BATCH-FABRICATION OF NOVEL NANOPROBES FOR SPM

Ashish Midha

Submitted for the degree of Doctor of Philosophy
to the Nanoelectronics Research Centre
in the Department of Electronics and Electrical Engineering
at the University of Glasgow

October 1999

PAGE NUMBERING AS FOUND IN THE ORIGINAL THESIS

मेरे माता-पिता को समर्पित

(To my parents)

"Some ass has got to do the donkey work"

— a quote from somewhere.

"This entire story has been fabricated"

— a pun from somewhere.

PRÉCIS

This thesis describes the development of a commercially viable technique for batch-fabricating novel nanoprobes to act as the imaging component in a variety of scanning probe microscopy (SPM) techniques.

In order to accomplish this a method for combining bulk silicon micromachining and electron-beam lithography has been developed. A special micromachined substrate of basic atomic force microscopy probes for both contact and non-contact mode has been designed, complete with the necessary features for electron-beam lithography to be carried out on the cantilevers and probes. The design of these probes has also been optimised for ease of cleaving, allowing access to the specimen and easy mounting onto a scanning probe microscope.

A micromachining method has been developed for fabricating 20μm tall silicon atomic force tips with flat tops less than 2μm wide suitable for defining nanosensors upon, and with low aspect ratio sides suitable for defining electrical connections to the sensor.

Methods have been developed to allow flat substrate processing techniques to be applied to such non-planar micromachined substrates. This has necessitated the development of a novel resist-coating technique and the use of defocused electron-beam lithography. Methods for through-wafer alignment by electron-beam lithography and accurate alignment to the tips using micromachined alignment markers have also had to be developed.

The fabrication process has been designed to enable a wide variety of sub-micron sensors to be defined on the atomic force probes, with little additional development beyond that of the sensors themselves. This flexibility has enabled very different sensors meant for very different scanning probe microscopy techniques to be designed without significant redevelopment of the underlying fabrication process. The main restrictions on the type of sensor that can be used are the physical dimensions of the sensor, the number of alignment levels necessary, the degree of alignment accuracy required and the choice of sensor materials. However, within these constraints it has been found that probes optimised for scanning near-field optical microscopy (SNOM), scanning thermal microscopy, modulation differential scanning calorimetry (MDSC) and scanning Hall-probe microscopy can be fabricated.

For the SNOM probes three methods for fabricating sub-100nm diameter apertures have been developed, analysed and compared with each other to evaluate both the process latitude, and, the size and reproducibility of apertures that can be fabricated, as a function of electron beam dose, pattern shape and size, and metallisation material and thickness. Two methods, both utilising multilayer 'resist' schemes have been found suitable for this purpose, one based on conventional electron-beam lithography with

PMMA and a new dry etching process for titanium, and the other based on a novel electron-beam lithography technique utilising cross-linked PMMA for lifting off nichrome. A simple analytical model has also been developed for these probes allowing the effects of changes in the sensor design parameters on the light throughput to be compared qualitatively, if not quantitatively.

For the scanning thermal probes a method for lifting-off sub-100nm, thin-film thermocouple sensors on silicon tips without the loss of electrical continuity has been developed. For the MDSC probes, a similar method has been developed for defining thermal resistors. A method has also been presented for fabricating sensors for scanning Hall-probe microscopy based on an evaporated germanium sensing layer. This has been found to require annealing and optimisation of sensor design and geometry to reduce sensor resistance to acceptable levels.

ACKNOWLEDGEMENTS

Many personages must be acknowledged for their hand, whether direct or indirect, in the production of this thesis. An attempt will now be made to name those who must be held responsible for their role in the work described herein. Alas, the large number of people involved means that inevitably, some will not be given their due credit, but such persons should rest assured that their contribution, no matter how insignificant it may have seemed to them, has been much appreciated.

Firstly, Peter J. R. Laybourn and his successor, Steve P. Beaumont are thanked for the provision of excellent fabrication facilities and ancillary support within the Department of Electronics and Electronic Engineering, and without which the work described here could not have been carried out, and indeed would never have even been attempted.

In particular, John H. Marsh is acknowledged for the existence of the photolithographic facilities, Steve P. Beaumont is acknowledged for the existence of the electron-beam lithography facilities and Chris D. W. Wilkinson is acknowledged for the existence of the dry etching facilities within the department.

The somewhat acrimonious process of learning the ropes when it came to photolithography was carried out under the watchful eyes of Lois Hobbs, Joan Carson and Linda Pollock. Special mention must be made here to both Karen Thayne, for general advice on photolithography, annealing and cleaving, and Alan W. Griffith for advice on wet etching, and chemical vapour deposition. Credit for my acquired knowledge of the nuances of photolithographic mask making must go principally to Robert Harkins, whose patience in the face of my screw-ups was astonishing to say the least.

The even more acrimonious process of getting to grips with electron-beam lithography and lift-off was supervised by a 'beamwriter guru' masquerading under the guise of Steve Thoms, and almost equally guilty partner in crime, Douglas MacIntyre. In particular, Steve Thoms is thanked profusely for implementing defocused electron-beam lithography and making possible automatic alignment to the exotic varieties of alignment marker thrust upon him, without which most of the work in Chapters 5 through 8 would have been pointless. In this regard credit must also go to all the beamwriter staff: Dave Gourlay, Helen McLelland, Susan Ferguson, Lesley Donaldson, Robert Harkins and Douglas MacIntyre for their 'other-worldly' patience and understanding when attempting to run what must have seemed to them to be some of the worst looking, most fragile and least friendly substrates ever to grace the face of the planet, through the beamwriter.

Additional thanks must go to Dave Gourlay for teaching me how to use an SEM properly and for showing me his notes on 'dip-coating' and the patterning of LPCVD silicon nitride using orthophosphoric acid. The latter was extremely useful in the early stages of this project.

The design of the silicon wet etching kit owes much to the advice of Andrew Glydle and Douglas MacIntyre. The execution of the design was largely left in the capable hands of Dougie Irons and the mechanical workshop team and Stuart Fairbairn and the electronic workshop team. John Weaver must also be acknowledged here for his design of the temperature sensing circuit for platinum resistance thermometry.

John M. R. Weaver must also be thanked for the inspirational idea which resulted in the 'float-coating' technique – that he thought to compare the fabrication of nanoprobes with constructing an 'airfix' model may have seemed odd to some. Without this technique, the project would definitely have stalled at Chapter 4.

Simon Hicks is credited with the provision of a technique for the deposition of silicon nitride with adjustable residual stress by chemical vapour deposition, without which functional contact mode probes would have been that much more difficult to fabricate.

The development of new dry etching processes was only made possible with extensive hand-holding and advice from the dry etching 'guru' otherwise known as, Saad K. Murad. Without making use of his uncanny knowledge of the multi-dimensional etching parameter space, the development of SNOM probes would have progressed at snail's pace, if at all. Gillian Hopkins and Dave Clifton must also be thanked for advice on running substrates through the dry etching machines.

Hubert Pollock of Lancaster University is thanked for the opportunity to demonstrate the flexibility of the developed fabrication process.

The learned researchers, Nigel I. Cameron, David R. S. Cumming, Thomas Krauss and Ian Thayne are also thanked for their many insightful comments on fabrication, as are the more contemporary, Robert L. Nicol, Khaled Elgaid, Iman Khandaker, Elef Skuras and Babak Khamsepour.

Clivia M. Sotomayor Torres is thanked for advice on all things spectroscopic, Michael R. S. Taylor is thanked for advice on modelling SNOM probes and John H. Davies is thanked for explaining the behaviour of electrons (and holes) in very, very small devices.

An extremely huge thank you is reserved for Lesley Donaldson who stuck with all the wacky, weird and downright ugly processes I threw at her, through thick and thin, and without whom the work here would have taken twice as long.

Credit must also be given to the, now legendary, final year student, Shalim Rahman, who managed to convince us all of the devious simplicity with which these probes could be fabricated by carrying out an initial feasibility study as his final year project.

The three 'J's, John M. R. Weaver, John G. Williamson and John Cooper, are also thanked for supervising and enthusing about this work, keeping me insane, and keeping me sane respectively. With regards to my sanity, Mahfuzur Rahman, Kevin Lister and David R. S. Cumming must also be given due credit. I also feel moved to apologise to

fellow researchers, Derek Corcoran, Kevin Lister, Brendan Casey, Nicholas Green, Christopher Smith and Gary Ternent, who have had to bear more than their fair share of general 'puniness' during my time in the department.

The unsung heroes and inheritors of the fabrication process, Haiping Zhou and Gordon Mills, who reduced the burden of fabrication so much by willingly taking on the final development of the SNOM and scanning thermal microscopy probes respectively, are also well-deservedly thanked.

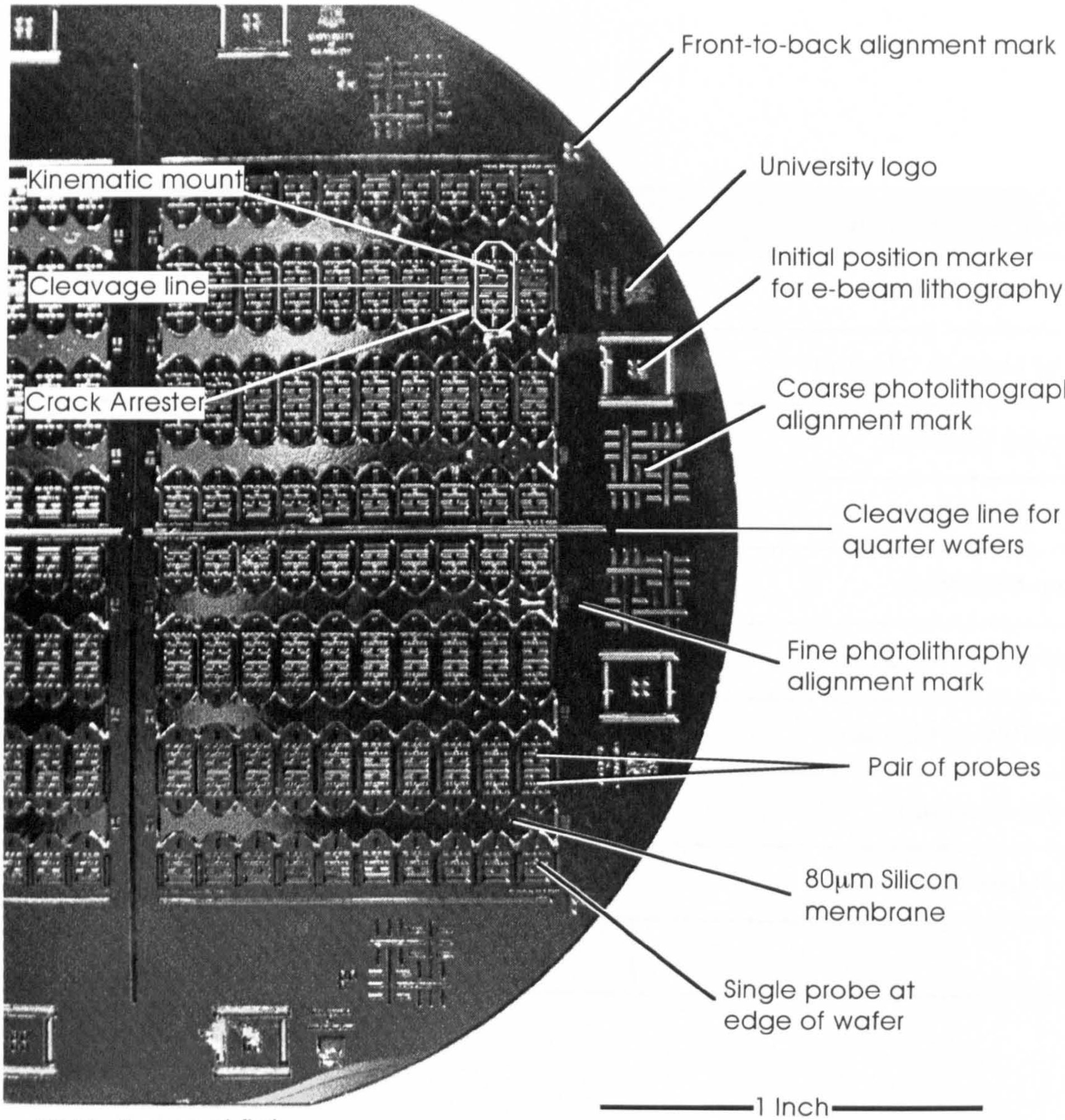
The extremely unfortunate task of having to under-write, direct and review this monstrosity has befallen the *formerly* enthusiastic John M. R. Weaver, to whom all the best is wished...

Financial support from the Engineering and Physical Sciences Research Council (EPSRC) and the University of Glasgow ('New Initiatives Fund') must also be acknowledged.

Finally, I must thank my whole family, especially my parents, for lending moral support and egging me on throughout the writing of this thing (thesis).

TABLE OF CONSTANTS

Quantity	Symbol	Value	Units
Planck's Constant	h	6.6260755 x 10^{-34}	J s
Speed of Light in Vacuum	c	2.99792458 x 10^{8}	m s^{-1}
Boltzmann Constant	k_B	1.380658 x 10^{-23}	J K^{-1}
Elementary Charge	e	1.60217733 x 10^{-19}	C
Electron Rest Mass	m_e	9.1093897 x 10^{-31}	kg
Permittivity in Vacuum	ε_o	8.85418782 x 10^{-12}	F m^{-1}
Permeability in Vacuum	μ_o	4π x 10^{-7}	H m^{-1}
Avogadro Constant	N_A	6.0221367 x 10^{23}	$mole^{-1}$
Atomic Mass Unit	amu	1.6606 x 10^{-27}	kg
Pi	π	3.141592654	—



A section of a 3 inch wafer after the second level of photolithography and the second dry and wet etches. After this stage, pyramids are defined on the other side of the wafer using electron-beam lithography, dry and wet etching. Stress-free silicon nitride is deposited and the cantilever shapes are defined by electron-beam lithography and dry etching. The substrates are completely generic up to the definition of the cantilevers, at which point the processes for contact mode and non-contact mode probes diverge. See Chapter 4 for further details. Figure courtesy of Dr. J. M. R. Weaver.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS

1. A. Midha, S. K. Murad and J. M. R. Weaver, *ANISOTROPIC PATTERN TRANSFER OF FINE RESIST PMMA FEATURES TO SILICON NITRIDE VIA AN INTERMEDIATE TITANIUM LAYER*, Microelectronic Engineering **Vol. 35,** pp. 99-102 (1997).

2. G. Mills, H. Zhou, A. Midha, L. Donaldson and J. Weaver, *SCANNING THERMAL MICROSCOPY USING BATCH FABRICATED THERMOCOUPLE PROBES*, Applied Physics Letters **Vol. 72,** No. 22, pp. 2900-2902 (1998).

3. H. Zhou, A. Midha, G. Mills, S. Thoms, S. K. Murad and J. M. R. Weaver, *GENERIC SCANNED-PROBE MICROSCOPE SENSORS BY COMBINED MICROMACHINING AND ELECTRON-BEAM LITHOGRAPHY*, Journal of Vacuum Science & Technology B **Vol. 16,** No. 1, pp. 54-58 (1998).

4. H. Zhou, A. Midha, L. Bruchhaus, G. Mills, L. Donaldson and J. M. R. Weaver, *NOVEL SCANNING NEAR-FIELD OPTICAL MICROSCOPY / ATOMIC FORCE MICROSCOPE PROBES BY COMBINED MICROMACHINING AND ELECTRON-BEAM NANOLITHOGRAPHY*, Journal of Vacuum Science & Technology B **Vol. 17,** No. 5, pp. 1954-1958 (1999).

5. H. Zhou, A. Midha, G. Mills, L. Donaldson and J. M. R. Weaver, *SCANNING NEAR-FIELD OPTICAL SPECTROSCOPY AND IMAGING USING NANOFABRICATED PROBES*, Applied Physics Letters **Vol. 75,** No. 13, pp. 1824-1826 (1999).

6. H. Zhou, G. Mills, B. K. Chong, A. Midha, L. Donaldson and J. M. R. Weaver, *RECENT PROGRESS IN THE FUNCTIONALISATION OF AFM PROBES USING ELECTRON-BEAM NANOLITHOGRAPHY*, Journal of Vacuum Science & Technology A **Vol. 17,** No. 4, pp. 2233-2239 (1999).

7. H. Zhou, B. K. Chong, P. Stopford, G. Mills, A. Midha, L. Donaldson and J. M. R. Weaver, *LITHOGRAPHICALLY DEFINED NANO AND MICRO SENSORS USING "FLOAT COATING" OF RESIST AND ELECTRON BEAM LITHOGRAPHY*, Journal of Vacuum Science & Technology B **Vol. 18,** No. 6, (2000).

CHAPTER 1

SCANNING PROBE MICROSCOPY

This chapter starts with a general introduction to microscopy and goes on to give an overview of those microscopy techniques that are the precursors and subject of this thesis. The first section describes concepts valid for microscopy in general. and introduces ideas such as resolution, artefacts and contrast mechanisms. The following sections describe various scanning probe microscopy techniques and report on their discovery, their current status, their main principles of operation, and their main applications and limitations. Where appropriate the concepts of the first section are related to the technique and suggestions are made as to what major improvements would be desirable.

1.1 INTRODUCTION TO MICROSCOPY

Microscope comes from the Greek words *mikros*, meaning small, and *skopos*, meaning focus. Historically, the first microscopes were optical and consisted of a single convex lens. Here an object was placed just inside the focal plane of the lens to form a magnified virtual image of the object which could be viewed directly. This 'magnifier' was superseded by the compound optical microscope which makes use of two convex lenses – the *objective* forms a magnified real image of the object just inside the focal plane of the *eyepiece*, which then magnifies this by forming a virtual image, just as in the simple magnifier. This is essentially the principle by which modern optical microscopes work.

Sight is the primary sense by which human beings interact with the world around them, so it was only to be expected that the first microscopes sought to enhance its ability. Due to everyday experience the human brain is well acquainted with the interactions of incoherent visible light, so there is less trouble in interpreting images obtained using this technique, and it is this, together with its simplicity, speed and non-invasive nature, which makes optical microscopes the first choice for examining any specimen.

As the understanding of other physical interactions increased, new techniques were developed to obtain spatially localised information using these phenomena. Consequently, the meaning of microscopy was extended to encompass all such techniques and not simply those based on the interaction of visible light. The use of these other physical interactions together with the drive for greater magnification has shaped the development of this field over the last two centuries. Microscopy has therefore become the science of measurement on a *small* scale, where *small* is continually being redefined as new techniques are invented and developed, and 'measurement' can be interpreted in its broadest sense.

The second half of this century has seen a huge increase in the number of techniques available to the microscopist. For any new technique to become established, it must offer certain advantages over those already existing, but, since microscopy is a practical science, the success of a technique is not purely governed by the kind of information it reveals or by its quality, or even by any improved magnification it may offer – it also depends on other criteria such as specimen preparation, examination conditions, the effects of the technique on the specimen being examined, the contrast mechanisms available, the ease of interpretation of the data, and the speed and cost of the technique. When making comparisons, all these must be taken into account in the context of the specific purpose for which the technique is required. In addition, it is not only the range of applications for which the technique is best suited under these criteria, but also the importance of the applications that determines whether the technique becomes established or simply remains a scientific curiosity until an important application is found.

As an example, consider the scanning electron microscope (SEM) which has been one of the biggest successes of the last seventy-five years [Cosslett[1] (1981); Cosslett[2] (1981)]. Here an electron beam is focused and scanned over the specimen and an image is formed from the intensity of secondary or backscattered electrons received by a detector. Although these interactions are very different to those used in optical microscopes, by presenting the information obtained from them visually, as a grey-scale image, the brain can be tricked into interpreting the information as though it was obtained using optical techniques.

Secondary electrons are knocked out of the specimen as a result of excitation by the incident (or backscattered) electrons. In general, many secondary electrons (depending on incident electron energy and the work function of the scattering material) are generated from each incident electron, each having much lower energy than the incident electron. The lower energy means that secondary electrons generated close to the surface are more likely to escape than those deeper within the specimen. As a result, this interaction is mainly sensitive to the topography of the specimen.

Backscattered electrons are incident electrons that have rebounded and escaped from the specimen. Such 'reflected' electrons have an energy close to that of the incident electrons. The likelihood of a backscattered electron escaping from the specimen without losing much energy (*e.g.* by producing secondary electrons) decreases rapidly with depth (although escape from a particular depth is much more likely for larger depths than with secondary electrons). For this reason, this technique is more sensitive to variations in the composition of the specimen with depth and therefore yields information below the surface of the specimen. Thus both of these interactions yields a different contrast and is termed a *contrast mechanism*.

Due to the difference between these interactions and those of incoherent visible light, the interpretation of these images when presented as a grey-scale image is not straightforward. This is due to the presence of *artefacts*. An artefact is the result of other interactions (or contrast mechanisms) acting simultaneously with the desired interaction (or contrast mechanism), and can often manifest itself as an 'easily misinterpretable' change in contrast. Thus any interactions which act so as to confuse the results of the interaction required result in artefacts.

It can be argued that shadows are an artefact of the interaction of visible light from localised sources with the environment. Artefacts that are common in optical microscopy are caused by diffraction, interference, and refraction. However, because the human brain is familiar with these artefacts (from a very early age we become familiar with the effects of illumination in producing shadows, with interference effects producing the rainbow patterns in soap bubbles, with refraction effects changing the apparent depth of a swimming pool, *etc.*), it will try to interpret images obtained by different techniques in terms of our experience with visible light phenomena.

For example, in SEM the charging up of areas of a specimen can result in secondary electron images that appear to show the presence of shadows but which are actually the result of the incident electron beam being deflected by the electrostatic potential of the charged object. In backscatter images, a thin membrane may appear darker than the surrounding supporting substrate due to the absence of scattering material beneath a certain depth, but without prior information could be interpreted as indicating the presence of a different material on the surface, when, in reality, the surface is homogenous.

It is also clear that any techniques which are rich in artefacts are also potentially rich in contrast mechanisms. However, of key importance to tapping into this potential is the ability to separate these contrast mechanisms, or at least be able to distinguish between data affected by each contrast mechanism. In the case of the optical microscope, which is rich in contrast mechanisms, many methods have been developed to promote one contrast mechanism at the expense of others with great success, and without a prohibitive increase in the complexity and expense of the instrument.

As another example, scanning electron microscopy has become established due to its reasonable speed (video-rate images), relatively low cost, well-understood artefacts, minimal sample preparation (at least in the case of non-organic specimens, which, if they are insulating, need to be coated with a charge conduction layer), high magnification (at least two orders of magnitude better than with an optical microscope), and low damage (electron-beam damage, pump-oil contamination and electrostatic discharge) caused to one of the most commonly viewed specimens (semiconductor devices). In general, even one application, if of sufficient importance or interest, can be enough to establish a microscopy technique, even if all the contrast mechanisms and artefacts are not well understood.

Over the last two decades the importance and profile of microscopy has been steadily increasing. This is due, in no small measure, to the world-wide growth of the consumer electronics industry. The rise of the computer and the onset of the information age have created a demand for faster and faster computers capable of handling more and more data in less and less time. These computers are powered by semiconductor microprocessors, use semiconductor circuitry for temporary data storage (random access memory or RAM) and use magnetic, magneto-optic or all-optical techniques for permanent data storage. In order to satisfy the demands of the consumer while achieving lower costs through increased volume, the chosen route has been to manufacture at smaller and smaller dimensions. Faster processors run at higher clock speeds and result in greater power consumption and therefore greater likelihood of overheating. In addition the demand for portable computers coupled with the limited life of batteries has also put pressure on reducing the power consumption of these processors. This is only possible by shrinking the dimensions of processor components and reducing the logic voltage level at which these components work. Add to this the demanded increase in functionality and

the situation becomes one of integrating a greater density of lower power components on an increasingly larger die (which must then have a lower defect density in order to make manufacture financially viable). Currently desktop computers contain microprocessors with component features patterned at dimensions down to 0.18μm, with features of 0.12μm already demonstrated as being industrially feasible. Industry projections suggest that features down to 0.1μm may be possible without requiring a radical change in manufacturing methodology (*i.e.* from photolithography to extreme ultra-violet lithography, *x*-ray lithography or electron-beam lithography).

However as the volume of data being processed increases, other bottlenecks appear – transmission of such data from one computer to another (which could be thousands of kilometres away), from temporary memory to the processor and from permanent storage to temporary memory. The first of these has resulted in the rapid growth of the communications industry in recent years and the increasing adoption of optical fibre networks for data transmission. This has also been mirrored by the increasing research into optoelectronic devices where optical components are integrated with very fast electronic circuitry using semiconductor technology to enhance communication throughput and bandwidth. The second has forced the random access memory manufacturers to follow the same route as the microprocessor manufacturers, *i.e.* to integrate more capacity on the same size of die so as to increase or maintain the rate at which data can be accessed. The third has forced media manufacturers to look at smaller storage dimensions in order to increase capacity, data access speed and reading or writing speed.

All these factors have caused the world's research efforts to focus increasingly on: (i) studying the effects of manufacturing and operating electrical components at even smaller dimensions running on even less power (in the extreme this includes quantum, ballistic and single-electron devices); (ii) the development of magnetic, magneto-optic and optical materials capable of offering high contrast at even smaller dimensions, for data storage purposes, along with the techniques necessary for reading and writing this data; and (iii) the development of techniques for examining specimens at these length scales for: (a) the failure and quality analysis of 'microchips' and materials during the manufacturing process; (b) the evaluation of the correctness of models used to optimise device and material designs; and (c) the understanding of the small-scale behaviour of new and existing materials under different external conditions. This thesis will be primarily concerned with item (iii).

From the previous discussion it should come as no surprise that much of the early work described in the remaining sections of this chapter was carried out in the industrial research laboratories of semiconductor manufacturing, media manufacturing and communication companies, as it is these companies who stand to gain the most from the development of such techniques. However, these 'surface science' techniques, have also been of use in other completely unrelated areas, such as biology, biochemistry,

electrochemistry, polymer science and metallurgy, which have also contributed to the recent growth in small-scale imaging techniques.

To take a microscopy technique down to the scale of a few tens of nanometres is not a trivial matter. It is tempting to assume that the factor limiting the magnification of any existing technique is the less than ideal performance of the components of the instrument. However, while this is true for certain techniques, for others, ideal components would still not give the desired improvement in magnification. This is because microscopy relies on interactions that are governed, and ultimately limited, by the laws of physics. For example, with an SEM it is possible to view gaps of 1nm but this can never be the case with a conventional optical microscope. In fact, higher magnification with the SEM is primarily limited by the quality of the lenses used, while in the optical microscope negligible improvements would be obtained by using 'perfect' lenses (such microscopes are already diffraction limited with today's lens technology).

The ability of a microscopy technique to operate at small dimensions is often described in terms of its magnification. However, this is not the best way of measuring the capability of a technique as can be seen by considering what happens if a small section of a photograph is enlarged (using photographic techniques), and then a small section of the enlargement is enlarged and so on... Then the final photograph is an enlargement of a very small area of the original, representing a substantial overall magnification. However, it is clear that this final photograph cannot contain more detail than in the original. Whilst, the first enlargement may appear to show more detail than the original, this is only because that detail was contained in the original photograph at a size difficult for the human eye to see. Placing the original under an optical microscope can rectify this without the need for a photographic enlargement. So it becomes clear that the power of a microscopy technique should be based on the level of *usable* and *useful* detail it delivers, as opposed to the magnification it produces. This attribute is usually expressed in terms of the *resolution* of the technique.

One definition of the resolution is the smallest difference in the actual result of a measurement that can be detected. This limit, may be set by the components of the microscope, or by the laws of physics (Heisenberg uncertainty principle). In the case of the optical microscope this definition is given numerical form through the Rayleigh criterion [Jenkins and White[3] (1981)]. A typical microscope can have many different resolutions corresponding to different directions or contrast mechanisms, *e.g.* the lateral resolution of an optical microscope may differ from its depth resolution.

