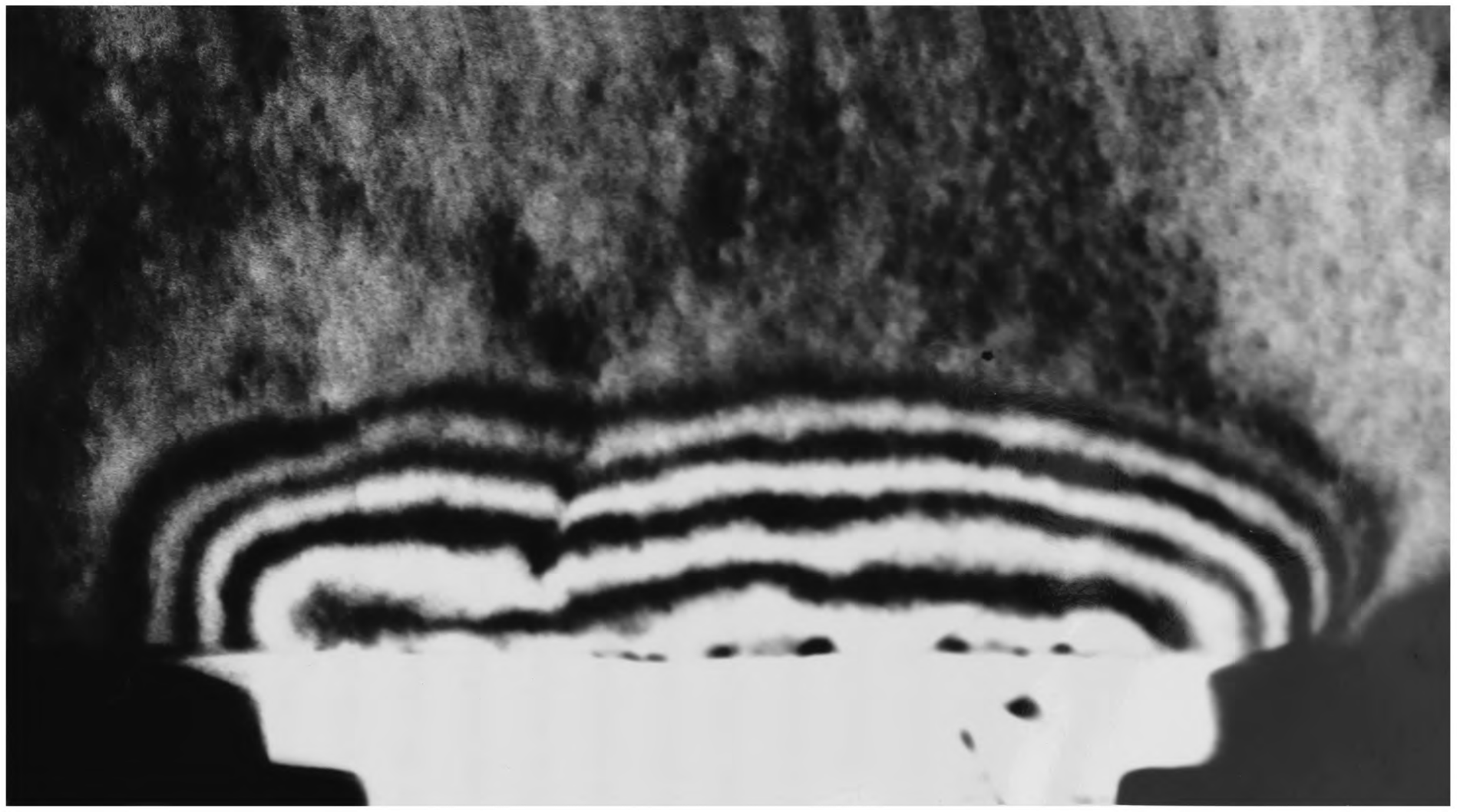


Fig.5.25 2D profile device 2



in fig.5.25 is actually an enhanced image produced by the Semper 6 processing package. An acetate sheet was then laid over the surface of the micrograph and the centre of the appropriate fringe was carefully hand traced onto the acetate sheet. The appropriate sheets are attached to their respective micrographs. The contours are labelled by reference to the 1-D profiles previously produced, and the appropriate labels are given on the corresponding acetate sheets. In the case of the second deep emitter investigated, the right hand side of the etched region is mirrored onto the left to produce a profile. This was carried out to avoid introducing the observed defect into the profile. A number of observations can be made by reference to these micrographs. These observations are made on the assumption that the shape effects observed are a consequence of the doping distribution and not an artefact of the technique.

From examination of fig.5.24, it is clear that the delineation is W shaped. However, the amount of variation from planarity varies for different doping concentrations. Near the surface, where the doping concentration is high, the non-planarity is greatest. In the weakest fringe, the observed non planarity is very slight, less than 10nm. The device in fig.5.25 can only be assessed in the broadest terms due to the effect of the defect on the delineated shape. Comparison with the data from the shallow emitters suggests that the presence of the polysilicon fillets, may result in a reduction in the depth of the delineation close to the surface, and this may explain in part why the deviation from planarity increases as the fringes approach the surface. In the shallow emitter devices, it is clear that the approximate shape of the deeper emitters is retained. The shallower device without polysilicon fillets appears to be totally planar, as for the deeper emitter without polysilicon fillets. Thus for the first time it is shown that planarity is retained in the emitter/base junction when the depth of the emitter is reduced to 800Å or less

An area of great interest to device engineers is the extent of diffusion under either side of the emitter window during fabrication. For the shallower emitters,

this distance was measured to be only 20nm. The situation is rather different for the deeper emitters. The distance that the emitter dopant has diffused under the window edges is approximately 100nm on either side. More importantly the fringes are not perpendicular to the interface at the point of contact, but bend inwards towards the window. The curvature becomes less extreme as the fringes become closer to the window. This curvature is not observed in the delineations of the shallower emitters, which are approximately perpendicular to the interface at the point of contact. Because this effect is only observed in the deeper junction, it is believed by the author to be a real effect of variation in dopant distribution. This is probably caused either by the relatively slow diffusion of dopant along the monosilicon/oxide interface compared with the bulk diffusion rate, or by segregation of dopant to the oxide or oxide/monosilicon boundary. Thus dopant concentration in the region adjacent to the interface is lower for a given distance from the window than the corresponding concentration elsewhere in the bulk.

Differences were observed between the doping profiles of the two samples from the same foil. This may be a real effect, however it seems more likely that it is the result of the presence of the observed defect within one of the delineated regions. Both regions however do exhibit reasonable agreement with each other compared with their agreement with the SIMS results.

The potential accuracy of the technique may now be assessed. Further work will be needed before repeatability from foil to foil can be demonstrated, however comparison of devices on the same foil may be made if the etch has acted uniformly, as detailed in chapter 2. Even when the etch acts non-uniformly, the delineated region was clearly visible, therefore the technique may be used to estimate emitter shapes in these devices. Emitters as shallow as 500Å can be imaged and there is no obvious reason why emitters as shallow as 100Å could not be imaged. Examination of the weak beam image of the emitter region reveals that the fringes are separated by as little as 10nm. This separation

could potentially be improved, either by varying the imaging conditions, or by etching a steeper slope in the delineated region. An emitter width of 400nm has been easily resolved by the technique and the author sees no reason why widths of 100nm or less could not be measured. Roberts measured the limit of sensitivity of the technique to be $2\times10^{17}cm^{-3}$. Wills however has shown that the potential sensitivity of the technique is approximately $5\times10^{16}cm^{-3}$. Low concentration sensitivity may be statistically limited in thin foils however. In a region 10nm×10nm×50nm (the estimated resolution of the technique), doped at $10^{17}cm^{-3}$, there will only be an average of just 1 dopant atom. For these doping levels, slight statistical variations in the local doping levels may have a significant effect on the resolution at low doping levels.

5.4 CONCLUSIONS

In this chapter, a two dimensional analysis of dopant distribution has been carried out using a dopant sensitive etch, for the first time in a real bipolar device structure. The shape of the emitter in a polysilicon contacted device has been revealed for the first time. It has been shown that the emitter profile is flat bottomed and not dished as suggested in the literature.