As this thesis is concerned with the engineering of microscopy techniques, all the concepts discussed in this section are applicable to the work presented here. The remainder of this chapter is concerned with describing the subset of microscopy techniques that define scanning probe microscopy. A full coverage of the field of microscopy would occupy many volumes and as such is beyond the scope of this thesis.

1.2 THE SCANNING TUNNELLING MICROSCOPE

The scanning tunnelling microscope (STM) was invented by Binnig, Rohrer, Gerber and Weibel [Binnig and Rohrer[4] (1982); Binnig, Rohrer, Gerber and Weibel[5] (1982); Binnig, Rohrer, Gerber and Weibel[6] (1982)] and introduced not just a new microscopy technique but a whole new concept for microscopy. It was this concept which inspired most of the other techniques described in this chapter. The Nobel prize was awarded to the inventors five years later, in 1986, for this very reason (jointly with the inventors of the SEM). This technique, its limitations and its applications have been described in some of the many excellent books and reviews of the subject [see, for example, Hansma and Tersoff[7] (1987), or Wiesendanger[8] (1994)]. However, it is instructive to discuss some aspects of this technique here, since it contains some elements common to all scanning probe microscopy techniques.

1.2.1 Historic Origins and Principles of Operation

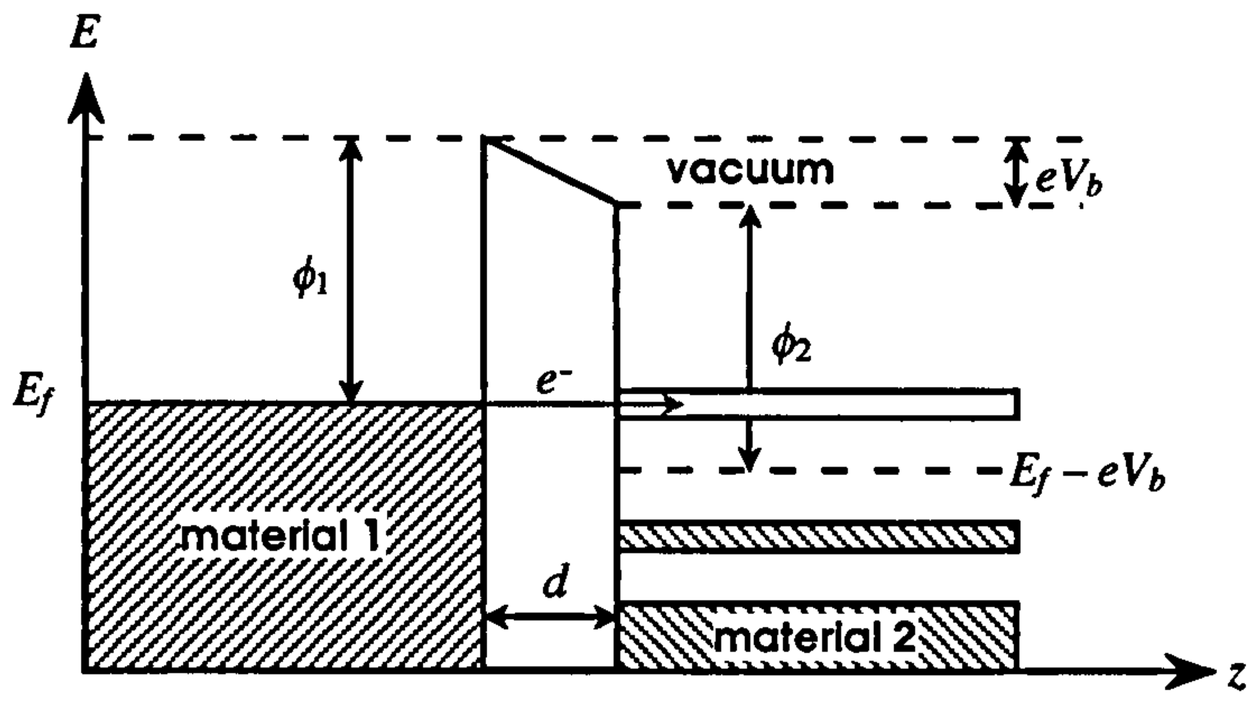


Figure 1.2.1: Tunnelling into a material with a bandgap. Spectroscopy is carried out by varying the bias and observing changes in current caused by variations in the density of occupied and unoccupied states as a function of energy.

The principle behind the technique is an extension of the Nobel prize winning work of Giaever, carried out in the late 1950s and early 1960s, which established electron tunnelling spectroscopy as a useful technique, particularly for measuring band gaps in superconductors [Giaever[9] (1960)]. [A more detailed discussion of the origins and historical context of STM is given by Walmsley[10] (1987).] In electron tunnelling spectroscopy, the conductance due to electrons tunnelling between two conducting films separated by an insulating layer much less than 10nm thick is measured as a function of the bias between them. Clearly the current can only flow from filled states in one material to empty states in the other (which are located at the *Fermi level* in the low temperature limit). The bias acts so as to shift the Fermi levels of the two materials with respect to each other (as shown in Figure 1.2.1) and thus scans through the energy states of the two materials. If there is a range of energies where there are no electron states, as in

Figure 1.2.1, then in the low temperature limit (where there is insufficient thermal energy for electrons to occupy higher energy states than necessary to fulfil the Pauli Exclusion Principle), the current will drop below the noise floor as the bias is scanned across the band gap.

Elementary quantum mechanics shows that the probability of an electron tunnelling with energy, E, through a barrier of height, V, and width, d, is given by [Fowler and Nordheim[11] (1928); Frenkel[12] (1930)]:

$$p(E) = A\exp\left(-\frac{4\sqrt{2m_e}\pi}{h}\sqrt{V-E}d\right), \qquad (1.2.1)$$

where A is a normalisation factor, h is Planck's constant and m_e is the mass of the electron. This exponential dependence on barrier width means that most tunnelling will take place at the 'point' where the inhomogeneities in the insulating material are such that the barrier is narrowest. Thus most of the early problems with this technique were due to the difficulty in obtaining thin insulating films free from defects allowing ohmic current flow between the two materials. For low bias voltages, the height of the barrier above the Fermi level is usually dominated by the *work functions* of the two materials (the energy required to remove an electron from a material to a point an infinite distance away) which is typically of the order of 4eV for metals. Putting this value into Equation 1.2.1 shows that the characteristic decay length for this interaction is about 0.5Å (*i.e.* a change in the barrier width of 0.5Å will alter the tunnelling current by a factor equivalent to the inverse natural logarithm of 1 or put another way, a 1Å change in barrier width causes the tunnelling current to change by almost one order of magnitude).

The scanning tunnelling microscope was an attempt to obtain (a) an instrument capable of varying the tunnelling gap in a controllable fashion; and (b) an instrument capable of investigating spatial variations in electron tunnelling spectra. To this end the instrument consists of a sharpened tip, usually tungsten, held in close proximity to the specimen being examined, but making use of piezoelectric actuators to allow the tip to be moved in all three dimensions. These piezoelectric actuators, whilst exhibiting non-linear behaviour and suffering from hysteresis and creep, permit very small displacements to be made by the application of small, but achievable, voltages – a particularly compact implementation allowing motion in all three dimensions uses the piezoelectric ceramic, lead zirconate-titanate (PZT), mounted in a tubular arrangement and has a response of a few nm/V [Binnig and Smith[13] (1986)]. [This arrangement has become popular in many commercial implementations of scanning probe microscope systems.] In order to maintain gaps as small as a few nanometres, the instrument has to be isolated from sources of vibration, such as building vibrations, acoustic vibrations, etc. This necessitates the use of damped springs, air (optical) tables, rubber mounts, *etc.*, upon which the critical parts of the instrument can be placed.

1.2.2 Modes of Operation

The STM can be operated in many modes. In the most commonly used mode, *constant current mode*, the tip is raster-scanned over the surface in such a manner as to maintain a constant tunnelling current between the tip and the specimen, achieved by altering the tip-specimen distance. Figure 1.2.2 shows a typical implementation of this mode. The axis joining the tip and the sample is conventionally taken to be the z-axis. As the tip is scanned over the surface, the tip-specimen distance is sampled at a set of discrete points (x, y), and this data is used to form an image of the specimen. The scan-rate is typically limited by the lowest mechanical resonance of the system (due to the use of an integrating amplifier to control the tip-specimen distance). The lowest resonance, usually associated with the translation stages and piezoelectric actuators, is of the order of a few tens of kilohertz. Typical tunnelling currents are of the order of 1nA for a bias of a few tens of millivolts and a tunnelling gap of 1nm (on metallic surfaces).

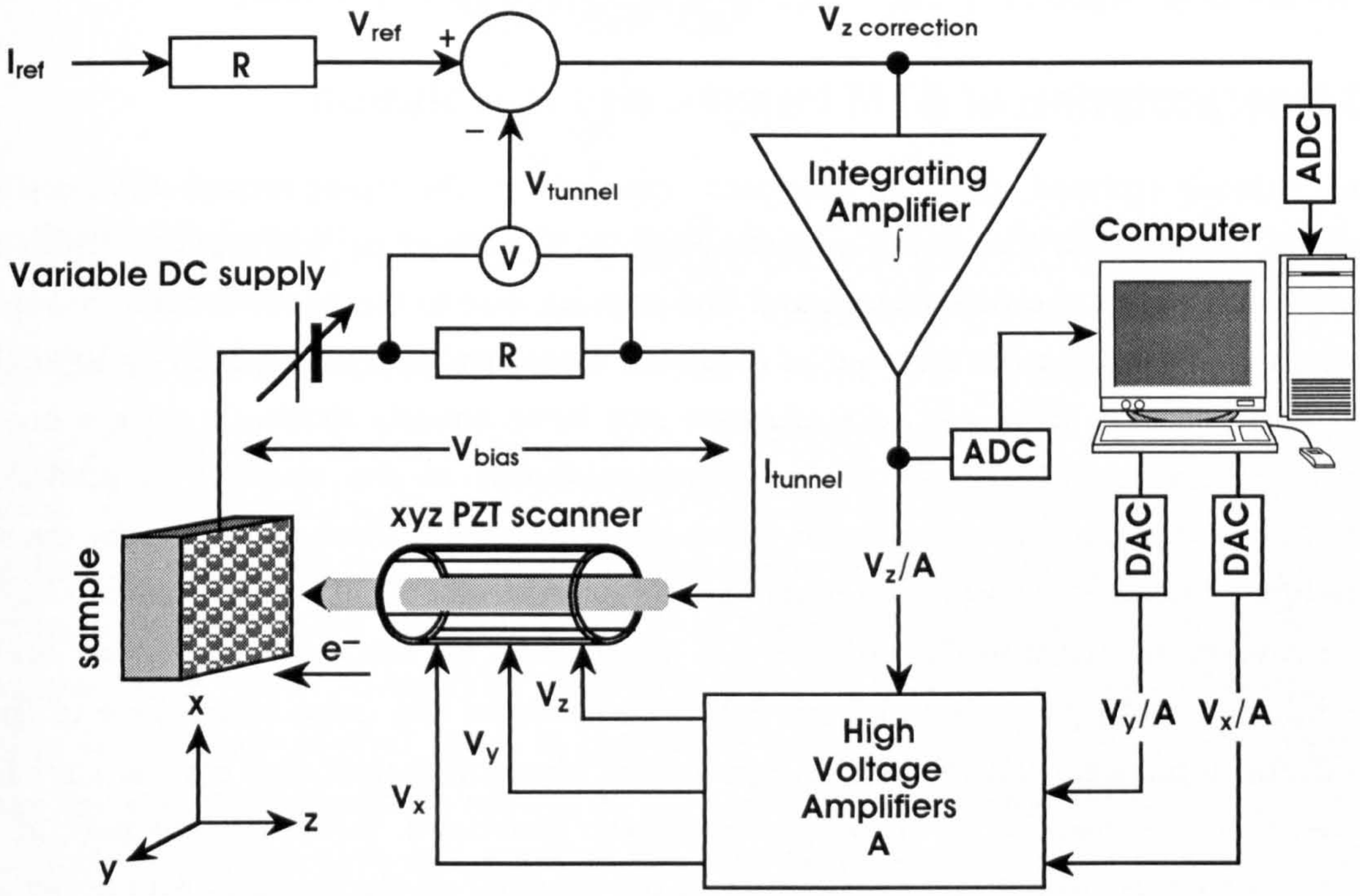


Figure 1.2.2: Implementation of an STM system for constant tunnelling current operation.

In another mode of operation, *constant height mode*, the tip is raster-scanned as before, but without feedback, and the image is formed by measuring the tunnelling current (at constant bias) or bias (at constant tunnelling current) at each position. The advantage of this is that faster operation is possible since the tip can be scanned at any frequency away from mechanical resonances (rather than being limited to the lowest mechanical resonance) while the measurements are made in a bandwidth between these resonances. However, the specimen must be relatively flat, not tilted appreciably and the scan area must be small since the tunnelling gap is typically less than 1nm.

The STM is also a tool for performing spatially resolved electron tunnelling spectroscopy, when the technique is often given the name scanning tunnelling spectroscopy or STS. Again several modes are possible. In one mode, based on the constant current mode described above, the feedback loop is briefly 'switched off' at each co-ordinate and an I-V curve is taken. From this the conductance (normalised to the conductance prior to breaking the feedback loop to avoid the effects of varying tunnelling gaps) can be plotted as a function of bias. This technique allows the normal STM image to be obtained simultaneously with the spectra and is often known as current-imaging tunnelling spectroscopy (CITS). The instrument needs to be fairly stable to perform this technique (to prevent the tip drifting while the spectra are obtained) and is often carried out under conditions of ultra-high vacuum (clean surface) and low temperature (minimum thermal energy). More refined methods are also possible [Tromp, Hamers and Demuth[14] (1986)].

As might be expected, numerous other STM variants have been developed in the years since its invention – good coverage is given by *e.g.* Wiesendanger[8] (1994).

1.2.3 Interpretation of STM Images and Resolution

In the constant current mode of operation, changes in the tip-specimen distance are caused by two effects. The first, and simplest, is a change in the topography of the specimen (and hence the tunnelling gap). The second, due to the spectroscopic nature of the STM, is a change in the number of available states for tunnelling from or to at that energy and position. Both of these changes can have equally dramatic effects on the tunnelling current and are therefore indistinguishable in the absence of additional information. [Consider for example, the effect of a topographic change of 2Å – the current will change by nearly two orders of magnitude. Consider the effect of a tip moving from a surface covered in metal to the surface of a semiconductor when the bias is within the semiconductor band gap – the current will drop to below the noise level.] Due to these dual effects, interpretation of STM images is not straightforward and can be very bias dependent. In a sense the STM images electronic structure, but only that part of the electronic structure which can participate in the transfer of electrons at an energy and momentum determined by the bias and the work functions of the tip and specimen. A more appropriate description of the STM might be that it is an instrument for imaging topography in the absence of variations in (specific parts of the) electronic structure and vice-versa. The problem is that both variations are present on most real surfaces.

Clearly, the STM as described above is not suitable for imaging insulating surfaces. However, the technique demonstrates incredibly high vertical resolution (along the *z*-axis). A tunnel current variation of 10% corresponds to a height change of only 5pm – less than the 'diameter' of a typical atom. This also results in high lateral resolution as illustrated in Figure 1.2.3: all things being equal, in terms of the nomenclature used in the figure, the T2-S1 and T2-S3 interactions are around 4 times weaker than the T2-S2 interaction while the T1-S2 and T3-S2 interactions are over 150 times weaker for the

values used in the figure. Thus with a tip of the correct profile and under the right conditions the STM is able to resolve individual atoms. However, the question of resolution in STM is not a simple one since it depends on the geometry of the tip and specimen, the electronic structure of the tip and the specimen, the bias between them and the tunnelling gap size. A simple discussion of some of the most important issues is given by Hansma, *et al.*[7] (1987).

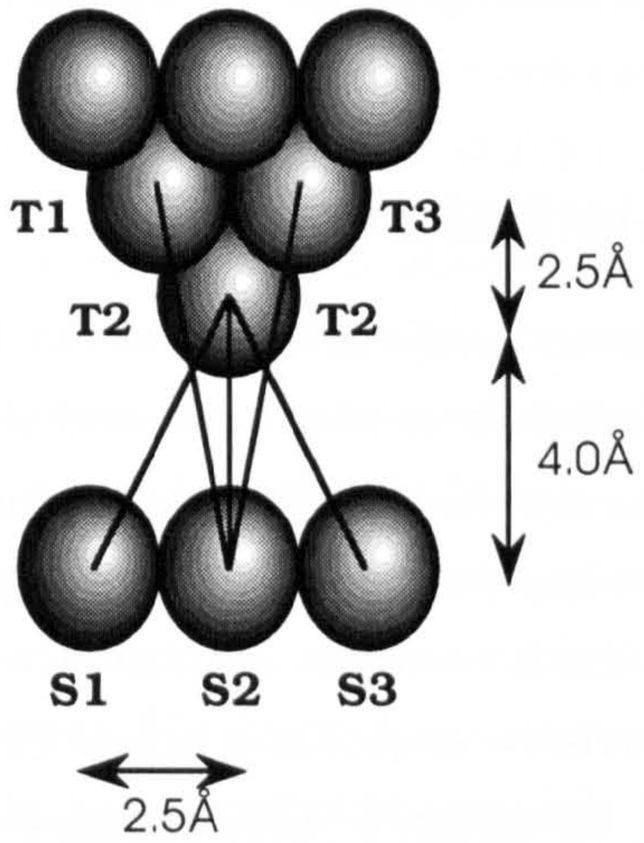


Figure 1.2.3: Diagram illustrating the origin of the high lateral resolution in the STM. 'S' atoms are part of the sample, 'T' atoms are part of the tip.

1.2.4 STM Tip Fabrication

The original STM tips were made from tungsten using electrochemical etching. Typically a solution of sodium hydroxide or potassium hydroxide is used as the electrolyte with the tungsten wire forming the anode and some other material (such as a graphite rod) forming the cathode. The application of a bias induces the etching process. A variety of techniques have been employed for making the tips as sharp as possible, from attaching weights to the wire so that when it reaches a certain thickness the weight falls and breaks the electrical circuit, to relying on the meniscus of the electrolyte and diffusion of the reaction products to form a taper in the wire until its own weight causes it to break and lift the suspended end out of the solution [Melmed[15] (1991)]. Such techniques can result in tips with a radius of curvature of less than 50nm.

However, tungsten is not chemically inert, making it unsuitable for use in chemical solutions, and its tendency to oxidise leads to problems in tunnel junction stability even in air so that it is often preferable or necessary to use tips made from other metals. The choices are limited to noble metals for which a good electrochemical etch can be found. Iridium is a good choice because of its mechanical strength (Young's modulus of 528GPa) but platinum or platinum-iridium are also commonly used. An alloy is usually preferable due to its improved plastic deformation properties so platinum-iridium (usually 90%/10% as used in thermocouples) is a popular choice. This can be electrochemically etched in, for example, a mixture of calcium chloride and hydrochloric acid [Musselman

and Russell[16] (1990)]. Alternatively, the wires may simply be cut or mechanically ground. The main advantages of electrochemically etched tips is that they can be made to taper thus producing a tip with a narrow cone angle (high aspect ratio). This makes the tip better able to access deep and narrow crevices on the specimen surface in comparison with a tip cut from a relatively large diameter wire. From a practical standpoint, mechanically ground and electrochemically etched tips are preferable to cut tips because it is much easier to judge the proximity of the tip to the specimen surface during the initial coarse approach.

1.2.5 Applications and Limitations

The applications of STM are many and varied in areas from superconductivity and semiconductors to biology, electrochemistry and nanolithography as reviewed by Hansma, *et al.*[7] (1987); Schneir, *et al.*[17] (1988); Jahanmir, Haggar and Hayes[18] (1992); Penner[19] (1993). The technique has been shown to work in gases (such as air) *e.g.* Miranda, *et al.*[20] (1985), liquids (such as water, electrolytes and alcohols) *e.g.* Carrejo, *et al.*[21] (1991); Song, Morch, Carneiro and Tholen[22] (1994) and under vacuum *e.g.* Gimzewski, Modesti, David and Schlittler[23] (1994), and at temperatures ranging from dilution fridge temperatures through to several hundred degrees above room temperature *e.g.* Gaisch, *et al.*[24] (1994); Ichinokawa, *et al.*[25] (1994). Each of these environments has its own advantages and disadvantages depending on the nature and requirements of the specimen being examined, and the information being sought. This flexibility is one of the advantages of STM over electron microscopy techniques. However, leakage currents due to the ionisation of air or the presence of ions in liquids means that the range of bias that can be used in these environments can be more limited (a 1V bias across a 1nm gap corresponds to an electric field strength of 10MV/cm).

Many biological specimens are better imaged in a liquid environment, where surface distortions due to the specimen 'drying out' can be reduced. As these specimens are typically soft and mobile and are not good conductors they are typically adsorbed onto a conducting substrate which is relatively free of electronic and topographic features, usually highly oriented pyrolytic graphite (HOPG) [Baro, Miranda and Carrascosa[26] (1986)], or coated with a thin metallic layer (which has the disadvantage of requiring vacuum conditions for deposition and may deform the shape of the specimen on an atomic scale). Electrodes can be studied by the STM while in their natural environment, but require an insulating layer to be deposited or coated over the tip, leaving just a small length at the end exposed to the ions in the solution, to minimise leakage currents, *e.g.* Carrejo, *et al.*[21] (1991).

The dual nature of STM images has led to it mostly being applied to the ordered surfaces of metals and semiconductors where the periodicity of the atomic structure can be postulated, calculated or measured by other means. Thus STM has mainly been used as a complementary technique for surface science studies of superconductors,

semiconductors, metals, and metal alloys. In particular, since atoms on the surfaces of many materials tend to reorganise or *reconstruct* to minimise surface energy [Myers[27] (1990)], the STM has proven very useful (when combined with theoretical calculations and other techniques) in determining the precise nature of these reconstructions. The behaviour of surfaces is very important to the semiconductor industry, particularly because as the dimensions of semiconductor devices decrease, the surface plays a larger role in determining electrical behaviour. In addition, the spectroscopic capabilities of the STM also allow the presence of surface states to be detected and spatial variations in the band gap to be measured across a semiconductor surface.

The problems with STM as a technique are almost as varied and numerous as the number of modes of operation. The first and foremost limitation is that the specimen being examined must either be conducting or adsorbed onto a conducting substrate – this makes it awkward when applying STM to semiconductor devices having conducting, semiconducting and insulating areas on the same specimen. The second problem is the difficulty in interpreting images. Obviously the STM is not the ideal metrological tool due to its spectroscopic sensitivity. However it is also not the ideal spectroscopic tool, not only because topography plays a part in determining the image, but also because the spectroscopic information it provides is often not sufficient to unambiguously identify the particular atoms which are present on the specimen surface [Wiesendanger[28] (1994)].

Some cases where this has led to confusion are discussed by Tromp[29] (1989), who also describes an experiment showing the bias dependent nature of images of the (110) surface of gallium arsenide, where the gallium atoms can be imaged with one bias and the arsenic atoms with another. This bias dependence has proven useful for studying the early (sub-monolayer) stages of the deposition of metals onto semiconductor surfaces since the metals show up clearly in both empty and filled state images, whereas the semiconductor shows up better in one rather than the other, depending on doping. A similar technique can be used to look at the adsorption of insulating atoms, such as oxygen, onto conducting surfaces.

One particular problem is caused by the geometry of the tips used. Since tunnelling is normally dominated by the atom closest to the surface, if a tip contains more than one large protrusion (more common for tips which are mechanically ground or cut), the atom involved in the tunnelling can switch between all the protrusions according to the geometry of the sample. This occurs even when there is only a single major protrusion since the atoms at the end of the tip are often mobile and can move around causing the tunnelling atom to switch. This artefact is known as *tip switching*, and in extreme cases, images can appear to be suffering from 'double vision' (or even higher orders). This artefact is a common problem with most scanning probe microscopy techniques, however with the STM, it is more difficult to judge when this is taking place if the tunnelling atoms are close together, since at small length scales the actual shape of the specimen is not known.

Another tip-related problem is caused by there being many atoms at the end interacting simultaneously. In this case the images may appear to show the expected atomic corrugations but a single atomic defect may go unnoticed due to the averaging effect caused by having multiple tips [Albrecht, *et al.*[30] (1988)]. This artefact is known as *multiple-tip averaging*. The situation is akin to rubbing two corrugated surfaces against each other – the surfaces will move up and down with the correct periodicity but if a single corrugation is missing it will go unnoticed. Again this is a common problem for most atomic resolution scanning probe microscopy techniques.

Additional problems are also caused by the tip modifying the surface by moving atoms around, by the tip picking up atoms from the specimen resulting in a change in the nature of the tip tunnelling atom, or by the tip crunching into the specimen due to either, a change in topography too sudden or large for the feedback loop to respond, or by a sudden reduction in conductivity. In common with most scanning probe microscopy techniques, tips are a consumable item and rarely last beyond a few days of use. In those techniques where the tip is extremely close to the specimen, they often suffer from continual modifications during scanning. Other problems of the technique include slow image acquisition rates and the need for verification by repetition, calculation and the application of other techniques.

Despite these problems, as a technique capable of producing 'atomic resolution' images, and with the dedication and the perseverance of its advocates, STM has helped to revolutionise our understanding in many key areas of surface behaviour and organisation. Perhaps even more importantly, it has introduced a new concept to the microscopy world – scanning probe microscopy.

1.3 THE ATOMIC FORCE MICROSCOPE

As discussed in the previous section, STM satisfies the criteria necessary for a useful scanning probe microscopy technique: it requires a probe which can be held in close proximity to the specimen and an interaction mechanism with a very strong characteristic decay length. If both these attributes are combined with a probe of sufficiently small dimensions, so as to limit the interacting regions of the specimen and the probe primarily to single atoms, then atomic resolution is possible. The greatest weakness of the STM is that the interaction is complex enough to support many comparable contrast mechanisms which can act simultaneously or individually and without control from the user. However, this is also its greatest strength and has led to the many modes of STM operation in use today. Ironically, understanding microscopic behaviour would have been easier if a 'simpler' interaction mechanism could have been used. In 1986 a 'simpler' technique was found by Binnig, Quate and Gerber[31] (1986). This technique was atomic force microscopy.

1.3.1 Principle of Operation

The idea stems from the fact that there are many types of forces in nature which act so as to bind materials together. Classically five main types of binding mechanisms are normally identified: ionic binding, covalent binding, metallic binding, hydrogen binding and Van der Waal's binding. However, forces are a very general expression of the presence of an energy gradient and so numerous other forces can be identified at the atomic level. These forces are, in general, very complex and will not be dealt with in detail here. For a more thorough overview see Sarid[32] (1991).

At first sight the complex nature of these forces (some of which are not additive in the normal sense) seems to suggest that any technique based on these interactions would be even more rich in conflicting contrast mechanisms than STM. However, in practice it turns out that integrating and averaging all the forces between two macroscopic bodies (tip and sample) results in a potential that can usually be approximated by a simple power law expression.

That this should be possible and the resulting interaction being simpler than that of STM is not immediately clear, however, it can be explained simplistically in the following manner: forces acting on components of a system can be viewed in a more general sense as the result of a difference in potential energy between spatially separated components and as the force is simply the rate of change in potential energy with position, it becomes a vector, with both magnitude and direction. However, the potential energy itself is scalar and so no matter how complex the system, these potential energies can be summed up individually for each system component, and its gradient will be the resultant force acting on each component.