It has been shown that the technique may be sensitive to effects other than dopant distribution. An offset has been observed in the profiles produced compared with SIMS due to a lower measured dopant concentration. It has been shown that the technique is capable of revealing differences in devices caused by small alterations in the surface structure of the device.

Lateral diffusion of dopant under the oxide mask edge is retarded compared with diffusion in the bulk, leading to the emitter profile curling back on itself in the vicinity of the interface.

It has been demonstrated for the first time that the general shape of the emitter is preserved during scaling of the emitter depth to 80nm or less.

The resolution of the technique has been shown to be better than 10nm in depth, with a minimum profile width of 100nm or less. Dopant concentrations down to approximately 2×10^{17} have been observed in accordance with the work of Roberts. Potential sensitivity as reported by Wills may be at least $5\times10^{16}\text{cm}^{-3}$. This may be statistically limited in TEM experiments.

References

[1] . P. Ashburn. Design and Realisation of Bipolar Transistors, Wiley 1988.

[2] . P.C. Hunt and M.P. Cooke, IEEE Custom Integrated Circuit Conference, Rochester New York, May 1988

[3] . M.C. Wilson, D. Gold, P.C. Hunt and G.R. Booker, 1988 Proc. IEEE BCTM Minneapolis USA p.128

[4] . S. Duncan, M.C. Wilson,, P.C. Hunt and D.J. Bazeley, IEEE VLSI Symposium, San Diego, 1988.

[5] . M.G. Dowsett, G. Cooke, C. Hill, E.A. Clark, I. Snowdon, B. Lewis and P. Pearson, Proc. SIMS VIII, Monterey California, Sept. 1989.

[6] . C. Hill, P.D. Augustus and A. Ward, Inst Phys. Conf. Ser. 76(11) pp.477-482, 1985

[7] . C. Hill, P.J. Pearson, B. Lewis, A.J. Holden and R.W. Allen, Sol. Stat. Dev., pp.147-150, 1988.

[8] . P.J. Pearson and C. Hill, J. de Phys., Colloque C4, supplement No.9, Tome 49, 1988.

[9] . Sheng T.T. and Marcus R.B., J. Electrochem. Soc., 106(6), p.881, 1981.

[10] . J.H. Wills, MSc Thesis, University of Oxford 1987.

[11] . M.C. Roberts, K.J. Yallup and G.R. Booker, Inst. Phys. Conf. Ser., 76(11), pp.483-488, 1985.

[12] . M.C. Roberts, D.Phil Thesis, Oxon, 1985.

[13] . B-Y. Hwang, P.J. Zdebel, R.J. Balda, G.G. Sweeney and M V. dl.Torre, IEEE BCTM, Minneapolis 1987.

[14] . S.M. Sze. Physics of Semiconductor Devices, Wiley 1981.

[15] . A. Cuthbertson and P. Ashburn. Trans. Electron Dev. ED-32(11), pp.2399-2407 1985.

[16] . P.M. Kelly, A. Jotsons, R.G. Blake and J.G. Napier, Phys. Stat. Sol.,(a) 31 p.771 1975.

[17] . M. Tanaka and M. Terauchi Convergent Beam Electron Diffraction, Publ. JEOL ltd, 1985.

Chapter 6

CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

6.1

In this thesis an investigation has been made of the emitter regions of state-of-the-art high speed bipolar transistors, primarily using the transmission electron microscope (TEM). The device technology involves the use of polysilicon contacted emitters in which the polysilicon acts as a diffusion source for the emitter dopant. This work forms part of a project which has resulted in the development of a silicon bipolar device for use in VLSI applications with a maximum frequency of oscillation of 22GHz. At the time that these devices were first tested, they represented the fastest of any published silicon bipolar devices in the world.

The importance of the interfacial oxide, the interface structure and the polysilicon layer on device performance has been suggested by many authors. In chapter 3 and chapter 4, a comprehensive investigation of the effect of varying oxide thickness, annealing conditions and dopant type has been carried out.