Whilst the expression for the potential energy itself may be complex, and may involve vectors as parameters and integrals for valuation, the procedure for accommodating additional contributions is very simple. Often, due to the huge differences in magnitude between the potential energy contributions of each interaction mechanism, the entire problem can be adequately approximated by the dominant interactions, at least one of which must be attractive and one repulsive to allow an equilibrium state to be defined. Usually it is not necessary to include more than two repulsive and attractive interaction mechanisms to obtain a good approximation to system behaviour. Since the dominant interactions for any given equilibrium situation can be readily found by comparing the size and distance dependence of their potential energy contributions, the system configuration can be readily optimised to enhance specific interactions. Unfortunately, in the case of STM, the tunnelling current can in no way be explained as arising from the distance dependence of a single scalar value, making it substantially harder to determine the dominant interactions within a particular system.

Based on this approach, the case of a two body system can be analysed in a simplistic manner. Using the fact that the potential energy must tend to a constant as the

separation tends to infinity (so that forces tend to zero), the potential energy function, $W(r)$, where r is the separation between the two bodies, can be viewed as being a function of the reciprocal of separation and so can be expressed using a Maclaurin expansion as:

$$W(r) = \sum_{i=1}^{\infty} \frac{\alpha_i}{r^i}, \tag{1.3.1}$$

for some α_i, where the arbitrary zero of potential energy has been taken to be numerically zero. The use of the reciprocal of separation automatically ensures that the potential energy tends to zero as the separation tends to infinity. For small distances only the first few terms will be significant and the rest can be neglected. This is a general method for expressing forces in power law terms.

For simplicity, the remainder of this section will illustrate AFM using a simple interaction mechanism for which only two terms of the expansion (one attractive and one repulsive) are an adequate approximation. This simplifies the integrals involved and allows the basic principles to be explained without focusing in detail on the specifics of any particular interaction mechanism. The chosen interaction will be that of the Van der Waal's force. This force dominates AFM strictly only in an ultra-high vacuum environment, in the absence of the capillary forces which tend to dominate in air due to the presence of a surface layer of water.

The Van der Waal's force acts so as to bring neutral atoms and molecules together. This has its biggest impact on the behaviour of gases. It is caused by the correlation of what would otherwise be random (thermally induced) dipole fluctuations which are a consequence of the separation of charge between the nucleus and surrounding electrons (in the case of atoms) or shifts in the positions of shared electrons from one atom to another (in the case of molecules).

If this force were allowed to act unrestrained, it would diverge to infinity as the atomic separation decreased to zero. This does not happen because of the electrostatic repulsion between nuclei caused by a reduction in screening due to the forced reconfiguration of the outer electrons in accordance with the Pauli Exclusion Principle. It is this force which is responsible for the lack of compressibility of solids. Between the attractive and repulsive forces lies a position where each cancels – this is the equilibrium position of atoms which are bonded together. Since the lengths of chemical bonds are of the order of a few angstroms, it is clear that the repulsive forces must be very short range and therefore very strong. This behaviour can be approximated by the Lennard-Jones or 6-12 potential [Gould, Burke and Hansma[33] (1989)]:

$$W(r) = -4W_o\left[\left(\frac{\sigma}{r}\right)^6 - \left(\frac{\sigma}{r}\right)^{12}\right], \tag{1.3.2}$$

where $W(r)$ is the potential energy variation with separation, r, and the minimum energy value, $-W_o$, is obtained at a separation of about 1.12σ, where σ is the

separation at which the energy first drops to zero. Figure 1.3.1 illustrates the variation of the potential energy and force with separation for some typical values of minimum energy and separation. It can be seen that for separations smaller than the minimum energy distance, the energy rises sharply with decreasing separation (corresponding to a very strong repulsive force), while for separations larger than this, the energy changes much more slowly with increasing separation (corresponding to the much weaker attractive forces). The force, $F_r(r)$, can be obtained by differentiating Equation 1.3.2:

$$F_r(r) = -24\frac{W_o}{\sigma}\left[\left(\frac{\sigma}{r}\right)^7 - 2\left(\frac{\sigma}{r}\right)^{13}\right], \tag{1.3.3}$$

and is plotted in Figure 1.3.1 for some typical values of minimum energy and separation.

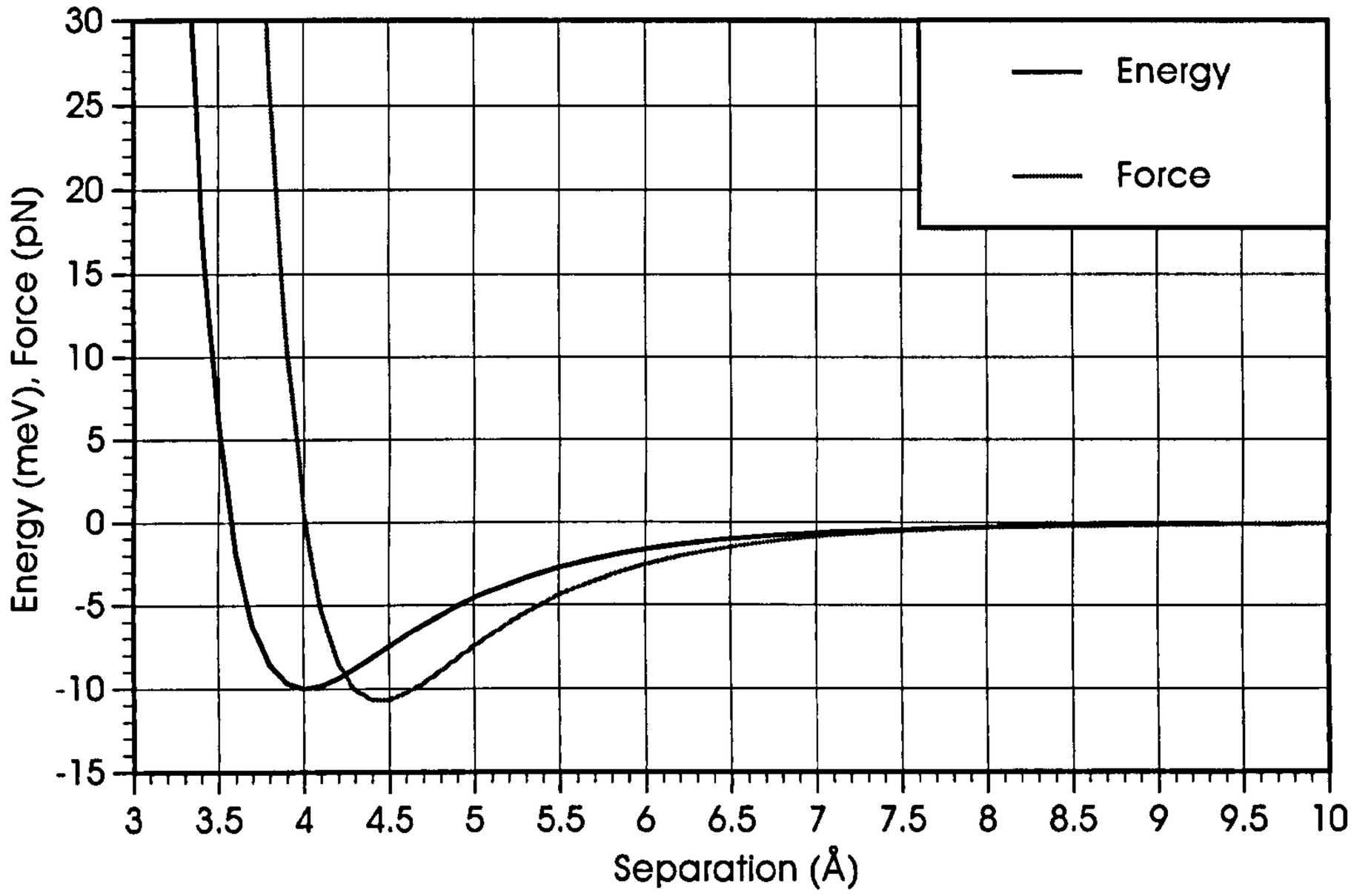


Figure 1.3.1: The Lennard-Jones potential for a minimum energy of -10meV at a separation of 4Å. Also shown is the negative of the force associated with this potential.

Due to its simplicity Equation 1.3.2 is often used to model general forces between objects even though it specifically applies only to Van der Waal's forces. To model the forces between two macroscopic bodies, such as a sample and a tip, the potentials must be integrated over the two interaction volumes. The regions are usually represented as a sphere (the tip) and planar surface (the sample). This then gives [see Sarid[32] (1991)]:

$$W(r) = -\frac{2}{3}\pi^2 W_o n_1 n_2 \sigma^5 R\left[\frac{\sigma}{r} - \frac{1}{210}\left(\frac{\sigma}{r}\right)^7\right], \tag{1.3.4}$$

where n_1 and n_2 are the number densities of the interacting atoms in the sample and tip, R is the effective radius of curvature of the end of the tip and all other parameters are as in Equation 1.3.2. [Note that, in general, Van der Waal's forces are not additive in this way, however, following on from Sarid[32] (1991), a first order approximation is made here.]

Then the force, $F_r(r)$, is given by differentiating Equation 1.3.4:

$$F_r(r) = -\frac{2}{3}\pi^2 W_o n_1 n_2 \sigma^4 R\left[\left(\frac{\sigma}{r}\right)^2 - \frac{1}{30}\left(\frac{\sigma}{r}\right)^8\right]. \qquad (1.3.5)$$

Using Equation 1.3.5 for a tip with an effective radius of curvature of 20nm, a nominal interatomic spacing of 4Å and with a bond strength of only 10meV (a typical value for the Van der Waal's interaction), it can be seen that the repulsive force has increased to 1nN when the separation is reduced to 1.9Å. This force would cause a deflection of about 1Å in a direction perpendicular to the largest plane of a rectangular piece of aluminium foil 10μm thick, 50μm wide and 450μm long. This foil cantilever would have a stiffness of about 10N/m while the repulsive force gradient is 130N/m so that a 0.1Å change in separation would cause a 1.3Å (or 130%) change in deflection. This is comparable to the sensitivity to separation achieved in STM and so atomic resolution should be possible. Figure 1.3.2 plots the energy, force and compliance for the values used above. As with STM, `it is the rapidity of the decay of the interaction with separation which allows such an amplification of deflection amplitude to be obtained. [Note that the separation is defined to be the distance between the surface and the end of the cantilever in the absence of any bending moment while the deflection is the departure of the interacting end of the cantilever from this position.] Most of the early problems with this technique were related to the manufacture of such compliant, yet robust cantilever probes and the measurement of cantilever deflection. Other problems, such as vibration isolation and fine displacement control had already been resolved during the development of the STM.

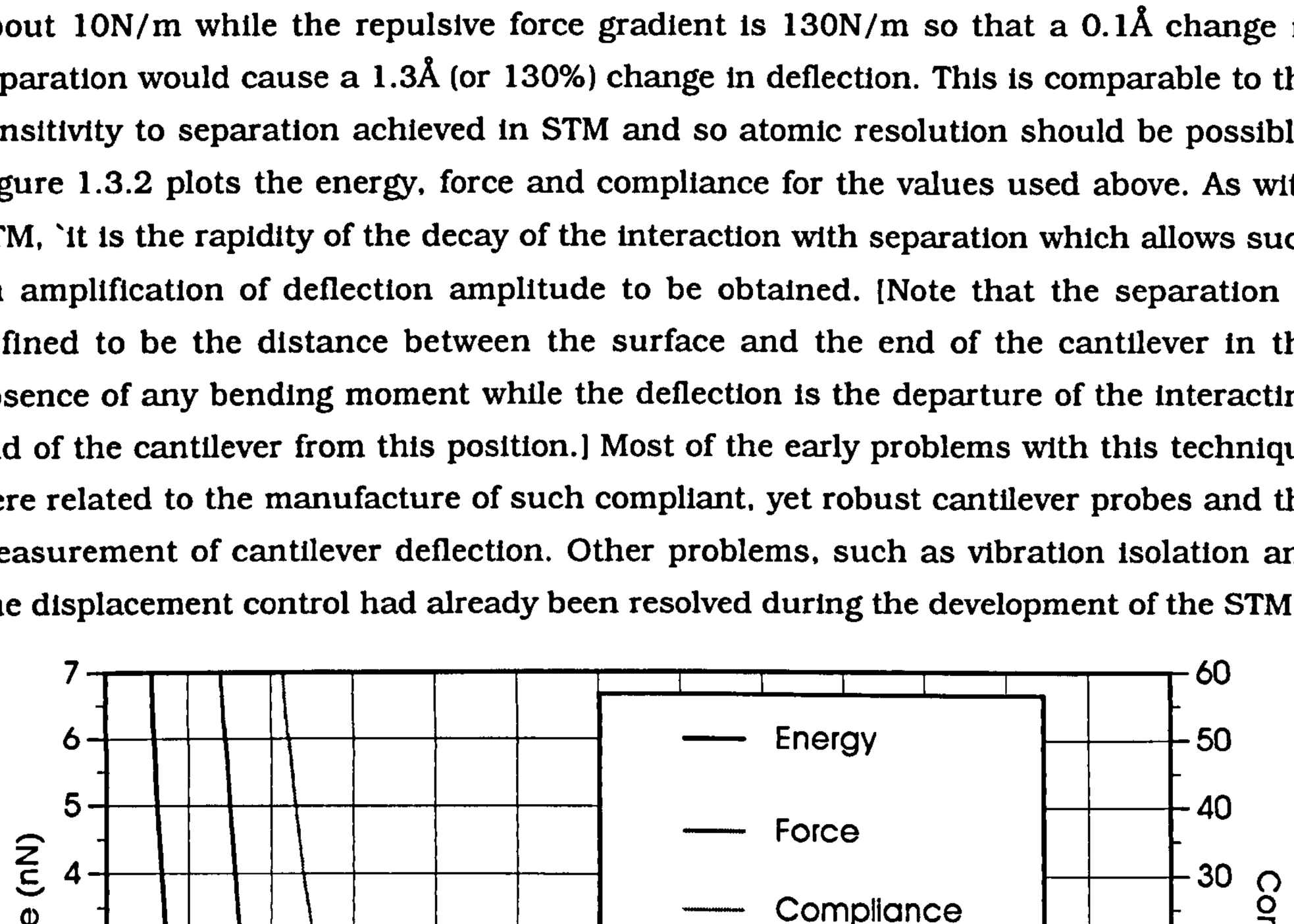


Figure 1.3.2: Typical energy, force and compliance curves for UHV AFM with Van der Waal's forces dominating. The graphs are for a minimum energy of -10meV and an interatomic spacing in both probe and specimen of 4Å, The probe is modelled as a sphere with a radius of curvature of 20nm while the specimen is modelled as an infinite plane. The minimum energy (equilibrium) position occurs at a separation of about 2Å.

1.3.2 Modes of Operation

Like the STM, the AFM can be operated in many modes. However, these modes fall into three main categories: *contact* modes, *non-contact* modes and *intermittent-contact* modes. The first of these are the most analogous to the normal STM modes, while the second makes use of the mechanical nature of the interaction dynamics to translate deflections to changes in resonant frequency. The last of these is a hybrid of the other two methods.

1.3.2.1 Contact Modes

In the contact modes, the cantilever is scanned across the specimen in close proximity to the surface (so that the interaction lies in the repulsive force regime of Figure 1.3.2) and changes in the cantilever deflection are monitored. The most commonly used mode in this category is *constant force mode*, where the deflection of the cantilever is held at a constant value as it is scanned across the specimen by varying the undeflected cantilever-to-specimen spacing. Figure 1.3.3 shows a typical implementation of this mode.

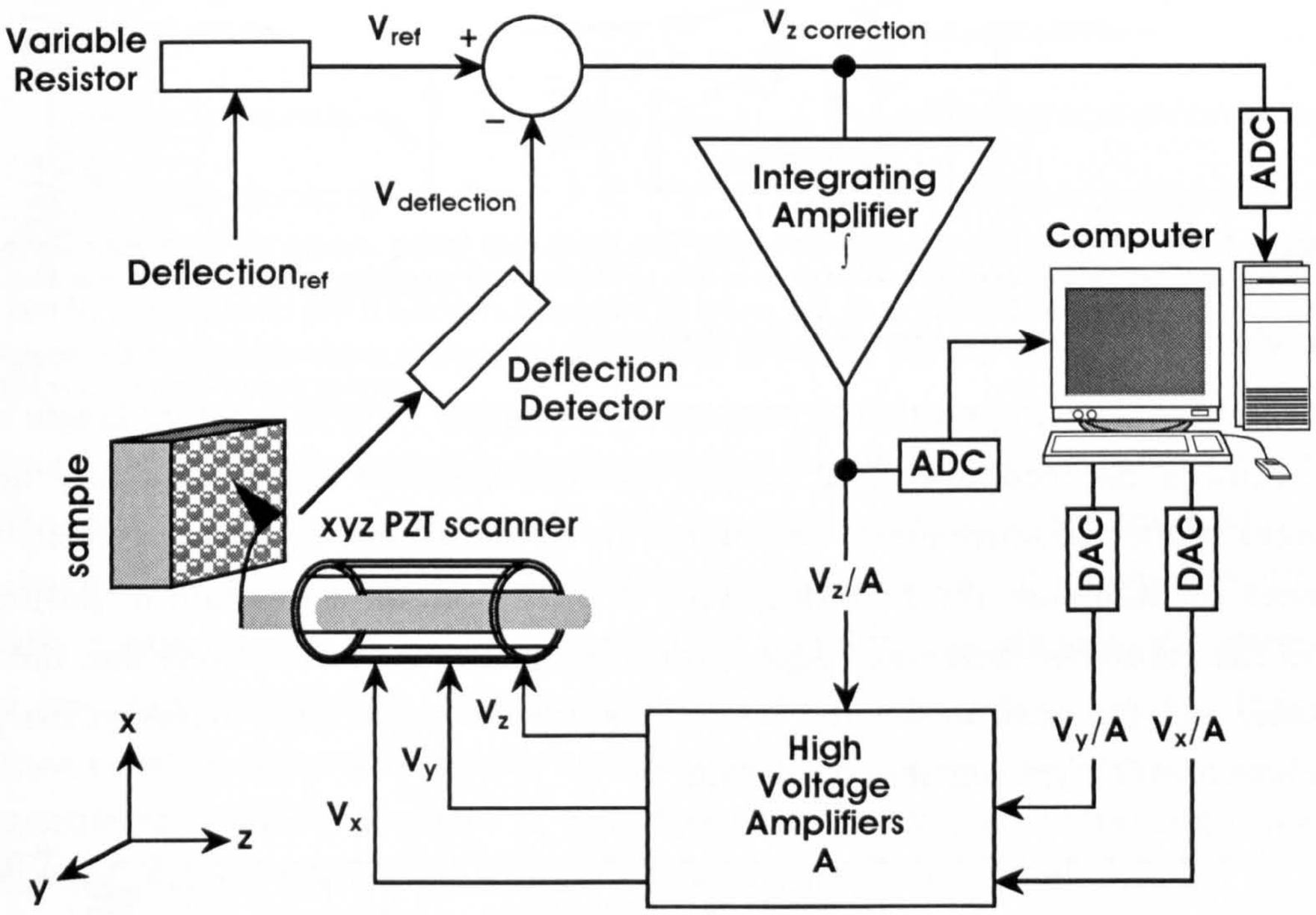


Figure 1.3.3: Implementation of an AFM system for constant deflection operation in contact mode.

As with the STM in constant current mode, the image of the specimen is formed by sampling the tip-specimen distance at a set of discrete points (x, y). Again the scan-rate is typically limited by the lowest mechanical resonance of the system (due to the use of an integrating amplifier to control the tip-specimen distance). Typical cantilever compliances are between 0.01 and 1N/m and operating forces range from ~10pN to ~100nN.

As noted earlier, the power law model for forces in UHV AFM (as described in §1.3.1) can be generalised to accommodate other forces and additional terms. For simplicity only the UHV AFM situation will be considered here. Since the restoring force of a cantilever is proportional to its compliance, k, and deflection from equilibrium, the equilibrium situation can be expressed as:

$$-k(u-d) = -\frac{2}{3}\pi^2 W_o n_1 n_2 \sigma^4 R\left[\left(\frac{\sigma}{d-s}\right)^2 - \frac{1}{30}\left(\frac{\sigma}{d-s}\right)^8\right], \qquad (1.3.6)$$

where all parameters are as given in Figure 1.3.4 and Equation 1.3.4. Thus maintaining constant force is equivalent to maintaining a constant deflection. Figure 1.3.5 represents this interaction as a force balance. From this figure it can be seen that the mechanical gain is given by the ratio of the spatial derivative of the force to the stiffness of the cantilever. As the force derivative is a function of the separation, this amplification is not constant but increases with decreasing separation. However, it is clear that the cantilever should be as compliant as possible for maximum sensitivity.

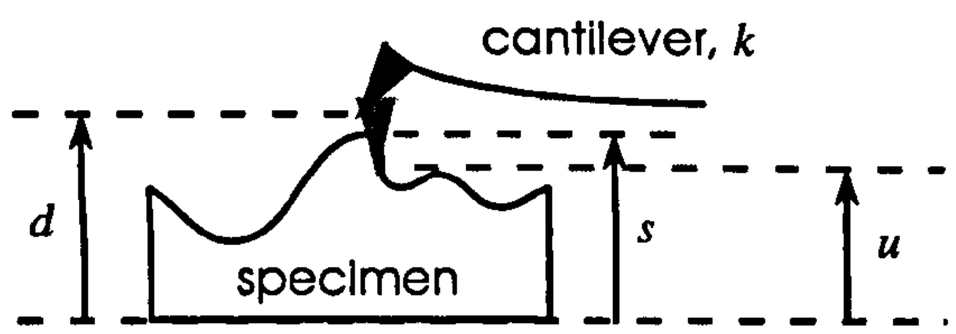


Figure 1.3.4: Parameters used to describe the constant force mode of operation. d is the deflected position of the probe, u is the undeflected position of the probe, s is the position of the sample surface at the point of interaction and k is the compliance of the cantilever in the direction normal to the cantilever.

As with the STM, faster operation is possible by using the AFM in *constant height mode*, which requires no feedback, and hence can be operated away from any system resonances while measurements are made in a bandwidth chosen to avoid these resonances. In this case the restoring force is monitored as a function of position by means of the cantilever deflection. Again, the specimen must be relatively flat, not tilted appreciably and the scan area must be small since the 'separation' between the probe and the specimen is typically less than 1nm.

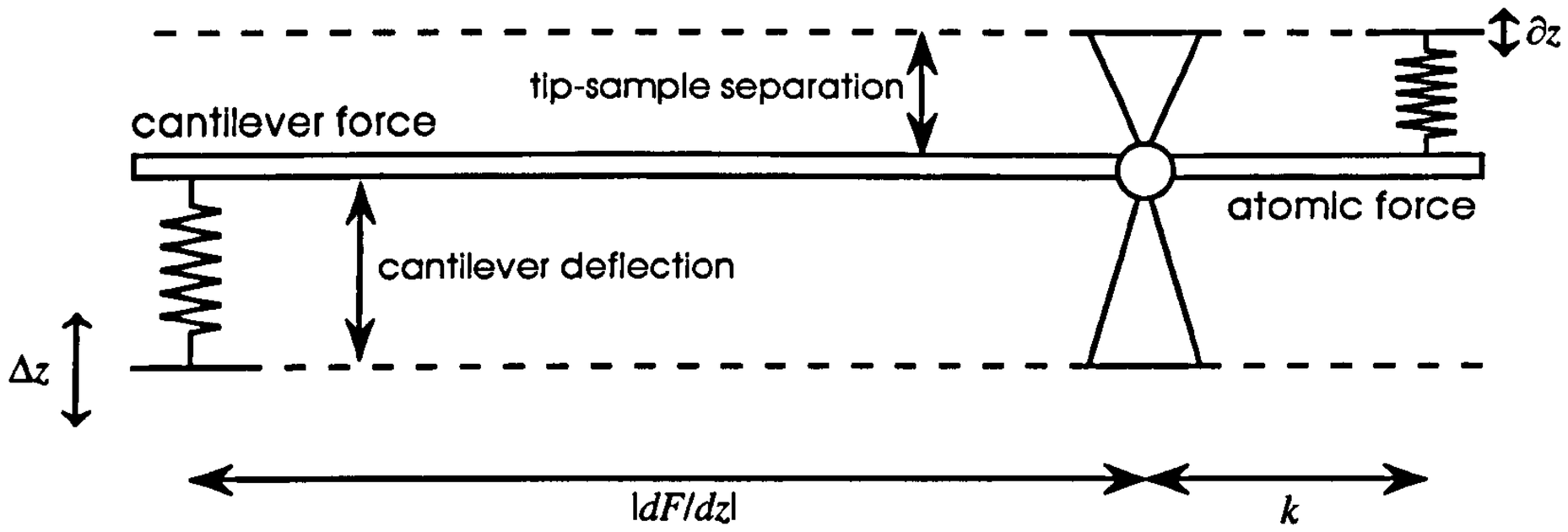


Figure 1.3.5: A force balance representation of the atomic force microscope. Both springs here have identical spring constants. Small changes in separation or atomic forces cause large changes in cantilever deflection if the cantilever stiffness is much smaller than the atomic force gradient.

1.3.2.2 Non-Contact Modes

In the non-contact modes, first demonstrated by Martin, Williams and Wickramasinghe[34] (1987), the situation depicted in Figure 1.3.4 is treated as a coupled spring system with the cantilever representing one spring (the test spring) and the atomic force or bond representing the other spring (the spring to be measured). Each spring has its own resonant frequency and energy dissipation mechanisms. The atomic spring can be viewed as a lossy spring, with losses being associated with deformation of the sample and due to the specific force mechanism involved. However, unlike an ideal spring, the compliance of the atomic spring is a function of position as can be seen from the compliance curve in Figure 1.3.2. Thus the effective spring constant and resonant frequency of the whole system is a function of the separation between the specimen and probe and by measuring this frequency, a measure of the separation can be obtained.

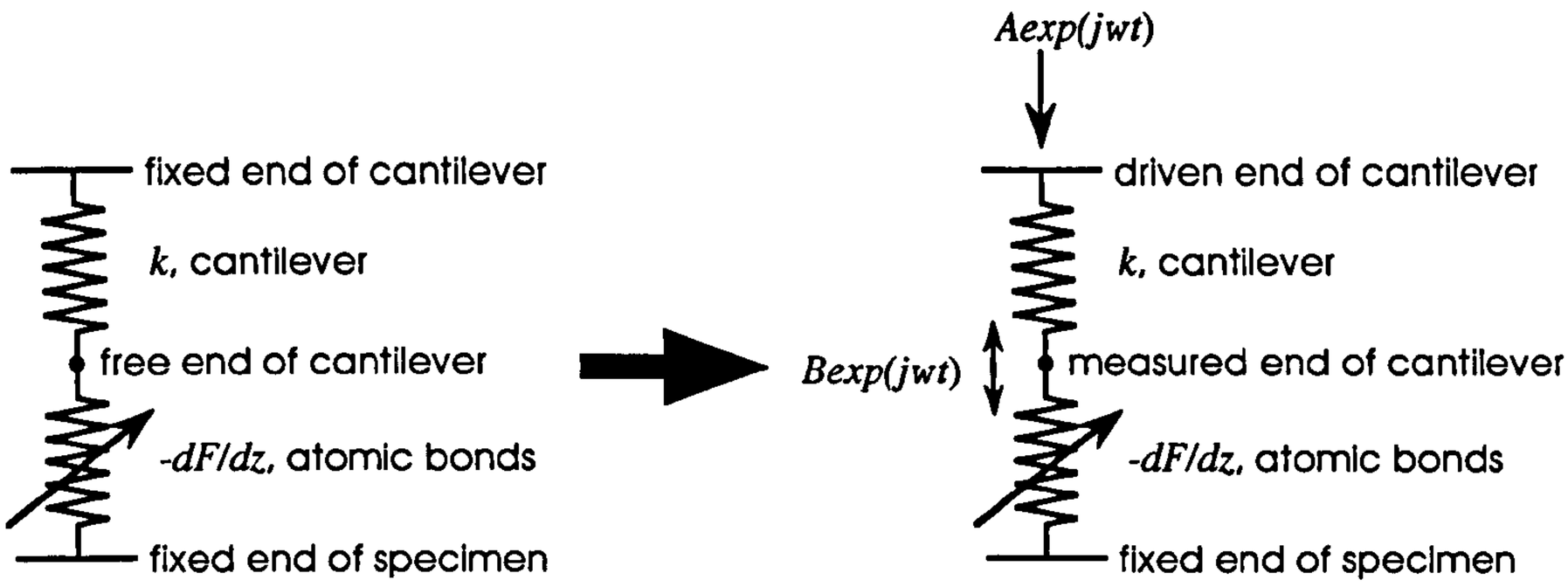


Figure 1.3.6: A schematic representation of the principle behind non-contact AFM in terms of mechanical springs.

The situation can be schematically represented as in Figure 1.3.6. Within the specimen the driving force upon the cantilever will lead to a group of atoms oscillating with respect to the rest of the specimen by varying amplitudes. However only the response of the free end of the cantilever can be easily measured. To simplify the analysis, the specimen is usually assumed to be a solid incompressible block in comparison with the atomic force interaction between the probe and the specimen at the resonant frequency of the free end of the cantilever. Then the analysis is no longer that of a system comprising two springs in series, but is closer to that of a system with two springs in parallel.

Consider the static equilibrium situation. If the probe is in the repulsive atomic force regime, both springs of Figure 1.3.6 can be considered to be compressed (by the force acting on the fixed end of the cantilever). If the probe is in the attractive atomic force regime, both springs can be considered to be extended. Quantitatively the analysis begins with Equation 1.3.6 from which the general force equation can be written as:

$$F_{atomic}(d-s) = -k(u-d) \equiv F_{atomic}(z_o) = -k(f-h-z_o-s), u = f-h, z_o = d-s, \qquad (1.3.7)$$

where k is the stiffness of the cantilever, u is the undeflected position of the probe, d is the deflected position of the probe, f is the position of the fixed end of the cantilever, h is the height of the probe, F_{atomic} is the atomic force between the sample and the probe, s

is the position of the interacting part of the sample and z_o is the distance between the sample and the deflected position of the probe. [Note that the compliance of the 'atomic spring' is given by the negative of the spatial derivative of the atomic force with respect to the z-axis, however, the atomic forces can act so as to increase deviations from the (metastable) equilibrium position and so this compliance can be *negative*, unlike with a classical spring.] In order to measure changes in compliance it is normal to measure the resonant frequency of the system. This can be done by using the piezoelectric driver to alter the position of the fixed end of the cantilever by a small amount, so that:

$$k\Delta f = k\Delta z - \Delta F_{atomic}(z_o) \approx k\Delta z - \left.\frac{dF_{atomic}}{dz}\right|_{z_o} \Delta z = \left[k + k_{atomic}(z_o)\right]\Delta z, \tag{1.3.8}$$

where the deviation of the atomic force has been expanded to the first order of the deviation in the sample-probe distance. With sinusoidal excitation of the fixed end of the cantilever by the piezoelectric driver of amplitude, a, and angular frequency, ω, and simply denoting the separation deviation by z, the equation of motion becomes:

$$ka\cos\omega t = m\frac{d^2z}{dt^2} + m\gamma\frac{dz}{dt} + \left[k + k_{atomic}(z_o)\right]z,\ k_{atomic}(z) = -\frac{dF_{atomic}}{dz}, \tag{1.3.9}$$

where m is the effective mass of the cantilever and probe and γ is the specific coefficient of velocity damping experienced by the probe (this includes losses within the cantilever and velocity damping effects caused by the force interactions being used). This is just the equation of a forced and damped simple harmonic oscillator. In terms of the amplitude response of the cantilever, Equations 1.3.8 and 1.3.9 are only valid for:

$$\left\|\tfrac{1}{2}\Delta k_{atomic}(z_o)\right\| << \left\|k_{atomic}(z_o)\right\| \Rightarrow \frac{\left\|k_{atomic}(z_o)\right\|}{\left\|\left.\frac{dk_{atomic}}{dz}\right|_{z_o}\right\|} >> \frac{\Delta z_o}{2}. \tag{1.3.10}$$

To proceed further, the UHV Van der Waal's force system of §1.3.1 will be used as an example – the analysis for other force mechanisms would proceed in a similar manner. Then from Equation 1.3.5 the compliance of the atomic force on the probe is given by:

$$k_{atomic}(r) = -\frac{4}{3}\pi^2 W_o n_1 n_2 \sigma^3 R\left[\left(\frac{\sigma}{r}\right)^3 - \frac{2}{15}\left(\frac{\sigma}{r}\right)^9\right] \tag{1.3.11}$$

and the compliance gradient is given by:

$$-\frac{dk_{atomic}(r)}{dr} = -4\pi^2 W_o n_1 n_2 \sigma^2 R\left[\left(\frac{\sigma}{r}\right)^4 - \frac{2}{5}\left(\frac{\sigma}{r}\right)^{10}\right], \tag{1.3.12}$$

where all parameters are as in Equation 1.3.4. From this, the ratio of compliance to compliance gradient is given by:

$$-\frac{k_{atomic}(r)}{\frac{dk_{atomic}(r)}{dr}} = \frac{1}{3}\sigma\frac{\left[1 - \frac{2}{15}\left(\frac{\sigma}{r}\right)^6\right]}{\left[\frac{\sigma}{r} - \frac{2}{5}\left(\frac{\sigma}{r}\right)^7\right]} \approx \frac{r}{3} \text{ for } r >> \sigma. \tag{1.3.13}$$

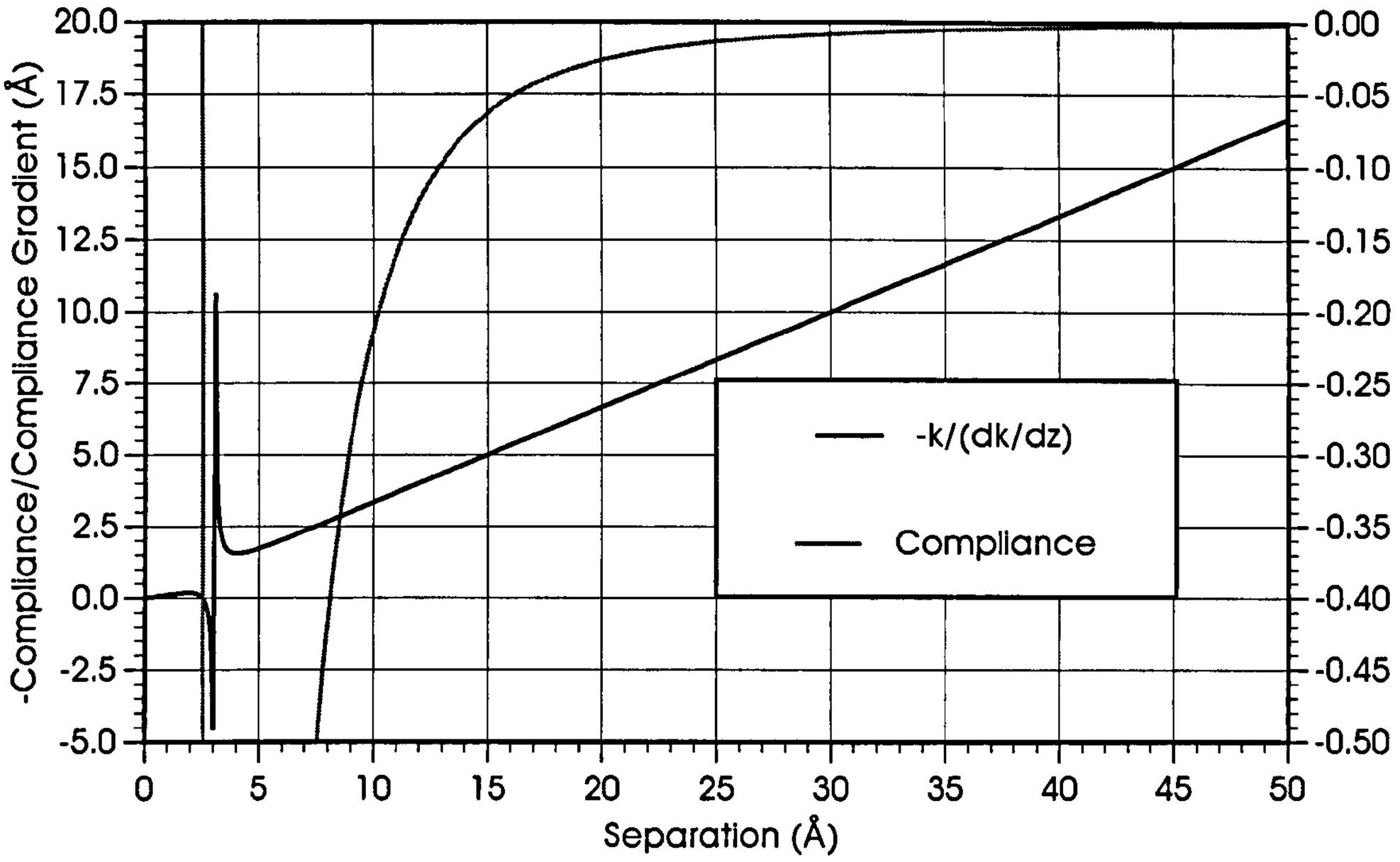


Figure 1.3.7: Typical compliance-to-compliance-gradient ratio and compliance curves for non-contact in UHV. The graphs are for a minimum energy of -10meV (Van der Waal's force) and an interatomic spacing in both the probe and specimen of 4Å, The probe is modelled as a sphere with a radius of curvature of 20nm while the specimen is modelled as an infinite plane. The error in the approximation of Equation 1.3.13 is down to 10% by 4.5Å.

Figure 1.3.7 plots the ratio of the compliance-to-compliance-gradient for the UHV Van der Waal's force system described by Figure 1.3.2. [Note that the behaviour at a separation of around 3Å is due to the second order term (compliance gradient) going to zero. This is a mathematical artefact and not a physical phenomenon. The correct behaviour for this region would be obtained by considering higher order derivatives.] If the condition expressed by Equation 1.3.10 is arbitrarily taken to mean that the second order term is 20% of the first order term and the total noise in the system is 0.5Å (typically limited by thermal noise in the cantilever), so that the minimum usable oscillation amplitude is 1.0Å, *i.e.* Δz_o is 2.0Å, (corresponding to a signal to noise ratio of two), then it is clear from Figure 1.3.7 that the condition can only be satisfied for $r >> \sigma$. In terms of the oscillation amplitude $2A = \Delta z_o$, and applying Equation 1.3.13, the operating separation condition becomes:

$$z_o \geq 7.5A \Leftrightarrow A \leq \frac{z_o}{7.5}, \tag{1.3.14}$$

so that the minimum operating separation is 7.5Å which is around 2σ, for the above values. This can be compared with contact mode AFM which can be viewed as AFM at an operating 'separation' of around σ.

From elementary mechanics the natural undamped angular frequency of the cantilever, ω_o, and the quality factor of its oscillations, Q, are given by the relations:

$$\omega_o = \sqrt{\frac{k}{m}},\ Q = \frac{2\pi \times \text{energy stored}}{\text{energy lost per cycle}} = \frac{\omega_o}{\gamma}, \tag{1.3.15}$$

free cantilever frequency	symbol	formula/value
natural undamped; velocity, phase and power resonance	ω_o	$= \sqrt{\frac{k}{m}}$
natural damped	ω_n	$= \frac{\omega_o}{Q}\sqrt{Q^2 - 1/4}, \approx \omega_o$ for $Q >> 1$
amplitude resonance	ω_r	$= \frac{\omega_o}{Q}\sqrt{Q^2 - 1/2}, \approx \omega_o$ for $Q >> 1$
acceleration resonance	ω_a	$= \frac{Q\omega_o}{\sqrt{Q^2 - 1/2}}, \approx \omega_o$ for $Q >> 1$

Table 1.3.1: The exact values of important free cantilever frequencies and useful approximations for cantilevers exhibiting a large quality factor.

the stiffness of the coupled system of Equation 1.3.9 is given by:

$$k'(z) = k + k_{atomic}(z) = k - \frac{dF_{atomic}(z)}{dz} \tag{1.3.16}$$

and the resulting natural undamped angular frequency of the system is given by:

$$\omega_o'(z) = \sqrt{\frac{k'(z)}{m}} = \omega_o\sqrt{1 + \frac{k_{atomic}(z)}{k}}\,. \tag{1.3.17}$$

Other important frequencies of the free and coupled cantilever are given in Tables 1.3.1 and 1.3.2. From Figure 1.3.7 it is clear that for separations of 2σ or more, the atomic force gradient is always *negative* and hence the resonant frequency of the coupled cantilever system is always *below* that of the free cantilever. In addition, as can be seen from Equation 1.3.17, the compliance of the cantilever must *exceed* the atomic force gradient for oscillations to take place. Physically, this corresponds to having a cantilever capable of resisting the attractive atomic force. Ignoring any change in the velocity damping coefficient from that in the uncoupled system for simplicity (valid for UHV Van der Waal's AFM), the effective quality factor of the system is given by:

$$Q'(z) = \frac{\omega_o'(z)}{\gamma} = Q\frac{\omega_o'(z)}{\omega_o} = Q\sqrt{1 + \frac{k_{atomic}(z)}{k}}\,, \tag{1.3.18}$$

and since the atomic force gradient is negative in the operating regime, this is *always* less than the quality factor of the free cantilever. The normal condition for a system to exhibit intrinsic oscillatory behaviour is that the effective quality factor exceed 0.5 (*i.e.* an underdamped system). If the compliance of the cantilever is at least double the atomic force gradient, this corresponds to the quality factor of the free cantilever exceeding $1/\sqrt{2}$.

The general solution of Equation 1.3.9 can be calculated from elementary mechanics:

$$z = A\cos(\omega t - \delta) + B\exp\left[-\omega_o t/2Q\right]\cos(\omega_n' t + \theta), \tag{1.3.19}$$

coupled cantilever frequency	symbol	formula/value
natural undamped; velocity, phase and power resonance	ω_o'	$= \sqrt{\frac{k'}{m}} = \omega_o \sqrt{1 + \frac{k_{atomic}}{k}}$, $\approx \omega_o \left[1 + \frac{k_{atomic}}{2k}\right]$ for $Q >> 1,\ k_{atomic} << k$
natural damped	ω_n'	$= \omega_n \sqrt{1 + \left(\frac{\omega_o}{\omega_n}\right)^2 \frac{k_{atomic}}{k}} = \omega_o \sqrt{1 + \frac{k_{atomic}}{k} - \frac{1}{4Q^2}}$, $\approx \omega_o \left[1 + \frac{k_{atomic}}{2k}\right]$ for $Q >> 1,\ k_{atomic} << k$
amplitude resonance	ω_r'	$= \omega_r \sqrt{1 + \left(\frac{\omega_o}{\omega_r}\right)^2 \frac{k_{atomic}}{k}} = \omega_o \sqrt{1 + \frac{k_{atomic}}{k} - \frac{1}{2Q^2}}$, $\approx \omega_o \left[1 + \frac{k_{atomic}}{2k}\right]$ for $Q >> 1,\ k_{atomic} << k$
acceleration resonance	ω_a'	$= \omega_a \frac{1 + \frac{k_{atomic}}{k}}{\sqrt{1 + \left(\frac{\omega_a}{\omega_o}\right)^2 \frac{k_{atomic}}{k}}} = \omega_o \frac{1 + \frac{k_{atomic}}{k}}{\sqrt{1 + \frac{k_{atomic}}{k} - \frac{1}{2Q^2}}}$, $\approx \omega_o \left[1 + \frac{k_{atomic}}{2k}\right]$ for $Q >> 1,\ k_{atomic} << k$

Table 1.3.2: The exact values of important coupled cantilever frequencies and useful approximations for cantilevers exhibiting a large quality factor in the presence of force gradients much smaller than the compliance of the cantilever.

where the first term is the response of the cantilever to the forced oscillations at an angular frequency, ω, and the second term is the transient response of the cantilever at the natural damped vibration frequency (as shown in Table 1.3.2). The amplitude, B, and phase, θ, of the transient response are the two integration constants necessary for a complete solution to Equation 1.3.9 and are determined from initial conditions.

The amplitude of the cantilever response to the forced oscillations, A, is given by:

$$A(\omega, \omega_o') = \frac{a\omega_o^{\,2}}{\left[\left(\omega_o'^{2} - \omega^2\right)^2 + \frac{\omega_o^{\,2}\omega^2}{Q^2}\right]^{1/2}} = \frac{a\omega_o^{\,2}}{\left[\left[\omega_o^{\,2}\left[1 + \frac{k_{atomic}}{k}\right] - \omega^2\right]^2 + \frac{\omega_o^{\,2}\omega^2}{Q^2}\right]^{1/2}}, \qquad (1.3.20)$$

where a is the amplitude of the forced oscillations driven by a piezoelectric actuator, with all angular frequencies as indicated in Tables 1.3.1 and 1.3.2.

The phase lag of the cantilever response, δ, behind that of the driving oscillations is implicitly given by the following relations:

$$\tan\delta = \frac{\omega_o\omega}{Q\left(\omega_o'^2 - \omega^2\right)} = \frac{1}{Q\left[\frac{\omega_o'^2}{\omega_o\omega} - \frac{\omega}{\omega_o}\right]} = \frac{1}{Q\left[\frac{\omega_o}{\omega}\left[1 + \frac{k_{atomic}}{k}\right] - \frac{\omega}{\omega_o}\right]}$$

$$\sin\delta = \frac{\omega_o\omega}{Q\sqrt{\Omega}} = \frac{\omega}{Q\omega_o}\frac{A\left(\omega,\omega_o'\right)}{a}, \cos\delta = \frac{\omega_o'^2 - \omega^2}{\sqrt{\Omega}} = \frac{\omega_o'^2 - \omega^2}{\omega_o^{\,2}}\frac{A\left(\omega,\omega_o'\right)}{a}. \quad (1.3.21)$$

$$\Omega = \left(\omega_o'^2 - \omega^2\right)^2 + \frac{\omega_o^{\,2}\omega^2}{Q^2}$$

From Equation 1.3.20 the gain (or transfer function) of the amplitude scheme is given by:

$$G\left(\omega,\omega_o'\right) = \frac{\omega_o^{\,2}}{\left[\left(\omega_o'^2 - \omega^2\right)^2 + \frac{\omega_o^{\,2}\omega^2}{Q^2}\right]^{1/2}} = \frac{\omega_o^{\,2}}{\left[\left[\omega_o^{\,2}\left[1 + \frac{k_{atomic}}{k}\right] - \omega^2\right]^2 + \frac{\omega_o^{\,2}\omega^2}{Q^2}\right]^{1/2}}. \quad (1.3.22)$$

There are then three possible feedback systems based on the change in important coupled system frequencies with sample-probe separation: those based on maintaining a constant oscillation amplitude at a fixed frequency; those based on maintaining a constant phase lag at a fixed frequency; and those based on maintaining a constant resonant frequency – all by altering the sample-probe separation. As with STM and contact mode AFM, each of these systems can also be run *open loop* (in a constant height mode) for faster scanning, where either the amplitude, phase lag or resonant frequency are measured as a function of lateral position.

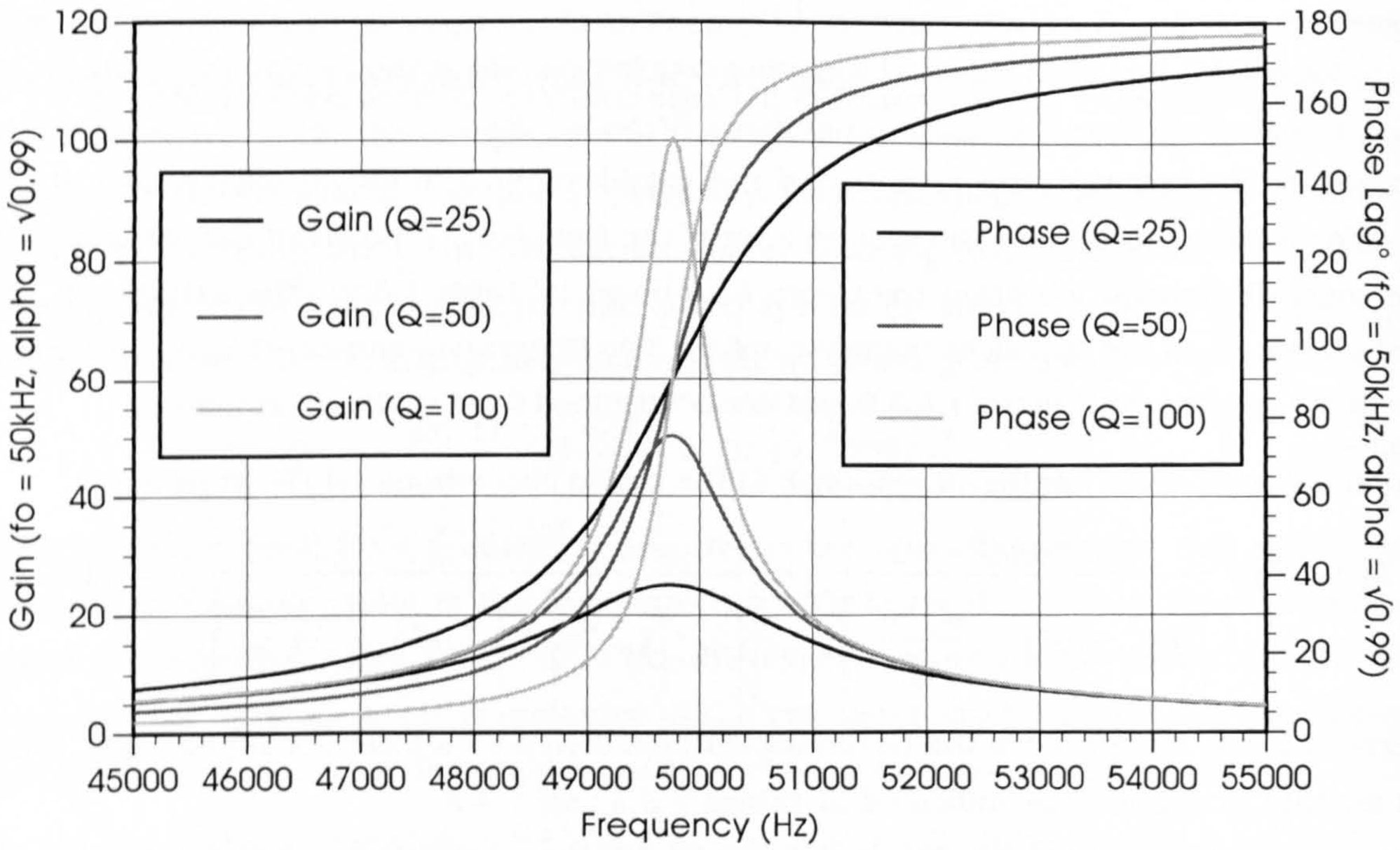


Figure 1.3.8: The variation in the cantilever response with frequency for a range of quality factors. Here the natural undamped resonant frequency is 50kHz and the frequency reduction factor is √0.99.

Taking into account the fact that the atomic force gradient is always negative, the condition that must hold over the range of operation is:

$$0 \geq \frac{k_{atomic}(z)}{k} \geq -\frac{1}{2} \Rightarrow \omega_o \geq \omega_o' \geq \frac{\omega_o}{\sqrt{2}}. \tag{1.3.23}$$

Thus in order to be sensitive to the probe-specimen separation, the detection scheme must be able to distinguish between changes in amplitude, phase lag or frequency over this range. If it is assumed that the operating undamped resonant frequency of the coupled system at the operating separation lies within the limits of Equation 1.3.23, then we can write:

$$\omega_o'(z_o) = \alpha\omega_o, \tag{1.3.24}$$

where α is a frequency reduction factor which lies within the range:

$$1 > \alpha \geq \frac{1}{\sqrt{2}} \Rightarrow 1 > \alpha^2 \geq \frac{1}{2}. \tag{1.3.25}$$

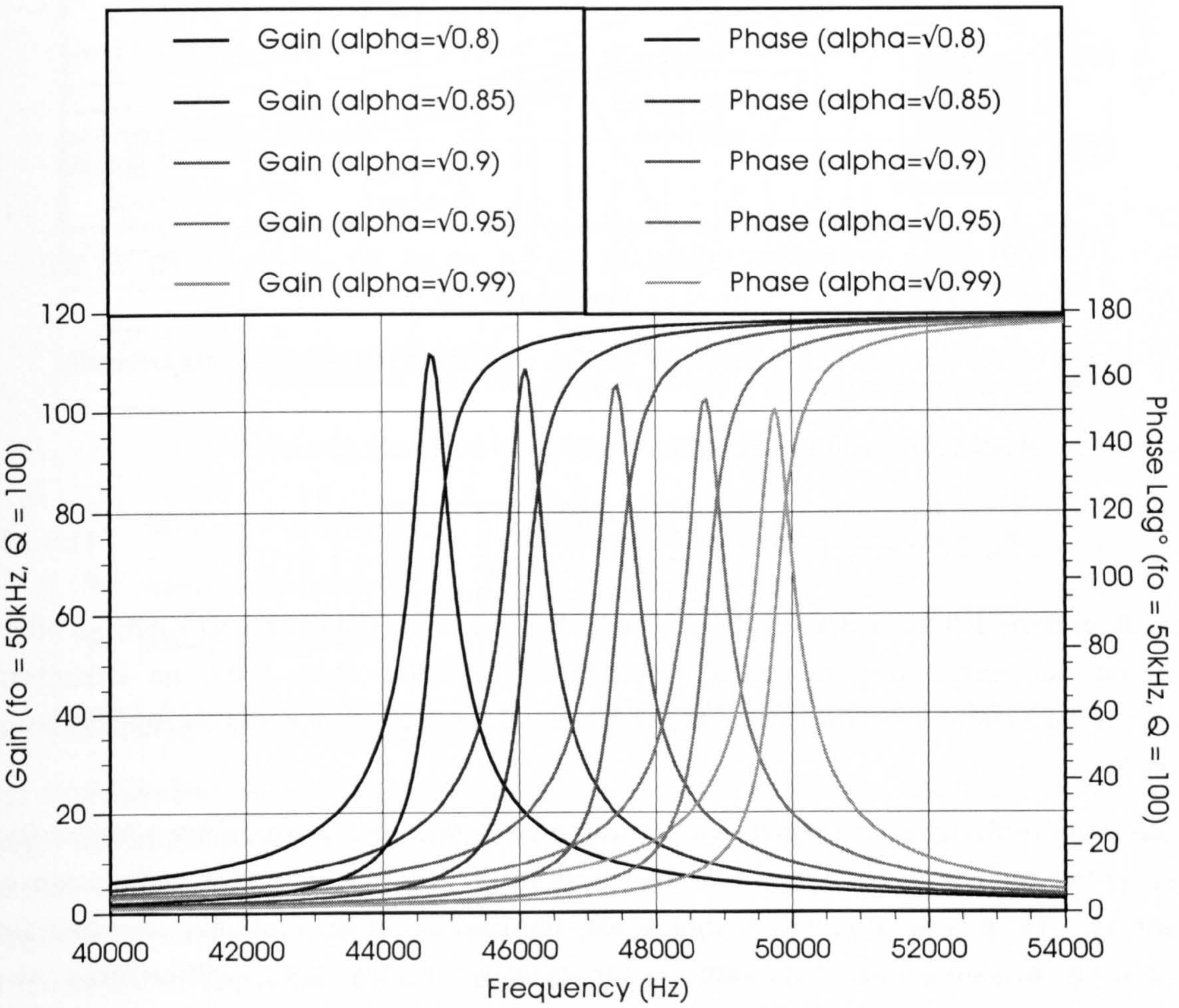


Figure 1.3.9: The variation in the cantilever response with frequency for a range of frequency reduction factors. Here the natural undamped resonant frequency is 50kHz and the quality factor is 100.