Neither phosphorus nor arsenic was observed to result in a significant change in the oxide structure from the undoped case. It is concluded that the mechanism resulting in accelerated regrowth is not the same mechanism controlling the break-up and balling-up of the oxide. The two oxides investigated were produced

either by an HF clean (4-7Å, irregular) or by an RCA clean (12-14Å, uniform). The balled-up oxides appeared similar in cross-section, the RCA oxide being thicker for any given degree of break-up. In plan-view however, the oxides exhibited differing structures. In all case, the balled-up HF oxide consisted of a series of worm-like particles whilst the balled-up RCA oxide resulted in oxide particles which were approximately circular in the plan-view images and disc shaped in three dimensions. This behaviour is explained in terms of the competing mechanisms driving the break-up of the oxide layer.

It is also concluded that for the doping levels investigated, in the presence of a thin oxide, phosphorus accelerates the epitaxial regrowth of the polysilicon layer, whilst arsenic does not produce significant variation from the undoped case. Further work in this area could usefully be carried out on different doping concentrations, specifically, on higher doping concentrations of As. The effect of phosphorus on the regrowth rate of the polysilicon layer is not observed in the presence of the thicker oxide. This is explained in terms of the driving force for epitaxial regrowth and the driving force for break-up of the thicker oxide layer.

In some samples unusual abrupt contrast was observed in the single crystal substrate adjacent to the oxide layer when imaged using high resolution electron microscopy. Similar contrast has been observed by other workers and has ben attributed to the existence of a crystalline sub-oxide. It is concluded that this cannot be directly inferred from HREM data, but that it provides strong circumstantial evidence for the existence of such a crystalline sub-oxide. HREM simulations may lead to a more complete understanding of the contrast mechanism, however this type of interface represents a particularly difficult system to model.

In chapter 4, the pre-anneal technique and structures investigated in chapter 3 have been applied to the development and characterisation electrically and structurally of a phosphorus doped emitter npn bipolar transistor. A model has been developed to explain the electrical behaviour of the devices as the

interfacial oxide breaks up and the polysilicon epitaxially regrows. This model has been compared with a sophisticated computer modelling package and good agreement has been found between these.

The supression of hole current observed in such devices is generally aknowledged to arise from two separate effects. The first of these is differential oxide tunnelling in which the barrier height of the oxide to holes is higher than the barrier height of the oxide to electrons, resulting in preferential supression of the hole current. This model fits the corresponding data well, with a hole barrier height of 1-2eV and an electron barrier height of approximately 0.4eV. The second mechanism is less well understood however and has been variously attributed to a thermionic barrier resulting from the segregation of active dopant to the polysilicon/monosilicon interface or to reduced mobility in the polysilicon layer. In this work it is shown that the model of reduced mobility does not explain the observed device behaviour. Neither does the model of active dopant segregation at the interface appear to fit the observed data. The experimental data supporting the importance of segregation to this barrier suggests that the polysilicon/monosilicon interface and the associated segregation are important in hole current supression, but for reasons other than those suggested by previous authors. This second mechanism of gain supression in these devices remains a matter of controversy, considerable further work may be necessary to resolve this mechanism. The study of pnp transistors is the most promising area of development to increase understanding of this mechanism.

Devices have been fabricated as described in chapter 4 with potentially useful DC charateristics, particularly a low emitter series resistance and some enhanced gain. The next stage in the development of such a device is to fabricate a similar device constructed to assess the high frequency characteristics of this structure. Enhanced gain was observed in a device with a broken oxide layer, and an epitaxially regrown polysilicon region. The mechanism for this is not clearly understood. This could be usefully investigated by thinning the polysilicon layer

of such devices and observing the point at which the base current begins to rise. This would enable the gain enhancement mechanism in this particular case to be more clearly identified.