Figure 1.3.8 plots the amplitude and phase response of a cantilever with a natural undamped resonant frequency of 50kHz and a frequency reduction factor of √0.99 for various values of the quality factor. As expected, the gain increases with increasing

quality factor. Figure 1.3.9 plots the amplitude and phase response of a cantilever with a natural undamped resonant frequency of 50kHz and a quality factor of 100 for various values of the frequency reduction factor. From this it is clear that the amplitude increases and the resonant frequency decreases as the frequency reduction factor decreases.

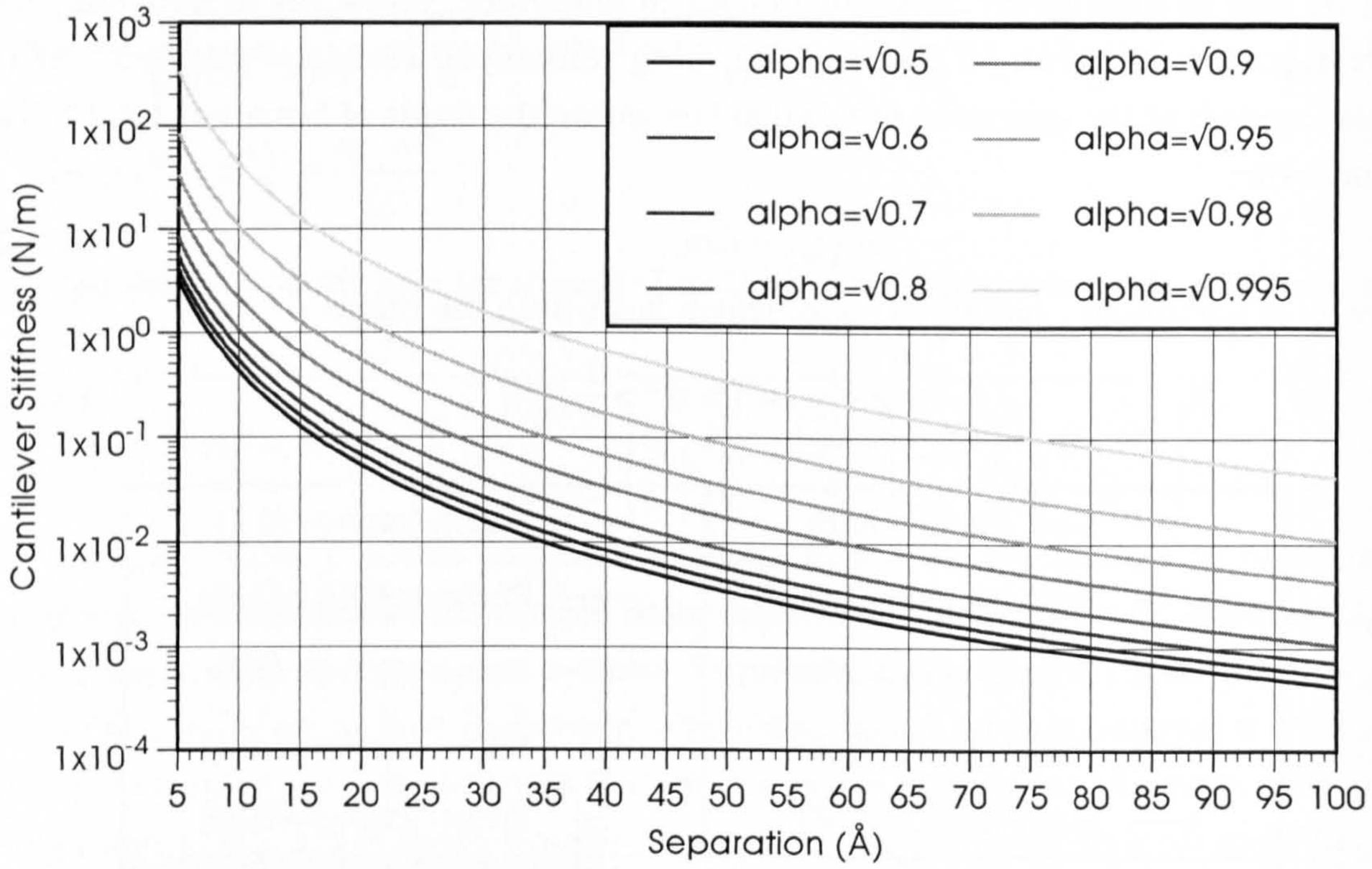


Figure 1.3.10: Variation of the cantilever stiffness with operating separation for a range of frequency reduction factors in Van der Waal's UHV AFM.

For Van der Waal's UHV AFM the cantilever stiffness is simply given by:

$$k_{z_o} = -\frac{k_{atomic}(z_o)}{1-\alpha^2} = \frac{4\pi^2 W_o n_1 n_2 \sigma^3 R}{3\left(1-\alpha^2\right)}\left[\left(\frac{\sigma}{z_o}\right)^3 - \frac{2}{15}\left(\frac{\sigma}{z_o}\right)^9\right] \tag{1.3.26}$$

from Equations 1.3.24, 1.3.17 and 1.3.11 which shows that the cantilever stiffness must increase for decreasing operating separations (greater than 2σ), as expected. Figure 1.3.10 shows how the choice of cantilever stiffness then depends on the operating separation.

In the amplitude detection scheme, the cantilever is vibrated at some fixed frequency, ω_A, which results in an amplitude response, $A\left(\omega_A, \omega_o'(z)\right)$, which is a function of separation. By maintaining a fixed amplitude, the position of the cantilever can be used to obtain an image of the sample topography. Figure 1.3.11 illustrates one implementation of this scheme.

The phase detection scheme is very similar and the set-up of Figure 1.3.11 only differs in that the phase output of the lock-in amplifier is utilised in place of the amplitude output. In the phase detection scheme, the cantilever is vibrated at some fixed frequency, ω_p, which results in a phase lag, δ, which is a function of separation. By maintaining a fixed

phase lag, the position of the cantilever can be used to obtain an image of the sample topography. The tangent of the phase lag can be rewritten from Equation 1.3.21 as:

$$\tan\delta = \frac{\dfrac{\omega}{\omega_o'}}{Q\dfrac{\omega_o'}{\omega_o}\left[1-\left(\dfrac{\omega}{\omega_o'}\right)^2\right]} = \frac{\dfrac{\omega}{\omega_o'}}{Q\alpha\left[1-\left(\dfrac{\omega}{\omega_o'}\right)^2\right]}, \tag{1.3.27}$$

from which it is clear that the phase lag is 90° when the oscillation frequency is equal to the undamped resonant frequency of the coupled system. Thus the most sensitive indicator of the phase lag is its cosine, which has its maximal rate of change at 90°.

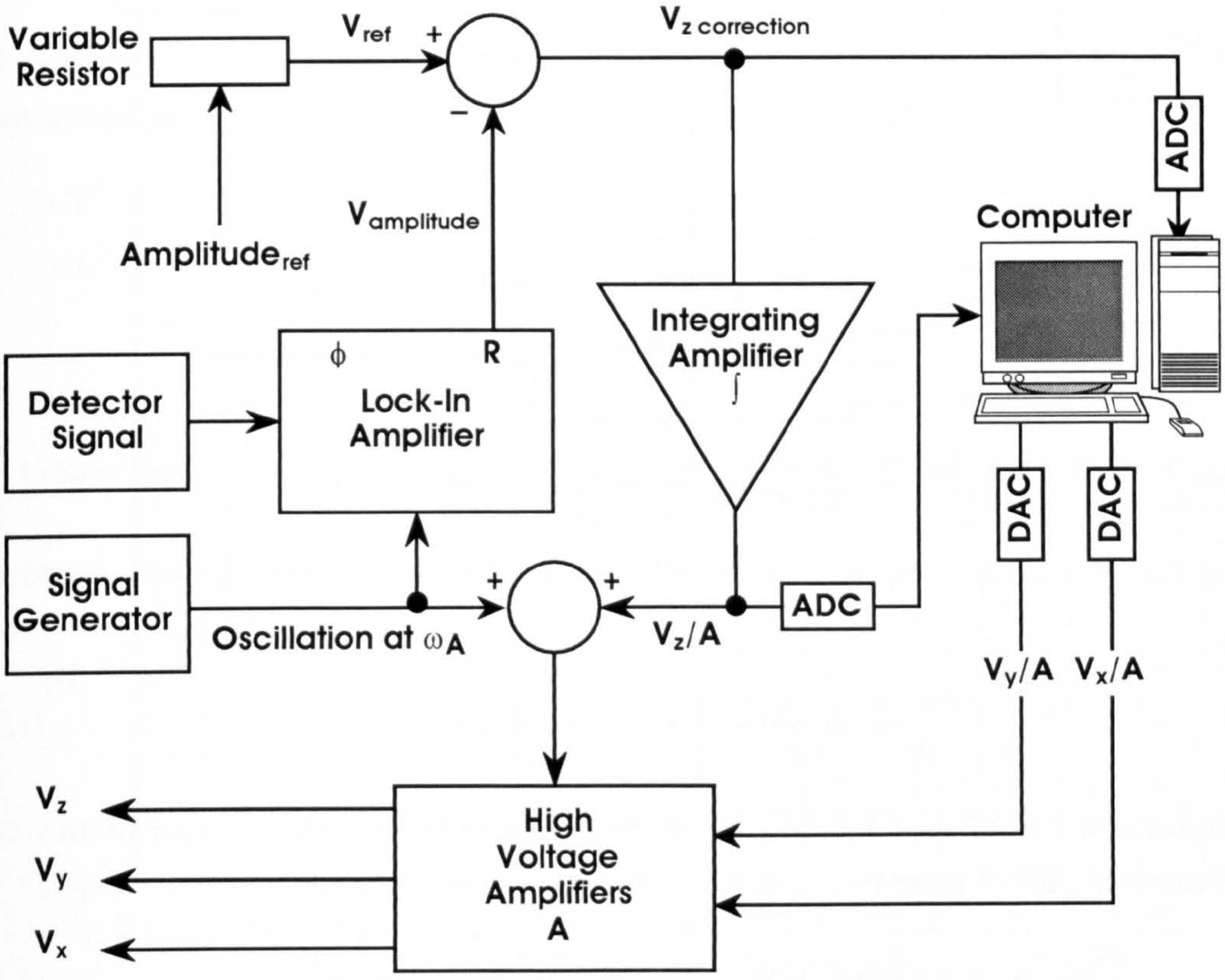


Figure 1.3.11: Implementation of the amplitude regulation scheme for non-contact mode AFM. The phase regulation scheme is very similar.

The final scheme is the frequency detection scheme [Albrecht, Grutter, Horne and Rugar[35] (1991)]. Here the cantilever becomes the controlling oscillator of a resonant system, *i.e.* the oscillation circuit follows the cantilever's resonant frequency. In the frequency detection scheme, the cantilever is vibrated at its resonant frequency, ω_o', which is a function of separation. This frequency is measured with reference to the set-point frequency, ω_f, and the cantilever position altered to maintain this resonant frequency. Figure 1.3.12 is an example of one such implementation.

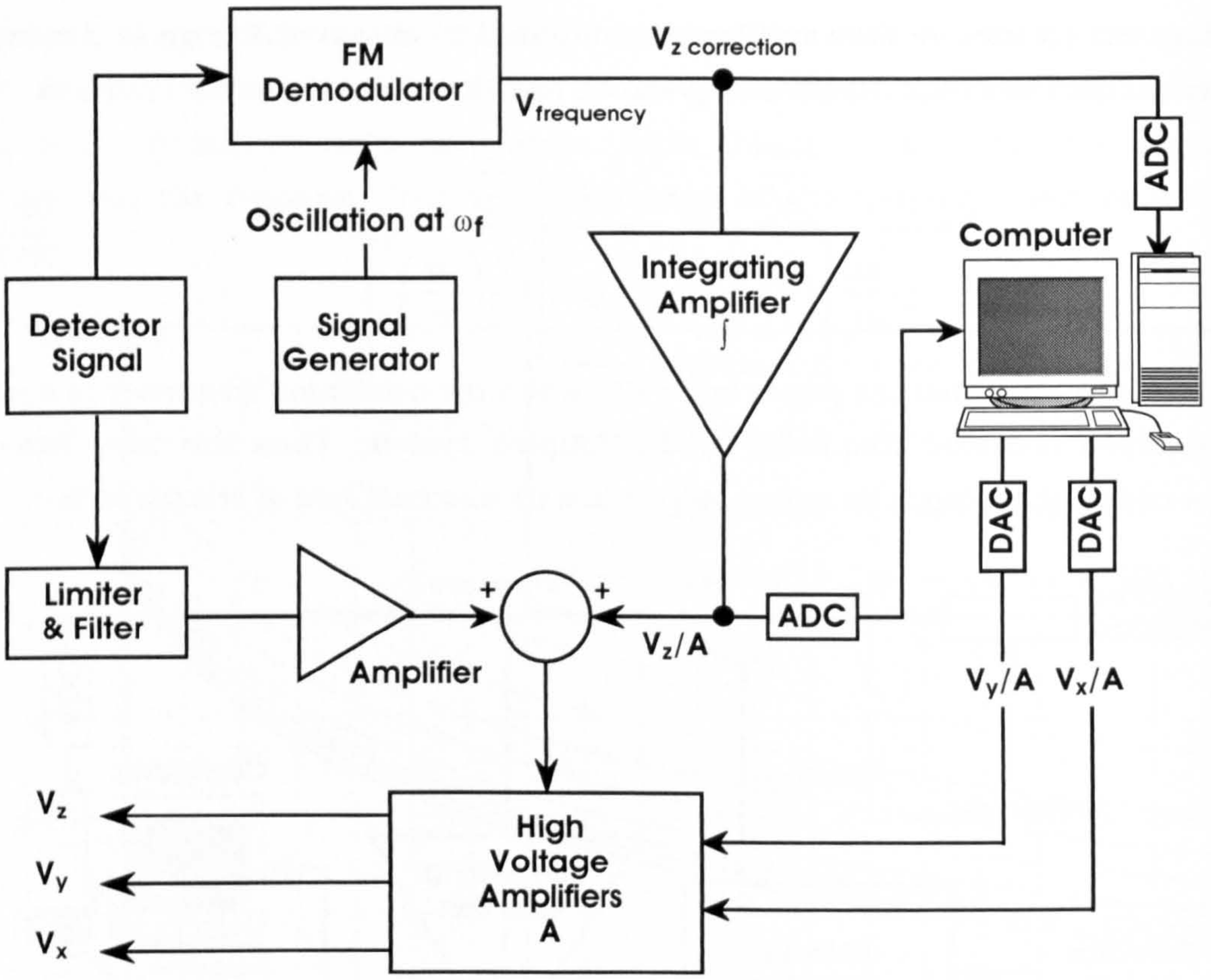


Figure 1.3.12: Implementation of the frequency regulation scheme for non-contact mode AFM.

All of the above schemes rely on the variation of the resonant frequency with separation, *i.e.*:

$$\left.\frac{d\omega_o'}{dz}\right|_{z_o} = \left.\frac{d\omega_o'}{dk_{atomic}}\right|_{k_{atomic}(z_o)} \left.\frac{dk_{atomic}}{dz}\right|_{z_o}. \tag{1.3.28}$$

From Equations 1.3.17 and 1.3.24, the rate of change in the resonant frequency of the coupled system with change in the atomic force gradient is given by:

$$\frac{d\omega_o'}{dk_{atomic}} = \frac{\omega_o}{2k\sqrt{1+\frac{k_{atomic}}{k}}} = \frac{1}{2m\omega_o'} \Rightarrow \left.\frac{d\omega_o'}{dk_{atomic}}\right|_{\omega_o'(z_o)} = \frac{\omega_o}{2\alpha k_{z_o}}, \tag{1.3.29}$$

where m is the effective mass of the cantilever. Combining this with Equations 1.3.12, 1.3.26 and 1.3.28 yields:

$$\left.\frac{d\omega_o'}{dz}\right|_{z_o} = \frac{3\omega_o\left(1-\alpha^2\right)}{2\alpha z_o}\frac{\left[1-\frac{2}{5}\left(\frac{\sigma}{z_o}\right)^6\right]}{\left[1-\frac{2}{15}\left(\frac{\sigma}{z_o}\right)^6\right]}, \tag{1.3.30}$$

for Van der Waal's AFM in UHV.

Equation 1.3.30 shows that the smaller the operating separation and frequency reduction factor, the more sensitive the system becomes. From this:

$$\left.\frac{\Delta\omega_o'}{\omega_o'}\right|_{z_o} \approx \frac{1}{\omega_o'}\left.\frac{d\omega_o'}{dz}\right|_{z_o}\Delta z = S_f(\alpha, z_o)\frac{\Delta z}{z_o}, \tag{1.3.31}$$

where the frequency deviation coefficient, $S_f(\alpha, z_o)$, has been defined as:

$$S_f(\alpha, z_o) = \frac{3(1-\alpha^2)}{2\alpha^2}\frac{\left[1-\frac{2}{5}\left(\frac{\sigma}{z_o}\right)^6\right]}{\left[1-\frac{2}{15}\left(\frac{\sigma}{z_o}\right)^6\right]} \approx \frac{3(1-\alpha^2)}{2\alpha^2}, z_o >> \sigma. \tag{1.3.32}$$

Figure 1.3.13 plots this as a function of the frequency reduction factor in the limit of large separations for Van der Waal's AFM in UHV.

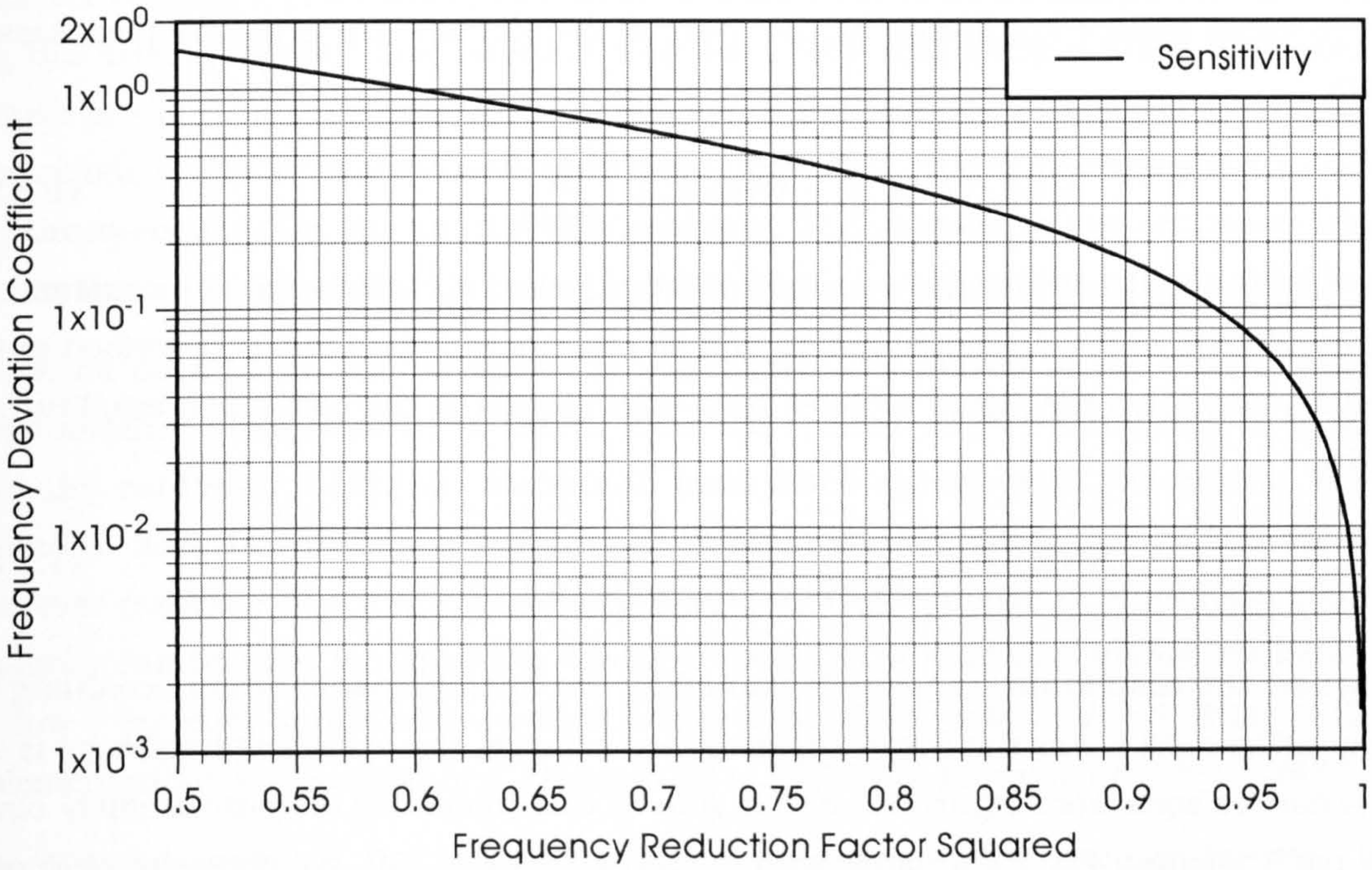


Figure 1.3.13: Variation of the frequency deviation coefficient with frequency reduction factor for non-contact mode Van der Waal's AFM in the limit of large operating separations in UHV.

It is clear from the preceding analysis that the frequency deviation factor is enhanced by utilising the steep variation of amplitude or phase with resonant frequency in the amplitude and phase detection schemes respectively. This comes at the expense of longer settling times and hence a slower scan speed. The amplitude and phase schemes are slower because a sudden change in resonant frequency requires the build up of energy at the driving frequency to close to its final value before it can be compared to the set-point amplitude or an accurate measure of the phase lag can be obtained. As can be seen from Figure 1.3.8 the sensitivity to any frequency deviation is greater for higher quality factors. However, it can be seen from Equation 1.3.19 that the amplitude (and phase) will take a time of the order of Q/ω_o to settle (around Q cycles).

It should also be noted that as the quality factor increases, the amplitude detection scheme is susceptible to the signal polarity changing for a sudden change in separation. This problem can be avoided by operating at the optimum upper frequency with a frequency reduction factor close enough to unity so that a polarity switch would require the coupled resonant frequency to shift beyond the free natural resonant frequency. The phase detection scheme is unipolar and so is not as susceptible to this problem. The frequency detection scheme is the fastest responding scheme, but the detection signal does not benefit as much from the high gain achievable using a cantilever with a high quality factor.

The analysis above suggests that non-contact mode AFM can be optimised by simply considering the quality factor, operating separation and cantilever stiffness. However, in practice the thermal vibrations in the cantilever can seriously limit the resolution. The limitations due to the thermal vibrations of the cantilever can be assessed by noting that the root-mean-square amplitude of the thermal oscillations are given by [see Sarid[36] (1991)]:

$$\sqrt{\left\langle A_{th}^{\,2}\left(\omega,\omega_o'\right)\right\rangle} = G\left(\omega,\omega_o'\right)\sqrt{\frac{4k_BTB}{Qk\omega_o}}, \tag{1.3.33}$$

where $\left\langle A_{th}^{\,2}\left(\omega,\omega_o'\right)\right\rangle$ is the mean-square amplitude noise, k_B is Boltzmann's constant, B is the bandwidth of the detection system and all other parameters are as described in the preceding text. Then the signal to noise ratio with respect to the oscillation amplitude is given by:

$$\mathrm{SNR}_A = a\sqrt{\frac{Qk\omega_o}{4k_BTB}} = \frac{A\left(\omega,\omega_o'\right)}{G\left(\omega,\omega_o'\right)}\sqrt{\frac{Qk\omega_o}{4k_BTB}}. \tag{1.3.34}$$

In practice, the amplitude response is limited by the operating separation according to, *e.g.* Equation 1.3.14. For the frequency detection scheme, the operating frequency is ω_o' and as can be seen from Figure 1.3.8, the gain at this frequency is approximately equal to the quality factor. For a cantilever with a quality factor of 100, a stiffness of 10N/m, a free resonant frequency of 50kHz measured in a bandwidth of 1kHz, with an operating separation of 15Å and an oscillation amplitude of 2Å, at 300K, the signal to noise ratio is about 3.5. This can be improved by operating at a lower temperature, which will also result in an increase in the quality factor. The amplitude and phase detection schemes benefit from increasing quality factors due to an increased rate of change of amplitude and phase with driving frequency near resonance. The frequency detection scheme does not benefit from the increased quality factor to the same degree.

In all the regulation schemes a very high quality factor can be a problem because the oscillation amplitude of the piezoelectric drivers must decrease for a fixed oscillation amplitude response. For the cantilever above the piezoelectric driver will be oscillating with an amplitude of only 0.02Å corresponding to an applied voltage of about 6mV peak-peak (based on a piezoelectric response of 6Å/V). Since the driving voltage may also

control the position of the cantilever (which requires 100V or more), the noise in the driving circuit would have to be better than 30ppm. For this reason, a different piezoelectric driver with a more limited driving circuit voltage range is often used to oscillate the cantilever with a separate piezoelectric driver controlling the sample position (or vice versa).

1.3.2.3 Intermittent-Contact Modes

The intermittent contact mode is a hybrid of the contact and non-contact modes, also sometimes known as 'tapping mode'. In this mode the cantilever is driven at a *fixed* frequency close to its free resonant frequency. At each scan point the cantilever is forced to approach the sample until the oscillation amplitude has decreased to a set-point level, which could be close to zero. This occurs because as the separation decreases the resonant frequency falls. [Alternatively, the phase lag could be monitored till the cantilever reaches a point where its oscillations are close to being completely out of phase with the driving signal.] The resonant frequency will drop to zero when the cantilever stiffness is equal in magnitude to the atomic force gradient (from Equation 1.3.17) as this corresponds to the attractive forces overcoming the cantilever and bringing it into contact with the specimen (*i.e.* the equilibrium position). The position of the cantilever when it reaches the set-point amplitude is then used to form an image of the sample topography.

After each data point, the cantilever is withdrawn and moved to the next data point, where another measurement is taken. [One convenient way of doing this is to simply drive the cantilever using an amplitude modulated signal, with the carrier being a frequency close to resonance so that the modulation results in an oscillation of the cantilever response at a lower frequency. The resulting response in amplitude, phase or frequency can be used to adjust the cantilever position so that the average response over the low frequency modulation period remains at a set-point value.] Many different implementations are possible [*e.g.* Hansma, *et al.*[37] (1994); Putman, *et al.*[38] (1994)].

This method protects the cantilever against damage from specimens with large topography (as the cantilever is withdrawn to a safe distance before moving to the next data point), but allows a resolution comparable to that in contact mode to be achieved. Its other advantage is that since the cantilever is not moved in as close proximity to the sample between data points, the tip and specimen are less likely to be damaged during scanning.

1.3.3 Interpretation of AFM Images and Resolution

It is immediately clear by comparing Equations 1.2.1, 1.3.5 and 1.3.31 that the resolution of STM is much higher than that of AFM. In addition the resolution of contact mode AFM is higher than that of non-contact AFM. This is purely because the response of each technique to changes in topography depends on the strength of the dependence of the interaction on separation. STM has an exponential dependence, contact mode Van

der Waal's AFM (in the repulsive regime) depends inversely on the seventh power of the separation while the non-contact mode equivalent depends inversely on the first power of separation (or marginally higher using the amplitude or phase regulation schemes). However, all these techniques are potentially capable of atomic resolution (albeit with non-contact mode AFM this may only be possible on samples exhibiting a large atomic corrugation). As with STM, a high vertical resolution results in a high lateral resolution (for the same reasons).

However, whilst the atomic resolution abilities of the STM were never in doubt (the standard test is to obtain images showing single atom point-defects), the same was not true of contact mode AFM. As with STM, it is possible to obtain data as a function of separation with the AFM. These force-distance curves tend to exhibit hysteresis due to the presence of two energy minima separated by a small energy barrier (the magnitude and existence of which depends on cantilever and specimen – see Sarid[32] (1991)).

However, the observed hysteresis is much larger than this and is primarily due to the presence of a thin film of water on the surface of specimens in air. This introduces attractive capillary forces which can range from 1nN to 10nN depending on ambient humidity, cantilever geometry and specimen. The hysteresis is caused by the need for a certain amount of repulsive force to overcome the surface tension as the cantilever penetrates the film, and a certain amount of attractive force to then release the tip from the film. This means that in order to operate in the (net) repulsive force regime, the repulsive forces have to exceed the attractive capillary forces. As these forces are relativity large, the repulsive forces are large enough to cause the end of the AFM probe to be distorted so that any protruding atoms are pushed into the tip. The tip is then flattened to the point where single atom defects cannot be observed (compare this with multiple tip averaging in STM).

Thus for many years images showing anything but perfectly ordered atomic surfaces (free from any defects) were rare. Once the origin of this problem was understood [Ohnesorge and Binnig[39] (1993)], it was shown that in order to obtain true atomic resolution with contact mode AFM, the system had to be operated with the cantilever within the film on the specimen's surface but with an attractive force being applied so as to neutralise the effect of the capillary forces (*i.e.* the cantilever had to be actively pulled away from the surface with a force just insufficient to pull free of the water film on the surface). This is because for true atomic imaging an operating force of 0.01nN to 0.1nN is required in order that neither the specimen or tip are distorted.

The other alternative is to operate the AFM under liquid, where Van der Waal's forces are much reduced [Weisenhorn, Hansma, Albrecht and Quate[40] (1989)], or in ultra-high vacuum where the thickness of liquid films on the specimen surface are much reduced.

The interpretation of contact mode AFM images is much simpler than STM images due to the existence of a pseudo-constant, known as the *Hamaker* constant. In terms of the parameters of Equation 1.3.4, the Hamaker constant, H, is defined as:

$$H = 4\pi^2 W_o n_1 n_2 \sigma^6 . \tag{1.3.35}$$

This constant appears as a proportionality constant in the equations governing the magnitude of the force interaction between the cantilever and specimen with separation. Since this constant only varies over a small range of one or two orders of magnitude, the force interaction only varies by this much. Due to the high power dependence of separation in contact mode AFM, a two order of magnitude change in interaction strength corresponds only to a factor of two or so change in separation. Thus the spectroscopic influence due to the varying identity of the atoms interacting is much reduced in comparison with STM (where momentum conservation plays almost as big a role as energy conservation). Hence contact mode AFM renders a much more true image of the topography than STM, although at the expense of resolution and spectroscopic information.

Since AFM depends on forces, the presence of magnetic or electric dipoles or trapped charges in the specimen or tip introduces other elements to the interaction. Images from such specimens can be difficult to interpret. However, these interactions can also be utilised to provide information on charge or dipole distribution, etc. The other techniques described in the rest of this chapter deal with some of these techniques.

1.3.4 AFM Probe Fabrication

The original AFM cantilevers were simply modified STM tips or very thin pieces of metal foil. Unfortunately, the STM tips had a very high stiffness (not desirable for contact mode AFM) while the metal foils lacked a physical tip for access to the specimen. However, with the application of silicon micromachining technology [Binnig, *et al.*[41] (1987); Albrecht and Quate[42] (1988); Akamine, Barrett and Quate[43] (1990); Albrecht, Akamine, Carver and Quate[44] (1990)], AFM probe manufacture was revolutionised to the point where the whole spectrum of semiconductor processing technology could be used to produce AFM probes for a variety of applications. Mass produced AFM probes are now available commercially with a range of compliances, resonant frequencies and quality factors. The advantage of silicon is that its mechanical strength combined with low density allow cantilevers with relatively high resonant frequencies but low compliances to be made. In general most commercially available probes are made using a combination of wet and dry etching and anodic bonding. Probes with a radius of curvature as small as 10nm and a cone angle of only 10° have been made in this way. These can be coated with metal if required or further modified for use in some of the other AFM techniques described in the remainder of this chapter. The inert and non-toxic nature of silicon also makes it suitable for operation in biological and chemical environments.

1.3.5 Applications and Limitations

The applications of AFM are many and varied in areas from metrology and semiconductors to biology and tribology. Just as with the STM, AFM techniques have been made to work in a variety of environments from liquid [Sonnenfeld and Hansma[45] (1986)] to air. Due to the much reduced spectroscopic dependence of the force interaction, AFM has found use as a metrological tool [Griffith and Grigg[46] (1993); Griffith, Marchman, Miller and Hopkins[47] (1995); Martin and Wickramasinghe[48] (1995)]. It has also been used to evaluate surface roughness and the quality of deposited coatings during manufacturing processes. In the semiconductor industry AFM is routinely used to measure the roughness induced by various fabrication processes [e.g. Feenstra, *et al.*[49] (1995)]. Its ability to image non-conducting [Albrecht and Quate[50] (1987)] as well as conducting areas has made it particularly useful on samples where both are present.

Its ability to image insulating specimens in a liquid environment has also made it suitable for biological applications, such as the imaging of DNA [Gould, *et al.*[51] (1988); Rees, *et al.*[52] (1993); Erie, Yang, Schultz and Bustamante[53] (1994)], proteins [*e.g.* Radmacher, Fritz, Hansma and Hansma[54] (1994)] and cells [Radmacher, Tillmann, Fritz and Gaub[55] (1992); Henderson[56] (1994)] – no coating or special substrate is required [Drake, *et al.*[57] (1989); Haberle, Horber and Binnig[58] (1991)]. However, the forces can be large enough to move such specimens around so that innovative methods for securing the specimens have had to be found.

The wearing, lubrication and frictional properties of surfaces is yet another area where the AFM has found use. A modification of the detection system and AFM set-up can allow cantilever deflections (and hence forces), perpendicular to the scan direction to be measured [Correia, Garcia, Massanell and Costakramer[59] (1996)], or alternatively, the use of shear forces (experienced by oscillating the cantilever perpendicular to the scan direction) can also yield such information. This ability of the AFM to measure forces laterally as well as horizontally has given rise to the science now commonly known as nanotribology [Bhushan, Israelachvili and Landman[60] (1995)]. In an extension of this the AFM has also been used to study interfaces between solids and liquids, particularly with regard to the role of surfactants [Oshea, Welland and Rayment[61] (1992); Manne and Gaub[62] (1995)].

Atomic resolution imaging has also been useful for imaging ionic crystals, either in atmosphere or in a liquid environment where the dissolution process can be monitored. The study of Langmuir-Blodgett and other organic thin films have also been a popular application of AFM [Marti, *et al.*[63] (1988); Meyer, *et al.*[64] (1991); Chi, *et al.*[65] (1993)] as has the interaction force between an AFM probe coated with a few monolayers of one of a pair of binding molecules and a surface coated with the other [Florin, Moy and Gaub[66] (1994); Moy, Florin and Gaub[67] (1994)].

However, one of the most important applications of AFM has been its use as the principal separation regulating mechanism in many of the other SPM techniques described in the remainder of this chapter. This has been due, in no small part, to: the ease with which it can be combined with other 'sensors' using similar semiconductor processing technologies as those used to manufacture the probe itself; its ability to work in all kinds of environments and with insulating specimens, and; the reduced spectroscopic interaction between sample and probe.

It should be noted that AFM suffers from similar tip problems to STM. As with STM tips, AFM tips can suffer from multiple-tip averaging and tip switching. However, the probe in contact mode AFM is usually scanned much closer to the surface than with STM, which means that contact mode AFM is more likely to modify the specimen than STM. Non-contact mode AFM uses separations much greater than in STM and is sometimes preferred where the highest resolution is not required.

1.4 SCANNING 'X' MICROSCOPIES

As can be seen from the previous two sections, both STM and AFM are capable of being scanned very close to the surface of a specimen. This ability has allowed the development of techniques which measure other physical parameters at close proximity to the specimen. These *near-field* techniques improve on the resolution of *far-field* techniques by making use of the fact that most interactions decay inversely with separation so that measurements close to the specimen are in a region where the interaction varies more rapidly with separation. This in turn means that small changes in separation result in a large change in the physical parameter being measured. Such techniques are closely related to the scanning probe microscopy concept and are sometimes called 'scanning x microscopies' where the 'x' represents an arbitrary physical parameter such as temperature, magnetic field, etc. This section describes some of the more important techniques.

1.4.1 Near-Field Optical Microscopies

Optical techniques still constitute the bulk of the microscopy methods used today due to its many desirable characteristics such as speed of measurement, ease of use, non-invasive nature and its ability to provide detailed spectroscopic information. However, as with all far-field optical techniques, the spatial resolution of the information provided is limited to the diffraction limit of light. Therefore, the interest in applying such techniques in the near-field so that higher resolution can be achieved is justified. High resolution can be achieved by making use of evanescent fields which decay rapidly with separation in the manner first proposed by Synghe in 1928 [Synghe[68] (1928)]. The first experimental demonstration of this technique only occurred as recently as 1972 using microwaves [Ash and Nicholls[69] (1972)].

The difficulty has always been that the evanescent waves need to be created or detected in the near-field and the near-field signal is attenuated extremely sharply with respect to the original far-field signal so that it is necessary to operate at a separation of much less than half a wavelength using very sensitive detection and filtering schemes. One of the most important goals has been to obtain spectroscopic information with high spatial resolution.

The simplest method of creating a source of evanescent waves is to bounce a beam of light off the surface of glass prism at an angle to the normal greater than its critical angle. Since the reflection does not take place at the infinitesimal line marked by the boundary between glass and air (Uncertainty Principle), the volume of space outside the illuminated part of the boundary contains non-propagating waves or evanescent waves. If an object is placed in this region, these non-propagating waves can be converted to propagating ones through scattering by the object. However, this probability decays exponentially with distance from the boundary in a similar fashion to tunnelling current. The object used to convert the evanescent waves into propagating waves is the sensor. An obvious choice is a silicon nitride based AFM probe. This can be scanned over the back of the prism, within the evanescent field, so that light effectively tunnels from one side of the prism to the AFM probe. Such a scan would reveal information about imperfections on the back surface of the prism, while simultaneously providing a topographic image. To make this technique more generally useful, a very thin optical sample can be placed on the back of the prism so that the light has to tunnel through the specimen. The effects of stray light can be simply eliminated by modulating the illumination. This is the principle behind the photon scanning tunnelling microscope or PSTM [Reddick, Warmack and Ferrell[70] (1989); Reddick, *et al.*[71] (1990); Ferrell, *et al.*[72] (1991)].

The main problems with this technique stem from the fact that the area of prism illuminated is of the order of the diffraction limit of the light within the prism, so that scattering at the interface between the prism and the sample can result in the formation of propagating waves which are picked up by the sensor in addition to those tunnelling directly to the sensor through the specimen. The two effects are indistinguishable and for this reason, high resolution has never been satisfactorily demonstrated using this technique. In addition such propagating modes may not be picked up in the evanescent region, but also further away as the sensor does not confine its detection to the near-field region. An alternative to using an AFM probe, which has the advantage of collecting more of the light, is to use a tapered optical fibre, typically made by wet etching or with a micropipette puller.

However, here the regulation of the probe-sample separation is more difficult and is typically done using scanning friction force microscopy (often called 'shear force' microscopy [Betzig, Finn and Weiner[73] (1992)]), where the optical fibre is dithered laterally at close to its resonance, and the resulting amplitude, phase or frequency of the response is measured using either a quartz crystal tuning fork or an additional

interferometer. As the optical fibre approaches the specimen it experiences frictional forces which alters the resonant frequency in a manner similar to that in non-contact mode AFM. The disadvantages of this method are the reliance on friction, which varies dramatically with surface and environmental conditions, and the averaging of the interaction over the amplitude of the dither motion.

In order to improve upon the above technique, it is necessary to limit the light collected by the sensor to the evanescent region. One way of doing this is to coat the sensor with a metallic film whilst leaving a sub-wavelength aperture at its apex. The earliest such sensor was made from a mechanically ground and chemically polished quartz tip which was coated with an aluminium film [Durig, Pohl and Rohner[74] (1986)]. An aperture was defined in this film by pressing the apex against a flat, transparent surface while illuminating the tip from the uncoated side. The formation of the aperture could be detected by looking for light emission from the apex. Using such a sensor the problem of scattering at two interfaces can be eliminated by using the 'sensor' as the light emitter and collecting light which has tunnelled through both the sub-wavelength aperture and specimen using conventional far-field optics. This technique is known as scanning near-field optical microscopy or SNOM [Betzig, *et al.*[75] (1988)] (other acronyms include NSOM and NFOM). In practice the lifetime of these tips was found to be only a few hours as the aperture had a tendency to grow steadily with time. In this implementation, which marginally pre-dates the AFM, the distance regulation was carried out using electron tunnelling, making it necessary to coat the specimen with a very thin conducting layer.

Many attempts have been made to design better probes for this technique and the one that has become most popular is the coated, tapered optical fibre [Betzig and Trautman[76] (1992); Shalom, Lieberman, Lewis and Cohen[77] (1992); Betzig, *et al.*[78] (1993)]. The aperture is defined by evaporating metal at an oblique angle with respect to the optical fibre while rotating the fibre. Apertures as small as 12nm [Betzig, *et al.*[79] (1991)] have been formed although the technique is a 'black art'. The distance regulation scheme used is that of friction force microscopy, as described earlier in §1.3.5.

An alternative method which tries to improve on the PSTM technique is to use a metallic STM tip to scatter light in the evanescent region. The tip-sample separation can be modulated so that only light scattered by the tip is detected using far-field optics [Inouye and Kawata[80] (1994)]. The problem with this is that as the tip-sample separation decreases, quenching of the evanescent radiation dramatically reduces the signal level. It is also difficult to detect the resulting small scattering signal on such a large background level. In addition light may be scattered by the tip outside the evanescent region. A refinement of this is to use the techniques of confocal microscopy to limit light illumination to a 100nm region around the tip-sample area while illuminating the sample from behind with two beams, only one of which interacts with the tip, so that the reflection from the other, reference beam can be subtracted from the beam associated with the scattering (eliminating the problem of a small signal on a large background). In

this scheme only light scattered back through the sample is detected using a complex interferometric technique [Zenhausern, Oboyle and Wickramasinghe[81] (1994); Zenhausern, Martin and Wickramasinghe[82] (1995)]. The separation is regulated using intermittent-contact mode microscopy and the use of a sharp tip has allowed a resolution of 1nm to be obtained. This is the highest resolution achieved to date. This technique has been named scanning interferometric apertureless microscopy or SIAM. Its main disadvantages are that it cannot be used with optically opaque samples and the incredible complexity of the detection system required to achieve low noise and adequate amplification. It is also not apparent how the technique can be extended to provide spectroscopic information.

Yet another approach has been to form a photodetector, usually a semiconductor diode, at the apex of a tip. The principal difficulty here has been the fabrication of photodiodes with a small active area. Many schemes have been devised from using large area detectors in close proximity to the sample to coating these diodes with a metallic layer so as to leave an aperture at the apex [Davis, Williams and Neuzil[83] (1996); Yamada, *et al.*[84] (1996)]. These probes, which use standard AFM distance regulation, rely on the generation of carriers by light scattered into the detector. However, the number of carriers also has an exponential dependence on temperature so that eliminating thermal effects becomes a major concern. In addition, the technique cannot be extended to optically opaque samples and rules out many spectroscopy techniques.

Attempts have also been made to directly form a sub-wavelength aperture in an AFM probe coated in a metallic film [Lewis, Isaacson, Harootunian and Muray[85] (1984)]. Again the small dimensions of the aperture required have made this extremely difficult. One group have succeeded in making a probe with a 250nm aperture using a combination of micromachining and boron doping [Tortonese, Yamada and Quate[86] (1991)]. However, a reliable method for making such sensors repeatedly has not yet been demonstrated.

Finally, the possibility of putting the light source at the end of the tip has also been looked at [Lieberman, Harush, Lewis and Kopelman[87] (1990); Lewis and Lieberman[88] (1991); Lieberman and Lewis[89] (1992); Gottlich, Lang and Heckl[90] (1995)]. This involves coating the apex of an AFM probe with some light emitting material which can be excited externally using light of a different wavelength to that emitted. The main problem with these techniques is that as the separation between probe and sample decreases, it becomes increasingly difficult for enough of the excitation signal to reach the apex of the probe, while the probability of emission also decreases due to quenching effects.

Many applications have been found for SNOM based on the extension of far-field techniques to the near-field domain. This is due to the large variety of contrast mechanisms that have been found [Betzig, *et al.*[76] (1992); Trautman, *et al.*[91] (1992)] including polarisation sensitivity, refractive index sensitivity, localised fluorescence and localised luminescence. Some applications include high-density data storage on magneto-

optic materials [Betzig, *et al.*[92] (1992); Betzig, *et al.*[93] (1993); Terris, *et al.*[94] (1994)]; single molecule fluorescence by chemical staining [Betzig and Chichester[95] (1993); Trautman, Macklin, Brus and Betzig[96] (1994)]; spectroscopy of semiconductor surfaces and laser diodes [Hollricher, *et al.*[97] (1995)]; local orientation determination of liquid crystal polymers using polarisation [Nagahara and Tokumoto[98] (1996)]; and photoluminescence of quantum wells and wires [Grober, *et al.*[99] (1994); Hess, *et al.*[100] (1994)]. In particular, localised fluorescence looks to be a very useful technique for biologists since imaging can be carried out with live cells, proteins, bacteria, etc., in the appropriate environment, without the need for the metal coating or vacuum environments associated with other techniques. By tagging molecules of interest with fluorescent molecules, SNOM can locally excite emission, allowing a more precise determination of the position of the fluorescent tags. Another very useful technique is that of localised photoluminescence. SNOM can carry out spectroscopy on individual quantum wells and wires, thus eliminating the problems associated with inhomogeneous broadening. This then allows the comparison of one quantum well structure with another.

SNOM is a technique which is just beginning to take off. It is still hampered by the difficulty in obtaining suitable probes. Despite demonstrations of very small apertures, it is still more common for SNOM probes used in real experiments to have aperture diameters in excess of 100nm. This is because the geometry of the fibre probes in common use result in large signal losses as the aperture diameter becomes smaller. This is caused by the long length of taper which extends the length of the region over which light is cut-off and decaying exponentially. In principle, the resolution achievable with SNOM should be approximately two or three times the decay length of the light in the surrounding metallic shielding. For aluminium the decay length (intensity) is of the order of 6.5nm in the visible, so that resolutions of 15 to 20nm should be possible. This can only be achieved when small apertures with much lower attenuation of the incident beam are available. Matters are further complicated with those probes requiring shear-force distance regulation as the probe is vibrated in a direction parallel to the sample surface.

For SIAM, resolutions down to atomic level may be possible, however the difficulty of the experiment and the high specifications required of the apparatus have meant that very few groups have been able to duplicate the original work, while the original group has also managed to achieve relatively few images using their apparatus.

For both SNOM and SIAM, the interaction between tip and sample is very complex and no widely accepted theoretical models exist. Even the calculation of the attenuation of light in SNOM probes has proven difficult. Whilst the microwave community have performed many calculations concerned with the degree of attenuation in cut-off waveguides (such as in antennae), the calculations have always assumed that the electric field at the metal-dielectric interface has no tangential component (and the magnetic field has no normal component) [see for example Barlow[101] (1947)]. Then the attenuation is simply that of a waveguide operating below its cut-off frequency. In cylindrical co-

ordinates (r,θ,z), the field amplitudes decay exponentially according to [Bleaney and Bleaney[102] (1976)]:

$$\mathbf{E}(r,\theta,z)=\mathbf{E}_o(r,\theta)\exp\left[-j\frac{2\pi z}{\lambda_g}\right],\ \mathbf{H}(r,\theta,z)=\mathbf{H}_o(r,\theta)\exp\left[-j\frac{2\pi z}{\lambda_g}\right], \tag{1.4.1}$$

where $\mathbf{E}(r,\theta,z)$ and $\mathbf{H}(r,\theta,z)$ are the electric and magnetic field distributions at (r,θ,z), respectively, $\mathbf{E}_o(r,\theta)$ and $\mathbf{H}_o(r,\theta)$ are the cross-sectional electric and magnetic field distributions at $z=0$ and λ_g is the apparent wavelength in the waveguide given by the waveguide equation [Barlow[101] (1947)]:

$$\frac{n^2}{\lambda_o{}^2}=\frac{1}{\lambda_g{}^2}+\frac{1}{\lambda_c{}^2}, \tag{1.4.2}$$

where λ_o is the wavelength in vacuum of the radiation, λ_c is the cut-off wavelength of the waveguide and n is the refractive index of the waveguide medium. Notice that λ_g is purely imaginary for wavelengths longer than the cut-off wavelength.

Mode	$\eta_{p,q}$, $\eta'_{p,q}$	${\lambda_c}^{p,q}/R$
$TE_{1,1}$, $H_{1,1}$	1.841183	3.412580
$TM_{0,1}$, $E_{0,1}$	2.404826	2.612741
$TE_{2,1}$, $H_{2,1}$	3.059329	2.053779
$TM_{1,1}$, $E_{1,1}$	3.831706	1.639788
$TE_{0,1}$, $H_{0,1}$	3.831706	1.639788

Table 1.4.1: Bessel function constants and cut-off wavelengths for a circular waveguide.

For a circular waveguide with perfectly conducting walls the field solutions involve Bessel functions so that the cut-off wavelengths for the transverse electric modes, $TE_{p,q}$, and the transverse magnetic modes, $TM_{p,q}$, are given by [Reich, Ordung, Krauss and Skalnik[103] (1953); Waldron[104] (1970); Combes[105] (1991)]:

$$TE_{p,q},\ H_{p,q}\colon \lambda_c=\frac{2\pi R}{\eta'_{p,q}},\quad TM_{p,q},\ E_{p,q}\colon \lambda_c=\frac{2\pi R}{\eta_{p,q}}, \tag{1.4.3}$$

where the inner radius of the waveguide is R_o and $\eta_{p,q}$, $\eta'_{p,q}$ are defined by:

$$J_p\left(\eta_{p,q}\right)=0,\ \frac{dJ_p\left(\eta'_{p,q}\right)}{dr}=0 \tag{1.4.4}$$

and J_p is a Bessel function of the first kind (the q subscript refers to the order of the zero, a zero subscript indicates the first zero, *etc.*). Table 1.4.1 gives the values of these for the first few modes.

Field	Transverse Electric $TE_{p,q}$, $H_{p,q}$	Transverse Magnetic $TM_{p,q}$, $E_{p,q}$
$E_z^{\,o}(r,\theta)$	0	$A_{p,q}J_p\left(k_c^{\,p,q}r\right)\cos p\theta$
$E_r^{\,o}(r,\theta)$	$jA'_{p,q}Z\frac{np}{k_c^{\,p,q}r}\frac{\lambda_c^{\,p,q}}{\lambda_o}J_p\left(k_c^{\,p,q}r\right)\sin p\theta$	$-jA_{p,q}\frac{\lambda_c^{\,p,q}}{\lambda_g^{\,p,q}}\frac{dJ_p\left(k_c^{\,p,q}r\right)}{dr}\cos p\theta$
$E_\theta^{\,o}(r,\theta)$	$jA'_{p,q}Z\frac{n\lambda_c^{\,p,q}}{\lambda_o}\frac{dJ_p\left(k_c^{\,p,q}r\right)}{dr}\cos p\theta$	$jA_{p,q}\frac{p}{k_c^{\,p,q}r}\frac{\lambda_c^{\,p,q}}{\lambda_g^{\,p,q}}J_p\left(k_c^{\,p,q}r\right)\sin p\theta$
$H_z^{\,o}(r,\theta)$	$A'_{p,q}J_p\left(k_c^{\,p,q}r\right)\cos p\theta$	0
$H_r^{\,o}(r,\theta)$	$-jA'_{p,q}\frac{\lambda_c^{\,p,q}}{\lambda_g^{\,p,q}}\frac{dJ_p\left(k_c^{\,p,q}r\right)}{dr}\cos p\theta$	$-jA_{p,q}\frac{1}{Z}\frac{np}{k_c^{\,p,q}r}\frac{\lambda_c^{\,p,q}}{\lambda_o}J_p\left(k_c^{\,p,q}r\right)\sin p\theta$
$H_\theta^{\,o}(r,\theta)$	$jA'_{p,q}\frac{p}{k_c^{\,p,q}r}\frac{\lambda_c^{\,p,q}}{\lambda_g^{\,p,q}}J_p\left(k_c^{\,p,q}r\right)\sin p\theta$	$-jA_{p,q}\frac{1}{Z}\frac{n\lambda_c^{\,p,q}}{\lambda_o}\frac{dJ_p\left(k_c^{\,p,q}r\right)}{dr}\cos p\theta$

Table 1.4.2: Field distribution amplitudes for modes in a perfect cylindrical metallic waveguide after **Waldron**[104] **(1970).**

The cross-sectional electric and magnetic field distributions are given by:

$$\begin{aligned}\mathbf{E}_o(r,\theta) &= E_r^{\,o}(r,\theta)\mathbf{e}_r + rE_\theta^{\,o}(r,\theta)\mathbf{e}_\theta + E_z^{\,o}(r,\theta)\mathbf{e}_z\\ \mathbf{H}_o(r,\theta) &= H_r^{\,o}(r,\theta)\mathbf{e}_r + rH_\theta^{\,o}(r,\theta)\mathbf{e}_\theta + H_z^{\,o}(r,\theta)\mathbf{e}_z\end{aligned}, \tag{1.4.5}$$

where $\mathbf{e}_r$, $\mathbf{e}_\theta$, and $\mathbf{e}_z$ are unit vectors in cylindrical coordinates, (r,θ,z); the component amplitudes are defined in Table 1.4.2; and:

$$Z = \sqrt{\frac{\mu}{\varepsilon}}, \tag{1.4.6}$$

is the characteristic impedance of the dielectric in terms of the absolute permittivity and permeability;

$$k_c^{\,p,q} = \frac{2\pi}{\lambda_c^{\,p,q}}, \tag{1.4.7}$$

defines the cut-off wavenumber in terms of the cut-off wavelength; $A_{p,q}$, $A'_{p,q}$ are normalisation constants which have the value [Waldron[104] (1970)]:

$$\begin{aligned}A'_{0,q} &= j\frac{\eta'^{\,2}_{0,q}\lambda_o}{n\pi R^2 J_0\left(\eta'_{0,q}\right)}\sqrt{\frac{\lambda_o\lambda_g}{2n\pi Z}}, \quad A'_{p,q} = j\frac{\eta'^{\,2}_{p,q}\lambda_o}{n\pi R^2\sqrt{\eta'^{\,2}_{p,q}-p^2}J_p\left(\eta'_{p,q}\right)}\sqrt{\frac{\lambda_o\lambda_g}{n\pi Z}},\ p\neq 0\\ A_{0,q} &= j\frac{\eta_{0,q}\lambda_o}{n\pi R^2\dfrac{dJ_p\left(\eta_{0,q}\right)}{dr}}\sqrt{\frac{\lambda_o\lambda_g Z}{2n\pi}}, \quad A_{p,q} = j\frac{\eta_{p,q}\lambda_o}{n\pi R^2\dfrac{dJ_p\left(\eta_{p,q}\right)}{dr}}\sqrt{\frac{\lambda_o\lambda_g Z}{n\pi}},\ p\neq 0\end{aligned}, \tag{1.4.8}$$

for a total mode power of:

$$P = \frac{\lambda_o^2}{n^2} \text{ Watts}, \tag{1.4.9}$$

and n is the refractive index of the dielectric defined by:

$$n = \sqrt{\mu_r \varepsilon_r}, \tag{1.4.10}$$

in terms of the relative permittivity and permeability.