In chapter 5 an investigation of the dopant distribution in fully processed device-like emitters has been carried out for the first time, based on the work of Roberts. It has been demonstrated that the shape of the emitter dopant profile has a planar bottom with approximately semi-circular ends. This vindicates the belief that the polysilicon/monosilicon interface acts as a nearly infinite dopant source as suggested previously. The 1-D dopant profiles were found to deviate from the SIMS profiles for equivalent bulk regions of material. It is not clear if this effect is the result of a real difference in doping distribution, or an artefact of the etching technique. The presence of a different surface geometry on the devices was observed to result in a non-planar delineation.

It has been demonstrated that the approximate shape of the emitter is retained when the junction depth is reduced from approximately 130nm to approximately 80nm or less. This suggests that scaling of emitter junction depths will be a relatively simple task as the next generation of bipolar devices are developed. In the deeper emitters, it was observed that the emitter edges were bent near the interface by the relatively slow diffusion rate of the monosilicon/oxide interface.

The etching technique was observed to be suitable for the observation of emitter doping shape. Under suitable conditions, a doping profile could be calculated from the image. However, considerable further work will be necessary to ensure repeatability of the technique. This may be divided into two areas, accurate measurement of the material removed, and the resolution of the failure to etch observed in some samples. It may be possible to apply the technique of convergent beam electron diffraction to this problem if the polymerisation of contaminants can be avoided. For example by utilising an ultra-clean vacuum system such as in a dedicated STEM. Alternatively the use of a cold stage TEM

holder would help to reduce the migration of contaminants on the surface into the beam. Ecob suggests that one source of error when calculating foil thickness is positioning the minima and maxima within the appropriate bands of contrast (these are not symmetric). This could be avoided in future by taking micro-densitometer traces across the discs from the negatives.

The investigation of the occasional non-linearity and failure to etch observed is probably the most complex single piece of work which can be carried out in this field. The next step in this work is to electrically isolate the gold grid from the foil and observe the effect of this on the etching of the sample. However the number of potential causes of the non-linearity and failure to etch are limited only by the imagination of the observer, thus the problem is highly complex.

The effect of stress on the etch requires further investigation. There are two approaches to this. Either the stress may be directly measured, for example using selected area diffraction patterns (SACP), or a similar device might be investigated using the beveled SIMS technique of Dowsett et al. The necessary removal of surface structure for this technique would also provide an interesting opportunity to etch a sample with no surface structure and observe whether the non-planarity of the emitter profile continues.

It may be possible to reduce the speed of the etch still further without modifying the etch proportions further, by diluting the etch in an inert solvent such as methanol. This would help to improve the accuracy of termination of the etch, but would require recalibration of the etch. Recalibration of the $0.3\%HF/99.5\%HNO_3$ is necessary as the scatter in the work of Roberts is unacceptably high, and the resulting etch calibration curve does not explain the observed results in this thesis sufficiently well.

The data listed below is the non-default data input into the BIPOLE computer modelling package, when using the polysilicon contacted emitter model. This model assumes only Auger recombination in the polysilicon and assumes gaussian dopant profiles. A base width of 0.32um was chosen. The base width was used as the fitting parameter to the electrical results.

```
B      =0.7e-3
BPB    =0.14E-2
DELTA=0,8,14..... OXIDE THICKNESS
ELEM   =0.7E-3
ELPB   =0.34E-2
ESB    =0.14E-2
FACN   =0.1E-2
ICI    =2
ICON   =1
IEDL   =-10
IEND   =5
IPOLY  =-3
NB1    =0.155E19
NE1    =0.7E20..... DOPING LEVEL AT INTERFACE
NEPI   =0.1E17..... DOPING LEVEL OF EPITAXIAL LAYER
NPOL   =0.7E20..... POLYSILICON DOPING LEVEL
NTOT   =1
PEE    =0.1E1
SME    =0.1E4
TAUL   =0.1E-4
TAUE   =0.1E-5
TE     =0.22E2
VCIN   =0.5E1
XEND   =0.2E-3
XFS    =0.1E-3
XJ1    =0.1E-4..... WE(FROM INTERFACE)
XJ2    =0.42E-4... WB(FROM INTERFACE)
XE     =0.4..... ELECTRON BARRIER HEIGHT
XH     =1.0..... HOLE BARRIER HEIGHT
```

Data for computer modelling using the BIPOLE package

APPENDIX 1