More accurate solutions for real metals can also be calculated and show that the critical wavelengths differ slightly from those given by the above equations [Waldron[104] (1970)], however, the calculation is involved and is beyond the scope of this thesis. Which modes are excited in the waveguide depends on matching the fields in the unshielded fibre with those in the shielded part at the boundary. Even then the calculation is not straightforward since as the fibre tapers, the radius decreases, the critical wavelength changes and the mode field structure changes. Then an overlap integral needs to be performed to calculate how the fields evolve as they travel down the fibre. The calculation is further complicated because the exact angle of the taper has an effect on how much and in what direction energy is reflected off the metal walls as it travels down the fibre.

Losses will also be experienced in the metallic walls of the waveguide. The equation for an electromagnetic plane wave travelling in a non-conducting, homogeneous, isotropic and linear medium in terms of its electric and magnetic fields is:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0 e^{j(\omega t - \mathbf{k}.\mathbf{r})}, \mathbf{H}(\mathbf{r}) = \mathbf{H}_0 e^{j(\omega t - \mathbf{k}.\mathbf{r})}, \tag{1.4.11}$$

where ω is the angular frequency of the radiation. However, the wavenumber can be generalised to include conducting media according to:

$$k = \frac{2\pi N}{\lambda_o} = \frac{\omega N}{c}, N = n - j\kappa, \tag{1.4.12}$$

where N is the complex refractive index with n and κ representing its real and imaginary parts respectively and c is the speed of light in vacuum. Both the real and imaginary parts are functions of frequency and have been tabulated for frequencies in the visible [Weaver and Frederikse[106] (1996)]. Sometimes data is presented in terms of the complex relative permittivity, where the values can be converted as shown below. The complex refractive index can also be defined in terms of the conductivity:

$$\varepsilon_r' = \varepsilon_1 - j\varepsilon_2 = \varepsilon_r\left(1 - \frac{j\sigma}{\omega\varepsilon}\right) \Rightarrow n^2 - \kappa^2 = \mu_r\varepsilon_1 = \mu_r\varepsilon_r, 2n\kappa = \mu_r\varepsilon_2 = \frac{\mu_r\sigma}{\omega\varepsilon_o}, \tag{1.4.13}$$

where σ is the conductivity of the metal and is a function of frequency. It is convenient to split the wavenumber into real and imaginary parts, β and α respectively:

$$k = \beta - j\alpha, \beta = \frac{2\pi n}{\lambda_o}, \alpha = \frac{2\pi\kappa}{\lambda_o}, \mathbf{k} = k\mathbf{e}_k, \tag{1.4.14}$$

so that Equation 1.4.11 becomes:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0 e^{-\alpha\mathbf{e}_k.\mathbf{r}} e^{j(\omega t - \beta\mathbf{e}_k.\mathbf{r})}, \mathbf{H}(\mathbf{r}) = \mathbf{H}_0 e^{-\alpha\mathbf{e}_k.\mathbf{r}} e^{j(\omega t - \beta\mathbf{e}_k.\mathbf{r})}. \tag{1.4.15}$$

From this it is clear that the wave amplitude attenuates with a characteristic decay length given by:

$$\delta = \alpha^{-1} = \frac{c}{\omega\kappa}, \tag{1.4.16}$$

where δ is often known as the skin depth of radiation in the metal. Note that the power attenuation coefficient is larger by a factor of two since the power flow is proportional to the product of the electric and magnetic field amplitudes.

To calculate the attenuation in the walls of the guide, the fields are assumed to be perturbed very little from the ideal case, so that the field components at the metal interface can be used to calculate the displacement and eddy currents flowing in the walls, which when combined with the resistivity of the metal, allows an approximation to the losses in the walls to be made. To find the attenuation coefficient, these losses are compared with the power flowing along the waveguide (as calculated from the Poynting vector). For the circular metallic waveguide this is given by [Reich, *et al.*[103] (1953)]:

$$\alpha_{TE} = \frac{R_s}{Z_o}\frac{1}{R\sqrt{1-\frac{\lambda_o^{\,2}}{n^2\lambda_c^{\,2}}}}\left[\frac{p^2}{\eta'^{\,2}_{p,q}-p^2}+\frac{\lambda_o^{\,2}}{n^2\lambda_c^{\,2}}\right], \alpha_{TM} = \frac{R_s}{Z_o}\frac{1}{R\sqrt{1-\frac{\lambda_o^{\,2}}{n^2\lambda_c^{\,2}}}}, \tag{1.4.17}$$

where Z_o is the characteristic impedance of vacuum:

$$Z_o = \sqrt{\frac{\mu_o}{\varepsilon_o}}, \tag{1.4.18}$$

and R_s is known as the skin resistance, defined by:

$$R_s = \frac{1}{\sigma\delta} \approx \frac{Z_o}{2n}, \text{ if } \mu_r \approx 1. \tag{1.4.19}$$

The total attenuation is then the sum of both attenuation effects. Unfortunately, in the case of SNOM, the fields are very severely disturbed from the ideal metal case. The approximations above are only valid if the skin depth in the metal walls is much smaller than the diameter of the guide – this is unfortunately not the case with SNOM probes at optical frequencies. The whole situation is further complicated by the presence of both the core and cladding dielectrics in the optical fibre.

One alternative is to use finite element modelling to calculate the fields in a SNOM probe. However, this is complicated because the profile of the fibre is not well known and varies dramatically from probe to probe. In addition, the attenuation is so great (in excess of six orders of magnitude) that a very large number of elements must be used and the accumulative error in the numerical algorithm (and even in the binary representation of floating point numbers) may well exceed the actual light output from the probe.

Both SNOM and SIAM give information which is very strongly dependent on spectroscopic effects as well as refractive index and geometry. Thus, as with STM, the number of artefacts using these techniques is expected to be large. However, unlike STM,

the nature of the interaction is given an additional dimension of complexity by its dependence on the polarisation aspects of electromagnetic waves which means that the optical techniques may demonstrate even more artefacts than STM, although SNOM at least has the minor advantage that the interaction is averaged over the aperture diameter. Arguably, it is precisely because there are so many different contrast mechanisms, that there will be so many artefacts. However, whilst the artefacts in far-field optics can be understood because of humankind's familiarity with its sense of sight, this intuition does not readily extend to the near-field regime where electromagnetic fields behave in an entirely different manner.

1.4.2 Scanning Thermal Microscopies

The measurement of thermal properties is a very important area as is reflected by the existence of many far-field techniques such as optical pyrometry and laser interferometric thermometry for temperature determination [Saenger and Gupta[107] (1991)]; photothermal deflection spectroscopy for optical absorption determination [Jackson, Amer, Boccara and Fournier[108] (1981)]; laser-modulated thermoreflectance for the determination of material inhomogeneities [Guidotti and Vandriel[109] (1985)], and; differential scanning calorimetry for the measurement of heat capacities and phase transition temperatures [Wunderlich and Gaur[110] (1983); Compton, Johnson and Powell[111] (1993); Johnson, Stout, Hill and Krishnan[112] (1994)]. However, most of the above techniques are based on optical methods and therefore the resolution achievable is limited to the diffraction limit *i.e.* greater than half a wavelength. For this reason there has been much interest in making measurements in the near-field where the resolution is more likely to be limited by the physical dimensions of the sensor and its separation from the specimen, both of which can be made much smaller than the resolution limit of the far-field techniques.

One of the first attempts at achieving this was made by forming a simple thermocouple at the end of a conical tip [Williams and Wickramasinghe[113] (1986)]. The thermocouple junction measured 100nm and was scanned across the specimen. This experiment was carried out prior to the publication of the AFM technique and the motivation behind it was not the measurement of temperature, but of the topography of non-conducting samples. [The AFM was subsequently shown to be a better technique for this purpose.] To this end the probe was heated to a temperature above ambient and the probe was scanned across the specimen so as to maintain a fixed temperature. This temperature is determined by the thermal conductance between probe and sample and hence separation. Thus a topographic image could be obtained in this way. However, it is clear that variations in the thermal conductivity of the sample from one region to another will affect the image obtained in a similar manner to the sample topography, making it difficult to separate one effect from the other.

Since thermocouples measure temperature by measuring the change in contact potential difference at one junction with respect to another (at a different temperature), another

technique used an STM tip as one side of a thermocouple junction with the specimen acting as the other [Weaver, Walpita and Wickramasinghe[114] (1989)]. Here a modulated laser signal was used to uniformly heat the specimen, resulting in the modulation of the thermoelectric potential across the tunnel junction. Since this modulation amplitude is proportional to temperature variation and hence laser absorption, spectroscopic information could be obtained by measuring the modulation amplitude as a function of wavelength. To eliminate topographic effects, this amplitude is measured with respect to a second modulated reference laser at a fixed wavelength. This technique achieves a resolution comparable with STM. However, it is clear that since the composition of the thermocouple junction varies with position, the thermoelectric response (Seebeck coefficient) does not remain constant and hence spatial variations may be caused by a superposition of several effects including composition, absorption coefficient, sample uniformity, etc. In addition the high junction resistance results in a greater noise level on the signal which then must be measured in a wider bandwidth than would otherwise be desirable.

In order to eliminate some of these problems, another group combined both of the above techniques by forming a real thermocouple using a constantan STM tip with a gold coating confined to the last 10μm of the tip [Oesterschulze, *et al.*[115] (1996)]. The tip was then maintained at a temperature above ambient by heating it with a laser and the tip was scanned at a uniform height above the sample using a constant tunnelling current (and thus obtaining a topographic image) while simultaneously measuring the STM thermocouple temperature. This temperature will vary with the thermal conductance between the tip and sample, and assuming the physical separation is fixed, the temperature image gives a spatial measurement of the thermal conductivity across the sample. The main disadvantages are the large size of the thermocouple junction and that the technique is limited to those samples on which STM can be carried out *i.e.* electrically conducting samples.

To eliminate the last of these restrictions, another method replaced the STM with a coated AFM to form a modified Kelvin Probe Microscope [see Nonnenmacher, Oboyle and Wickramasinghe[116] (1991) for a description]. This microscope measures the contact potential difference between the probe and the sample. Then by heating the far end of the probe a constant force topographic image can be obtained simultaneously with contact potential difference (and hence temperature). The temperature depends on the thermal conductivity as with the preceding technique. However, in common with the earlier STM technique, the contact potential difference also depends on the composition of the sample and therefore varies with position even in the absence of changes in thermal conductivity.

Another group which has been prolific in this area began their attempts by forming a thermocouple using alumel and chromel wires [Majumdar, Carrejo and Lai[117] (1993)]. They used capacitive discharge between the two wires to bond them together and then sheared the resulting wire as close as possible to the junction. In order to form a sharper

tip they electrochemically etched the junction. They opted to use AFM for distance regulation by bending the wires and attaching a piece of aluminium foil at the back for optical deflection detection. This was then scanned so as to obtain simultaneous topography and temperature images. However, the smallest junction they could manufacture was 25μm in diameter and the resolution achieved was also of this order. In addition the weak geometry of the probe structure meant that deformation of the tip occurred during scanning. To improve on this, instead of bending their wires, they attached it to a diamond fragment [Lai, Chandrachood, Majumdar and Carrejo[118] (1995)]. This sharp fragment then acted as the tip for the AFM while its excellent thermal conductivity transmitted heat from the specimen to the thermocouple. The result was an improved resolution of 0.5μm.

As an alternative approach they attempted to convert an AFM cantilever to a bi-metallic strip as originally carried out by another group [see Barnes, *et al.*[119] (1994); Barnes, *et al.*[120] (1994)]. The idea was to evaporate gold or aluminium onto commercially available silicon nitride cantilevers, thus forming a bi-metallic strip [Nakabeppu, *et al.*[121] (1995)]. Then temperature changes would result in the silicon nitride and the metal film expanding by different amounts causing the cantilever to bend. However, it was found that thermal conduction through the air dominated over other routes so that the resolution they achieved was of the order of the length of their cantilever, *i.e.* 200μm. In order to circumvent this they transferred their microscope to a vacuum environment, but then found that the thermal response time (due to the amount of energy required to bend the cantilever detectably) was much too large to be practical. It is also apparent that separating deflections due to the AFM interaction from those due to temperature is not straightforward.

In a review of their work, they concluded that in order to obtain high resolution, the temperature sensor must be of the order of the radius of curvature of the tip [Majumdar, *et al.*[122] (1995)]. The problem is that this requires the sensor dimensions to be of the order of nanometres, rather than microns, which is pushing current fabrication technology close to its limit. In order to achieve this they have tried to form a thin film thermocouple at the end of an AFM tip using the electrical breakdown between two metal films through an insulating layer. However, they have found that such thermocouples do not appear to last the duration of a scan, probably due to the very tenuous contact between the two metal films.

In yet another approach, they also tried to form a Schottky diode at the end of an AFM probe by forming a small aperture in a silicon nitride insulating layer coated over a doped silicon AFM tip and depositing metal over this aperture. Carrier generation in such a diode varies exponentially with temperature and so should be sensitive to temperature. The main problem with such a sensor is the thermal conductance of heat away from the junction by the metal film coating the cantilever.

In summary, it is clear that essentially two types of thermal measurements are dominating the development of near-field thermal sensors, namely, temperature and thermal conductivity. For the former, the ideal sensor would be a nanometre sized thermocouple held in close proximity to the specimen using atomic force regulation. For the latter a heater-thermocouple combination would seem to be the best solution – again on a scale of nanometres. The principle behind both these measurements has been proven and it is only the manufacture of such sensors which is preventing these techniques from becoming mainstream.

1.4.3 Scanning Electric Force Microscopies

The measurement of electrical properties, such as potential, on a microscopic scale, is very useful for investigating many kinds of semiconductor devices. Such microscopies can reveal: the presence of trapped charges; information on surface potentials, relative donor densities, surface photovoltage and carrier recombination; the extents of depletion regions, and; carrier diffusion characteristics. This is possible because the cantilever-sample system can be viewed as a complex capacitor arrangement.

It is usual to simplify the problem by modelling the cantilever as a either a sphere or parallel plate with the sample as a parallel plate of the capacitor system. In both cases the capacitance varies inversely with separation (approximately). The energy stored, W, in the capacitor system is then simply given by:

$$W = \frac{1}{2}CV^2, \tag{1.4.20}$$

where C is the capacitance and V is the potential difference. Such a potential difference would induce a charge, q_{tip}, on the tip, given by:

$$q_{tip} = q_{sample} + CV, \tag{1.4.21}$$

where q_{sample} is the charge present on the sample surface.

Then in addition to the forces in Equation 1.3.7, the cantilever would experience an electrostatic force given by:

$$F_{electrostatic}(z) = \frac{q_{sample}q_{tip}}{4\pi\varepsilon z^2} + \frac{1}{2}V^2\frac{dC}{dz} = \frac{{q_{sample}}^2}{4\pi\varepsilon z^2} + \frac{q_{sample}CV}{4\pi\varepsilon z^2} + \frac{1}{2}V^2\frac{dC}{dz}, \tag{1.4.22}$$

where $F_{electrostatic}(z)$ is the electrostatic component of the force, z is the separation between the cantilever and sample and ε is the dielectric constant between the interacting charges. To separate the electrostatic forces from the others, the potential between probe and sample is modulated *i.e.*:

$$V = V_{dc} + V_{ac}\cos\omega_{el}t, \tag{1.4.23}$$

where V_{dc} is the d.c. component of the potential difference between sample and probe, V_{ac} is the amplitude of the potential difference modulation and ω_{el} is the angular modulation frequency.

Combining Equations 1.4.22 and 1.4.23 and separating the components by frequency gives:

$$F_{electrostatic}(z) = \frac{q_{sample}^{2}}{4\pi\varepsilon z^{2}} + \frac{q_{sample}C}{4\pi\varepsilon z^{2}}V_{dc} + \frac{1}{2}\left[V_{dc}^{2} + \frac{1}{2}V_{ac}^{2}\right]\frac{dC}{dz} + \left[\frac{q_{sample}C}{4\pi\varepsilon z^{2}}V_{ac} + V_{ac}V_{dc}\frac{dC}{dz}\right]\cos\omega_{el}t + \frac{1}{4}V_{ac}^{2}\frac{dC}{dz}\cos 2\omega_{el}t \quad (1.4.24)$$

This force modulation causes the cantilever to oscillate at ω_{el} and $2\omega_{el}$. Since these forces are attractive and strong, the distance regulation is provided using non-contact mode force microscopy for stability. This means that the cantilever is already being forced to oscillate at a frequency ω_{top} for distance regulation. The result is a superposition of these motions. However, unless ω_{el} (and $2\omega_{el}$) are within the resonance envelope of the cantilever, the amplitude response of the cantilever at these frequencies is likely to be very small. The response at ω_{top} depends on the total force gradient so that:

$$\omega_{o}'^{2} = \omega_{o}^{2}\left[1 + \frac{k_{atomic}(z)}{k} + \frac{k_{electrostatic}(z)}{k}\right], \quad (1.4.25)$$

where:

$$k_{electrostatic}(z) = -\frac{dF_{electrostatic}(z)}{dz}$$

$$= \frac{q_{sample}^{2}}{2\pi\varepsilon z^{3}} + \frac{q_{sample}C}{2\pi\varepsilon z^{3}}V_{dc} - \frac{q_{sample}}{4\pi\varepsilon z^{2}}V_{dc}\frac{dC}{dz} - \frac{1}{2}\left[V_{dc}^{2} + \frac{1}{2}V_{ac}^{2}\right]\frac{d^{2}C}{dz^{2}} + \left[\frac{q_{sample}C}{2\pi\varepsilon z^{3}}V_{ac} - \frac{q_{sample}}{4\pi\varepsilon z^{2}}V_{ac}\frac{dC}{dz} - V_{ac}V_{dc}\frac{d^{2}C}{dz^{2}}\right]\cos\omega_{el}t - \frac{1}{4}V_{ac}^{2}\frac{d^{2}C}{dz^{2}}\cos 2\omega_{el}t \quad (1.4.26)$$

For electrostatic force gradients which are small compared to the cantilever spring constant this results in the coupled damped resonant *frequency* oscillating at ω_{el} and $2\omega_{el}$. This means that for a fixed ω_{top} on the linear part of the amplitude response curve, the amplitude of the response is also modulated at these frequencies. Thus the cantilever oscillates at ω_{el}, $2\omega_{el}$, ω_{top}, $\omega_{top} \pm \omega_{el}$ and $\omega_{top} \pm 2\omega_{el}$. If ω_{el} is chosen small enough, then the amplitude of the oscillation will remain within the linear region of the amplitude response. The same oscillations will appear on the phase response too. Then simply demodulating the amplitude signal at ω_{top} will produce the signal corresponding to Equation 1.4.26. The advantage of these schemes is that the gain of the response of the cantilever at ω_{top} (quality factor) is used to enhance the signal level of the signals at ω_{el} and $2\omega_{el}$. Alternatively, a frequency detection scheme can be used where the amplitude of the frequency modulation corresponds to Equation 1.4.26. In this case it is important that the amplitude of the electrostatic components at ω_{el} and $2\omega_{el}$ are sufficiently low that narrowband FM demodulation techniques can be used. However, due to the nature

of FM modulation, the signal at $2\omega_{el}$ will be obscured by the second harmonic of the signal at ω_{el}.

To understand what information can be obtained from these frequencies, it is assumed for simplicity that:

$$C \propto \frac{1}{z} \Rightarrow C = \frac{\kappa}{z} \Rightarrow \frac{dC}{dz} = -\frac{\kappa}{z^2} = -\frac{C}{z}, \frac{d^2C}{dz^2} = \frac{2\kappa}{z^3} = \frac{2C}{z^2}. \quad (1.4.27)$$

Then:

$$k_{electrostatic}(z) = \frac{q_{sample}{}^2}{2\pi\varepsilon z^3} + \frac{3q_{sample}C}{4\pi\varepsilon z^3}V_{dc} - \frac{1}{2}\left[V_{dc}{}^2 + \frac{1}{2}V_{ac}{}^2\right]\frac{d^2C}{dz^2} + \left[\frac{3q_{sample}C}{4\pi\varepsilon z^3}V_{ac} - V_{ac}V_{dc}\frac{d^2C}{dz^2}\right]\cos\omega_{el}t - \frac{1}{4}V_{ac}{}^2\frac{d^2C}{dz^2}\cos 2\omega_{el}t \quad (1.4.28)$$

It is clear from this that many operating methods are possible. If the sample is a metal then it is clear that the charge on the sample will be equal and opposite to the charge on the tip and the voltage between the surface of the sample and the tip will be equal to the voltage applied between the sample and tip. This is not true of a dielectric or semiconductor where there may be subsurface capacitances and trapped charges. One particularly useful method is to alter the voltage on the tip actively so as to minimise the amplitude of the component at ω_{el} using negative feedback. This occurs when $V_{dc} = 0$ so that the voltage required to do this is equal to the voltage on the surface of the sample thus allowing a map of the surface potential to be obtained. Another benefit of this technique is that the amplitude at ω_{top} more accurately represents the topography of the sample since all components involving the d.c. potential difference disappear. In addition, the remnant amplitude of the signal at ω_{el} is then directly proportional to the surface charge present allowing a map of the charge distribution to be made simultaneously. Another useful measurement is of the amplitude of the component at $2\omega_{el}$ which gives a direct measurement of the capacitance (through Equation 1.4.27).

Many of these techniques have been used with varying degrees of success [Martin, Abraham and Wickramasinghe[123] (1988); Williams, Hough and Rishton[124] (1989); Williams, Slinkman, Hough and Wickramasinghe[125] (1990); Abraham, Williams, Slinkman and Wickramasinghe[126] (1991); Weaver and Abraham[127] (1991); Weaver and Wickramasinghe[128] (1991); Domansky, *et al.*[129] (1993); Huang and Williams[130] (1994); Cunningham and Weaver[131] (1995)]. The main problem with the technique is that although the sharp dependence of the interaction on the separation means that the forces between the tip and sample will dominate, the large surface area of the rest of the cantilever and the area surrounding the apex of the tip contribute significantly to the stray capacitance. Thus the interaction is not well localised resulting in a loss of resolution. To circumvent this problem what is required is a probe in which all but the apex of the tip is shielded with a metal layer kept at ground potential. Unfortunately the manufacture of such a probe is non trivial.

1.4.4 Scanning Magnetic Force Microscopies

The measurement of the magnetic properties of materials is very important, particularly as the drive to higher and higher data storage densities on magnetic and magneto-optic media continues. In chemistry, nuclear-magnetic resonance (NMR) and related techniques are used to identify molecular structure. There is also much interest in the study of magnetic flux vortices in superconductors and domain formation during the superconducting transition.

Traditionally, magnetic domains have been investigated using Bitter decoration, electron microscopy techniques such as transmission electron microscopy (TEM) and scanning electron microscopy with polarisation analysis (SEMPA) or optical techniques based on the Faraday rotation of light passing through a magnetic material [Scheinfein, *et al.*[132] (1990); Crooker, Awschalom and Samarth[133] (1995); Frost, *et al.*[134] (1996)]. All these techniques have their advantages and disadvantages. Optical techniques do not offer high resolution, while TEM requires the samples to be sliced into very thin membranes and SEMPA requires a high vacuum environment. Therefore there has been great interest in applying near-field probe techniques to allow imaging with high resolution, high sensitivity or in alternative environments.

One method being looked at involves combining aspects of SEMPA with STM. A sharp metal tip is held close to the sample (although about two orders of magnitude further than in ordinary STM) and is forced to field emit by applying a large bias between sample and tip. The current is used to maintain a constant height, allowing topography to be imaged simultaneously, while the electrons emitted have sufficient energy to knock out secondary electrons from the sample. These electrons then undergo the same polarisation analysis as with SEMPA, allowing magnetic domains to be visualised. A resolution comparable to that of SEMPA, around 30nm, has been achieved.

In another related technique, the fact that electrons produced by field emission from a magnetic material can be made to have a well defined spin polarisation is used in an STM arrangement [Wiesendanger, *et al.*[135] (1990); Wiesendanger, *et al.*[136] (1991)]. A tip coated with a magnetic material is used to scan over a sample in the usual constant current mode. Electrons are more likely to tunnel into a magnetic domain of favourable orientation than an unfavourable one, so that the different domains appear as topographic features of different heights. The problem here is that the sample surface must be known well enough to be able to distinguish between real topographic features and magnetic contrast.

Another approach which has been followed by many groups is to utilise the high magnetic flux sensitivity of the superconducting quantum interference device (SQUID) in a scanning microscope arrangement. This technique has been named scanning SQUID microscopy or SSM. SSM has been used in a constant height mode of imaging (no distance regulation) [Mathai, Song, Gim and Wellstood[137] (1992)] as well as with some

atomic force regulation (by forming a SQUID on a silicon cantilever) [Kirtley, *et al.*[138] (1995)]. Without distance regulation, the images are affected by changes in separation due to tilt, as well as topography. However, most of the samples imaged are relatively flat on the scale of the separation between SQUID and sample used (microns). The large separation naturally results in a blurring of the features as the magnetic fields decay with separation. While these devices have very high flux sensitivity, their physical resolution is not high because the size of the pick-up loop is typically of the order of 10μm or more. SQUIDs have also been used without a pick-up loop but here the need to bias the SQUID with a magnetic field can magnetically affect the sample itself. In addition, SQUIDs need to be operated under cryogenic conditions which limits their use to imaging other superconductors. Flux vortices have been imaged successfully using this technique.

A variation of this technique for imaging non-magnetic samples has also been developed, based on Eddy current microscopy [Black, *et al.*[139] (1994)]. Here an alternating magnetic field generated behind a SQUID is used to generate eddy currents in a metallic structure which then generates its own opposing magnetic field. By analysing the interaction between the two fields (actively altering the applied field amplitude to maintain constant flux through the SQUID) an image can be formed. A d.c. magnetic field is used to bias the SQUID which is also regulated using a feedback loop (by making use of the non-linear response of the SQUID to magnetic flux) allowing the presence of magnetic particles to be detected and imaged concurrently. Again the physical resolution is not high and cryogenic environments are necessary. Yet another variant allows r.f. fields to be imaged directly, again with low resolution and at cryogenic temperatures [Black, *et al.*[140] (1995)].

One of the oldest methods of magnetic field measurement is that of the Hall-probe. Here the voltage generated by the Hall effect is measured. The Hall voltage appears in a direction perpendicular to the flow of current and magnetic field. It arises due to the action of the Lorentz force on moving charged carriers. Assuming that the current is travelling in the x direction, and the magnetic field is in the z direction where the sensor lies in the xy-plane, the force acting on the current carriers is given by:

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \Rightarrow F_y = qv_x B_z = qE_y, \tag{1.4.29}$$

where F_y is the Lorentz force acting on the carriers, v_x is the dominant velocity of the charged carriers, q is the charge on the carriers, B_z is the magnetic flux intensity acting on the sensor and E_y is the resulting electric field once equilibrium has been reached.

Assuming also that the sensor dimensions are l, b and t in the x, y and z directions respectively, the Hall voltage is given by:

$$V_y = bv_x B_z = b\frac{J_x}{nq}B_z = \frac{R_H I_x B_z}{t}, \tag{1.4.30}$$

where n is the number of charged carriers per unit volume, J_x is the current density given by:

$$J_x = nqv_x, \tag{1.4.31}$$

I_x is the current given by:

$$I_x = J_x bt, \tag{1.4.32}$$

and R_H is the Hall coefficient defined by:

$$R_H = \frac{1}{nq}. \tag{1.4.33}$$

Since the carrier velocity is proportional to the mobility according to:

$$v = \mu E, \tag{1.4.34}$$

where E is the applied electric field, and the current density (and current) are proportional to the carrier velocity, using a high mobility material means that fewer carriers are necessary to generate the same current. This in turn means that the Hall coefficient and hence the Hall effect is larger for such materials. Therefore, by using Hall-bars made from two dimensional electron gas (2DEG) materials, high mobility, low carrier concentration probes can be made [Chang, *et al.*[141] (1992); Oral, Bending and Henini[142] (1996)]. These are typically made by molecular beam epitaxy (MBE) with materials such as GaAs/AlGaAs. While they are not as sensitive as SQUIDs to magnetic fields, they are still very sensitive and can operate at room temperature.

Unfortunately, the materials used make it difficult to place the Hall-probe in a cantilever geometry in which AFM regulation can be used. Thus, most such scanning Hall-probe microscopes have used STM regulation by tunnelling from the corner of the substrate with the Hall sensor. This has meant that the sensor is typically at an angle with respect to the sample and several microns distant, leading to a loss of resolution. However, by making two separated Hall junctions, the sensor can be used in a gradiometer arrangement which is useful for determining the magnetic field lines through extrapolation. Resolutions down to 0.25μm have been obtained, limited primarily by the size of the active area of the sensor as a result of the large mean free paths of electrons in the sensor. These sensors also have the advantage of being completely non-invasive.

The most popular scanning probe technique that has emerged is the magnetic force microscope or MFM [Martin and Wickramasinghe[143] (1987); Saenz, *et al.*[144] (1987); Hobbs, Abraham and Wickramasinghe[145] (1989)]. This is based on the fact that the energy of, and force on, a magnetic dipole in a magnetic field are given by:

$$W = -\mathbf{m} \bullet \mathbf{B}, \mathbf{F} = \nabla(\mathbf{m} \bullet \mathbf{B}), \tag{1.4.35}$$

where $\mathbf{m}$ is the magnetic dipole moment, $\mathbf{B}$ is the magnetic flux intensity, W is the potential energy of the magnetic dipole and $\mathbf{F}$ is the force on the magnetic dipole. To obtain the total force, this must be integrated over both interacting volumes. It has been found that tips made from materials such as nickel can be magnetised so that the apex of the tip consists of a single domain [Rugar, *et al.*[146] (1990)]. If such a tip is used as part of an AFM, a magnetic force, given by:

$$\mathbf{F}_{magnetic} = \nabla \int_{sample} d\tau_s \int_{tip} d\tau_t \mathbf{m} \bullet \mathbf{B}, \tag{1.4.36}$$

will be experienced by the tip as it passes over a magnetic sample. However, the magnetic force is not separable from the ordinary atomic forces so that the topographic and magnetic features are combined in the image.

In order to tackle this, one option is to first scan the probe in hard contact when essentially the atomic forces, and hence topography, will dominate, and then scan the probe in constant height mode (because the sudden changes in the sign of the magnetic force due to the domain structure of the sample magnetisation means that it is difficult for a feedback loop to follow these changes) further from the sample where the longer range magnetic interaction will dominate over the weaker Van der Waal's forces. Unfortunately, as noted in section §1.3.3, the capillary forces present are very strong so that a large separation is required if the cantilever is not to be overcome by the attractive forces. This naturally leads to a loss of resolution due to the decay of the magnetic fields. The presence of any tilt between the sample and probe is an additional problem. An alternative for relatively flat samples is to scan the probe in constant force mode, but to keep the sign of the force fixed by applying a d.c. bias between sample and tip. This introduces an electrostatic force which can be adjusted to be larger than any changes in the magnetic field.

In order to increase the resolution of features in the magnetic image at the expense of resolution in the topographic image, it is normal to use non-contact mode. Since the easiest method for forming a magnetic force probe is to use a bent piece of wire made from a magnetic material, the spring constant is usually larger than desirable for contact mode in any case. Non-contact mode is sensitive to force gradients in the direction of oscillation so that an additional magnetic spring constant given by:

$$k_{magnetic}(z) = -\frac{d\mathbf{F}_{magnetic}}{dz}, \tag{1.4.37}$$

appears in the expression for the resonant frequency of the probe. In general, this will be inseparable from effects due to the normal atomic force gradients. In addition the sudden changes in sign can make it difficult for any feedback loops to operate in a stable manner. However, for flat samples, electrostatic forces can be used to obtain topographic information if it is assumed that there are no trapped charges or varying surface potentials on the sample. This has the additional advantage of stabilising the frequency variation so that a feedback loop can be made to operate.

Electrostatic forces can be used in two ways. The first method is to apply a d.c. voltage between tip and sample [*e.g.* Mamin, *et al.*[147] (1989)]. By taking two scans successively, one with the d.c. voltage adjusted so that the electrostatic force gradients dominate allows an approximation to the sample topography to be obtained, and the other with the voltage lowered so that the magnetic force gradient is comparable to the electrostatic force gradient allows magnetic forces to be mapped out. Alternatively, the second scan can be obtained in constant height mode with the potential difference set to zero and the

feedback loop switched off. The other method is to apply an a.c. voltage between the tip and sample and use the resulting signal at ω_{el} or $2\omega_{el}$ to simultaneously obtain an approximation to the topography and a magnetic force image.

Using these methods, resolutions down to 10nm have been obtained. A variant of this technique is magnetising the tip at an angle to its axis or oscillating the tip laterally so that sensitivity to in-plane magnetic fields (such as those at domain boundaries) is obtained [Abraham, Williams and Wickramasinghe[148] (1988)]. Another interesting point is the difference in the images obtained by using low coercivity and high coercivity tips [Abraham and McDonald[149] (1990); Grutter, *et al.*[150] (1990); Sueoka, Okuda, Matsubara and Sai[151] (1991); Barnes, Oshea and Welland[152] (1994)]. A high coercivity tip's magnetic polarisation will be unaffected by the polarisation of the sample so that different domains will appear with different intensities in the magnetic image. A sufficiently low coercivity tip will switch its magnetisation at every domain so that the force between the tip and sample remains constant, however, at domain boundaries the signal will be different as the polarisations are parallel to the plane of the sample. Thus the latter gives good domain boundary contrast. There are, however, obvious disadvantages to the MFM technique such as the difficulty in imaging samples which are not flat due to the mixing of the topographic and magnetic signals and the fact that the field of the tip, of the order of 10mT, can affect the domains of the sample being looked at. In particular, soft magnetic materials cannot be imaged with MFM. Furthermore, it is almost impossible to get quantitative measurements of magnetic field using MFM.

Another technique worthy of mention is the magnetic resonance force microscope or MRFM [Rugar, Yannoni and Sidles[153] (1992); Zuger and Rugar[154] (1993); Zuger and Rugar[155] (1994)]. This technique is meant to apply some aspects of AFM to NMR related measurement techniques. Here the sample is deposited onto a silicon cantilever (without tip). The lever is placed in a non-uniform magnetic field so that a magnetic field gradient is present across the sample. This field can be provided by a permanent magnet or by bringing some magnetised wire, as used for MFM, in close proximity to the cantilever. The magnetic field results in the splitting of degenerate levels in the sample in proportion to the applied field. As this field is not uniform, the areas of equal energy level splitting form parabolas which follow the lines of constant magnetic field. The magnetic moment of the atoms in the sample can be flipped between energy levels by the application of radio frequency excitation of energy equal to the energy difference (using Planck's constant).

For a given frequency, this magnetic resonance will only occur in a narrow parabolic slice of the sample. The flipping of the magnetic orientation results in the force acting on the sample altering. This force causes the cantilever to bend towards the magnetic source. However, this movement causes a different slice of the sample to occupy the magnetic resonance position. Eventually equilibrium is reached between the restoring cantilever force and the magnetic force. By oscillating the r.f. frequency at the resonant frequency of the cantilever, the same slice that was excited into resonance can be followed by the r.f.,

signal resulting in a force modulation at the cantilever resonant frequency. This causes the cantilever to resonate, amplifying the signal by its mechanical quality factor so that the resulting response can be easily detected. By scanning through the r.f. frequencies different slices can be probed. In addition the magnetic tip can be scanned across the sample laterally allowing a three dimensional picture of the magnetic moments in the sample to be detected. In theory, this method is sensitive to the spin of an individual electron, although the lack of magnetic screening around the magnetic source means that a large area is probed at one time [Rugar, *et al.*[156] (1994); Sidles, *et al.*[157] (1995)]. This technique has applications in biochemistry where parts of proteins and cells can be tagged with a paramagnetic molecule. Here the three-dimensional capability of the technique can be used to locate those tags even when they are embedded within a cell or protein where conventional probe techniques cannot reach.

1.5 CANTILEVER DEFLECTION DETECTION

As the previous sections have shown, SPM techniques are being used in many different applications and many different environments. One of the key elements in all SPM techniques is the detection of cantilever deflection. Originally tunnelling detection was used to measure deflections [Binnig, *et al.*[31] (1986)]. However, its dependence on the surface conditions and topography of the back of the cantilever in addition to the increased complexity of the system meant that optical techniques soon became the preferred option.

Initially all the optical techniques were based on homodyne interferometry [*e.g.* Denboef[158] (1989)] with which very small deflections can be measured. In order to eliminate the effects of drift in the optics, some groups resorted to heterodyne [Martin, *et al.*[143] (1987); Martin, *et al.*[34] (1987)] or common path interferometry [Schonenberger and Alvarado[159] (1989); Watanabe, Hane and Goto[160] (1992)] based on having two beams, a reference beam and a probe beam, both of which travel through as much of the optics as possible. These beams have different wavelengths, usually with the aid of an acousto-optic modulator (AOM), and are made to mix with each other (optical down-conversion) to provide a signal free of any drift in their common paths [See, Appel and Somekh[161] (1988); Offside, Somekh and See[162] (1989)].

However, the complexity of these schemes meant that some groups preferred to look for simpler schemes based on fibre interferometry [Rugar, *et al.*[163] (1988); Rugar, Mamin and Guethner[164] (1989); Nonnenmacher, Vaeziravani and Wickramasinghe[165] (1992)] or cavity feedback [Sarid, Iams, Weissenberger and Bell[166] (1988)], possibly at the expense of some sensitivity under certain circumstances [Putman, Degrooth, Vanhulst and Greve[167] (1992)]. For commercial applications an even simpler and more inexpensive scheme became popular, that of optical beam deflection. Here a laser is bounced off the rear of a cantilever at an angle and focused onto a position sensitive photodetector (PSP) or a split

photodiode [Meyer and Amer[168] (1988)]. Any deflections are amplified by this optical lever in the optical path length between cantilever and detector. This scheme has the advantage that it can be extended to measure bending of the cantilever perpendicular to the scan direction (either using a PSP or four quadrant photodiode) for applications in friction force microscopy [Alexander, *et al.*[169] (1989); Ascoli, *et al.*[170] (1994)].

However, in many environments such as cryogenic and ultra-high vacuum it can be awkward to use an optical detection system. For this reason, schemes which can somehow integrate the detection system with the cantilever are desirable. This led to the development of capacitive detection, where the lever is used as one plate of a capacitor and its capacitance measured with respect to a reference plate [Goddenhenrich, Lemke, Hartmann and Heiden[171] (1990); Neubauer, *et al.*[172] (1990); Brugger, Blanc, Renaud and Derooij[173] (1994)]; the development of piezoresistive cantilevers, where a piezoresistor fabricated on the cantilever alters its resistance upon being stretched (or when the cantilever bends) [Tortonese, Barrett and Quate[174] (1993)]; and, the development of piezoelectric cantilevers where a piezoelectric material deposited onto the cantilever generates a voltage upon being stretched (or when the cantilever bends) [Tansock and Williams[175] (1992); Itoh and Suga[176] (1994); Itoh and Suga[177] (1994)]. These schemes, whilst not being as sensitive as their optical counterparts, are sensitive enough for many applications. However, their real benefit comes from the simplification of the whole microscope system achievable. This is an important factor where the microscope has to be placed in a restricted environment such as in a cryostat or vacuum chamber. Unfortunately, most of these sensors can only be fabricated on simple AFM cantilevers and therefore do not help users of techniques such as SNOM and MFM. Sarid[178] (1991) compares the sensitivity of some of these schemes.

More novel methods have also been used such as one where a piezoelectric is used to oscillate the cantilever and a piezoresistor is used to simultaneously measure the deflection [Minne, Manalis and Quate[179] (1995); Manalis, Minne and Quate[180] (1996)], one where a piezoelectric is not only used to measure the deflection but also to oscillate the cantilever [Minne, Manalis, Atalar and Quate[181] (1996)] and one where a phase grating is formed by turning the cantilever into a series of interlaced fingers [Manalis, Minne, Atalar and Quate[182] (1996)]. However, these techniques have yet to be integrated with many of the other SPM techniques.

In the case of SNOM, as carried out with an optical fibre probe, which uses shear force microscopy to regulate the tip-sample separation, an interferometer positioned close to the end of the probe was originally required to measure the lateral oscillation of the fibre, [Betzig, *et al.*[73] (1992)]. As SNOM is an optical technique in which very low light levels (often requiring the use of a photomultiplier tube (PMT), cooled charge coupled device (CCD) or avalanche photodiode (APD)) are measured, the use of an optical detection scheme is extremely undesirable. For this reason another technique based on attaching the optical fibre to one arm of a quartz crystal tuning fork was developed [Karrai and

Grober[183] (1995)]. Quartz is a piezoelectric material so that the fibre's lateral oscillation generates an oscillating potential across the tuning fork which can then be used to supply an amplitude or phase signal.

One commercial company [Nanonics Ltd, Havaad Haleumi 21, Jerusalem, Israel] offers SNOM probes consisting of a bent optical fibre upon which a piece of aluminium foil has been attached, so that ordinary non-contact force microscopy can be used to regulate the probe-sample separation. In a similar manner, the etched magnetic wire tips commonly used in MFM are often bent to allow optical techniques to be used to supply the deflection signal.

1.6 CONCLUSIONS

It is clear from this chapter that many SPM techniques have not become mainstream because of the difficulty in mass-producing the probes required. Whilst piezoresistive and piezoelectric cantilevers are now commercially available for AFM, (and STM tips are easy to produce) other techniques have not been commercialised because of the difficulty in obtaining a reliable supply of probes rather than a lack of demand. The result is that many potentially valuable techniques are only available to the relatively few prepared to invest substantial amounts of time and skill making probes using techniques more akin to handicrafts than mass-production. Unfortunately these hand-crafted probes do not offer a consistent level of performance from probe to probe, even when they are produced by the same person sequentially in one sitting, and their manufacture is commonly regarded as a 'black art'.

Furthermore it is also true that the resolution achieved by many near-field probe techniques are limited more by the currently achievable dimensions of the sensor, the magnitude of the separation between specimen and probe and the quality of its regulation, rather than the physics of the sensor contrast mechanism. It is therefore clear that much is to be gained from tackling the manufacturing problem. As the following chapters will show, the work presented in this thesis attempts to tackle this very problem by using a particular approach bound by several real-world constraints and with a clear objective in mind.

1.7 REFERENCES

1. V. E. Cosslett, *THE DEVELOPMENT OF ELECTRON MICROSCOPY AND RELATED TECHNIQUES AT THE CAVENDISH LABORATORY, CAMBRIDGE, 1946-79 (PART 1 1946-60)*, Contemporary Physics **Vol. 22,** No. 1, pp. 3-36 (1981).

2. V. E. Cosslett, *THE DEVELOPMENT OF ELECTRON MICROSCOPY AND RELATED TECHNIQUES AT THE CAVENDISH LABORATORY, CAMBRIDGE, 1946-79 (PART 2 1961-79)*, Contemporary Physics **Vol. 22,** No. 2, pp. 147-182 (1981).

3. F. A. Jenkins and H. E. White, *Section 15.10: RESOLVING POWER OF A MICROSCOPE* in Fundamentals of Optics, 4th edition, published by McGraw-Hill, Singapore, pp. 332-334 (1981).

4. G. Binnig and H. Rohrer, *SCANNING TUNNELING MICROSCOPY*, Helvetica Physica Acta **Vol. 55,** No. 6, pp. 726-735 (1982).

5. G. Binnig, H. Rohrer, C. Gerber and E. Weibel, *TUNNELING THROUGH A CONTROLLABLE VACUUM GAP*, Applied Physics Letters **Vol. 40,** No. 2, pp. 178-180 (1982).

6. G. Binnig, H. Rohrer, C. Gerber and E. Weibel, *SURFACE STUDIES BY SCANNING TUNNELING MICROSCOPY*, Physical Review Letters **Vol. 49,** No. 1, pp. 57-61 (1982).

7. P. K. Hansma and J. Tersoff, *SCANNING TUNNELING MICROSCOPY*, Journal Of Applied Physics **Vol. 61,** No. 2, pp. R 1-R 23 (1987).

8. R. Wiesendanger, *Chapter 1: SCANNING TUNNELING MICROSCOPY (STM)* in Scanning Probe Microscopy and Spectroscopy, published by Cambridge University Press, Cambridge, pp. 11-209 (1994).

9. I. Giaever, Physical Review Letters **Vol. 5,** p. 147 (1960).

10. D. G. Walmsley, *PRE-MICROSCOPE TUNNELLING - INSPIRATION OR CONSTRAINT?*, Surface Science **Vol. 181,** pp. 1-26 (1987).

11. R. H. Fowler and L. Nordheim, Proceedings of the Royal Society of London A **Vol. 119,** p. 173 (1928).

12. J. Frenkel, Physical Review **Vol. 36,** p. 1604 (1930).

13. G. Binnig and D. P. E. Smith, *SINGLE-TUBE 3-DIMENSIONAL SCANNER FOR SCANNING TUNNELING MICROSCOPY*, Review Of Scientific Instruments **Vol. 57,** No. 8, pp. 1688-1689 (1986).

14. R. M. Tromp, R. J. Hamers and J. E. Demuth, *QUANTUM STATES AND ATOMIC-STRUCTURE OF SILICON SURFACES*, Science **Vol. 234,** No. 4774, pp. 304-309 (1986).

15. A. J. Melmed, *THE ART AND SCIENCE AND OTHER ASPECTS OF MAKING SHARP TIPS*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 2, pp. 601-608 (1991).

16. I. H. Musselman and P. E. Russell, *PLATINUM/IRIDIUM TIPS WITH CONTROLLED GEOMETRY FOR SCANNING TUNNELING MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 8,** No. 4, pp. 3558-3562 (1990).

17. J. Schneir, O. Marti, G. Remmers, D. Glaser, R. Sonnenfeld, B. Drake, P. K. Hansma and V. Elings, *SCANNING TUNNELING MICROSCOPY AND ATOMIC FORCE MICROSCOPY OF THE LIQUID-SOLID INTERFACE*, Journal Of Vacuum Science & Technology A-Vacuum Surfaces and Films **Vol. 6,** No. 2, pp. 283-286 (1988).

18. J. Jahanmir, B. G. Haggar and J. B. Hayes, *THE SCANNING PROBE MICROSCOPE*, Scanning Microscopy **Vol. 6,** No. 3, pp. 625-660 (1992).

19. R. M. Penner, *NANOMETER-SCALE SYNTHESIS AND ATOMIC-SCALE MODIFICATION WITH THE SCANNING TUNNELING MICROSCOPE*, Scanning Microscopy **Vol. 7,** No. 3, pp. 805-813 (1993).

20. R. Miranda, N. Garcia, A. M. Baro, R. Garcia, J. L. Pena and H. Rohrer, *TECHNOLOGICAL APPLICATIONS OF SCANNING TUNNELING MICROSCOPY AT ATMOSPHERIC-PRESSURE*, Applied Physics Letters **Vol. 47,** No. 4, pp. 367-369 (1985).

21. J. P. Carrejo, T. Thundat, L. A. Nagahara, S. M. Lindsay and A. Majumdar, *SCANNING TUNNELING MICROSCOPY INVESTIGATIONS OF POLYSILICON FILMS UNDER SOLUTION*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 2, pp. 955-959 (1991).

22. J. P. Song, K. A. Morch, K. Carneiro and A. R. Tholen, *INVESTIGATION OF SCANNING TUNNELING MICROSCOPY TUNNELING BARRIER SIGNALS IN AIR AND WATER*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 3, pp. 2237-2242 (1994).

23. J. K. Gimzewski, S. Modesti, T. David and R. R. Schlittler, *SCANNING-TUNNELING-MICROSCOPY OF ORDERED C-60 AND C-70 LAYERS ON AU(111), CU(111), AG(110), AND AU(110) SURFACES*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 3, pp. 1942-1946 (1994).

24. R. Gaisch, R. Berndt, W. D. Schneider, J. K. Gimzewski, B. Reihl, R. R. Schlittler and M. Tschudy, *INTERNAL STRUCTURE OF C(60) ON AU(110) AS OBSERVED BY LOW-TEMPERATURE SCANNING-TUNNELING-MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 3, pp. 2153-2155 (1994).

25. T. Ichinokawa, H. Itoh, A. Schmid, D. Winau and J. Kirschner, *RECONSTRUCTED STRUCTURES IN METAL/SI(100) SURFACES AT HIGH TEMPERATURE OBSERVED BY SCANNING TUNNELING MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 3, pp. 2070-2073 (1994).

26. A. M. Baro, R. Miranda and J. L. Carrascosa, *APPLICATION TO BIOLOGY OF THE SCANNING TUNNELING MICROSCOPE OPERATED IN AIR AT AMBIENT PRESSURE*, IBM Journal of Research and Development **Vol. 30,** No. 4, pp. 380-386 (1986).

27. H. P. Myers, *Chapter 14: ASPECTS OF SURFACE PHYSICS* in Introductory Solid State Physics, published by Taylor and Francis, London, pp. 477-500 (1990).

28. R. Wiesendanger, *CONTRIBUTIONS OF SCANNING PROBE MICROSCOPY AND SPECTROSCOPY TO THE INVESTIGATION AND FABRICATION OF NANOMETER-SCALE STRUCTURES*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 2, pp. 515-529 (1994).

29. R. M. Tromp, *SPECTROSCOPY WITH THE SCANNING TUNNELING MICROSCOPE - A CRITICAL- REVIEW*, Journal Of Physics-Condensed Matter **Vol. 1,** No. 51, pp. 10211-10228 (1989).

30. T. R. Albrecht, H. A. Mizes, J. Nogami, S. I. Park and C. F. Quate, *OBSERVATION OF TILT BOUNDARIES IN GRAPHITE BY SCANNING TUNNELING MICROSCOPY AND ASSOCIATED MULTIPLE TIP EFFECTS*, Applied Physics Letters **Vol. 52,** No. 5, pp. 362-364 (1988).

31. G. Binnig, C. F. Quate and C. Gerber, *ATOMIC FORCE MICROSCOPE*, Physical Review Letters **Vol. 56,** No. 9, pp. 930-933 (1986).

32. D. Sarid, *Chapter 13: ATOMIC FORCE MICROSCOPY* in Scanning Force Microscopy with Applications to Electric, Magnetic and Atomic Forces, published by Oxford University Press, New York, pp. 181-230 (1991).

33. S. A. C. Gould, K. Burke and P. K. Hansma, *SIMPLE THEORY FOR THE ATOMIC-FORCE MICROSCOPE WITH A COMPARISON OF THEORETICAL AND EXPERIMENTAL IMAGES OF GRAPHITE*, Physical Review B-Condensed Matter **Vol. 40,** No. 8, pp. 5363-5366 (1989).

34. Y. Martin, C. C. Williams and H. K. Wickramasinghe, *ATOMIC FORCE MICROSCOPE FORCE MAPPING AND PROFILING ON A SUB 100-A SCALE*, Journal Of Applied Physics **Vol. 61,** No. 10, pp. 4723-4729 (1987).

35. T. R. Albrecht, P. Grutter, D. Horne and D. Rugar, *FREQUENCY-MODULATION DETECTION USING HIGH-Q CANTILEVERS FOR ENHANCED FORCE MICROSCOPE SENSITIVITY*, Journal Of Applied Physics **Vol. 69,** No. 2, pp. 668-673 (1991).

36. D. Sarid, *Chapter 3: SOURCES OF NOISE* in Scanning Force Microscopy with Applications to Electric, Magnetic and Atomic Forces, published by Oxford University Press, New York, pp. 39-53 (1991).

37. P. K. Hansma, J. P. Cleveland, M. Radmacher, D. A. Walters, P. E. Hillner, M. Bezanilla, M. Fritz, D. Vie, H. G. Hansma, C. B. Prater, J. Massie, L. Fukunaga, J. Gurley and V. Elings, *TAPPING MODE ATOMIC-FORCE MICROSCOPY IN LIQUIDS*, Applied Physics Letters **Vol. 64,** No. 13, pp. 1738-1740 (1994).

38. C. A. J. Putman, K. O. Vanderwerf, B. G. Degrooth, N. F. Vanhulst and J. Greve, *TAPPING MODE ATOMIC-FORCE MICROSCOPY IN LIQUID*, Applied Physics Letters **Vol. 64,** No. 18, pp. 2454-2456 (1994).

39. F. Ohnesorge and G. Binnig, *TRUE ATOMIC-RESOLUTION BY ATOMIC FORCE MICROSCOPY THROUGH REPULSIVE AND ATTRACTIVE FORCES*, Science **Vol. 260,** No. 5113, pp. 1451-1456 (1993).

40. A. L. Weisenhorn, P. K. Hansma, T. R. Albrecht and C. F. Quate, *FORCES IN ATOMIC FORCE MICROSCOPY IN AIR AND WATER*, Applied Physics Letters **Vol. 54,** No. 26, pp. 2651-2653 (1989).

41. G. Binnig, C. Gerber, E. Stoll, T. R. Albrecht and C. F. Quate, *ATOMIC RESOLUTION WITH ATOMIC FORCE MICROSCOPE*, Surface Science **Vol. 189,** No. OCT, pp. 1-6 (1987).

42. T. R. Albrecht and C. F. Quate, *ATOMIC RESOLUTION WITH THE ATOMIC FORCE MICROSCOPE ON CONDUCTORS AND NONCONDUCTORS*, Journal Of Vacuum Science & Technology A-Vacuum Surfaces and Films **Vol. 6,** No. 2, pp. 271-274 (1988).

43. S. Akamine, R. C. Barrett and C. F. Quate, *IMPROVED ATOMIC FORCE MICROSCOPE IMAGES USING MICROCANTILEVERS WITH SHARP TIPS*, Applied Physics Letters **Vol. 57,** No. 3, pp. 316-318 (1990).

44. T. R. Albrecht, S. Akamine, T. E. Carver and C. F. Quate, *MICROFABRICATION OF CANTILEVER STYLI FOR THE ATOMIC FORCE MICROSCOPE*, Journal Of Vacuum Science & Technology A-Vacuum Surfaces and Films **Vol. 8,** No. 4, pp. 3386-3396 (1990).

45. R. Sonnenfeld and P. K. Hansma, *ATOMIC-RESOLUTION MICROSCOPY IN WATER*, Science **Vol. 232,** No. 4747, pp. 211-213 (1986).

46. J. E. Griffith and D. A. Grigg, *DIMENSIONAL METROLOGY WITH SCANNING PROBE MICROSCOPES*, Journal Of Applied Physics **Vol. 74,** No. 9, pp. R 83-R 109 (1993).

47. J. E. Griffith, H. M. Marchman, G. L. Miller and L. C. Hopkins, *DIMENSIONAL METROLOGY WITH SCANNING PROBE MICROSCOPES*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 3, pp. 1100-1105 (1995).

48. Y. Martin and H. K. Wickramasinghe, *TOWARD ACCURATE METROLOGY WITH SCANNING FORCE MICROSCOPES*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 2335-2339 (1995).

49. R. M. Feenstra, M. A. Lutz, F. Stern, K. Ismail, P. M. Mooney, F. K. Legoues, C. Stanis, J. O. Chu and B. S. Meyerson, *ROUGHNESS ANALYSIS OF SI/SIGE HETEROSTRUCTURES*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 4, pp. 1608-1612 (1995).

50. T. R. Albrecht and C. F. Quate, *ATOMIC RESOLUTION IMAGING OF A NONCONDUCTOR BY ATOMIC FORCE MICROSCOPY*, Journal Of Applied Physics **Vol. 62,** No. 7, pp. 2599-2602 (1987).

51. S. Gould, O. Marti, B. Drake, L. Hellemans, C. E. Bracker, P. K. Hansma, N. L. Keder, M. M. Eddy and G. D. Stucky, *MOLECULAR RESOLUTION IMAGES OF AMINO-ACID CRYSTALS WITH THE ATOMIC FORCE MICROSCOPE*, Nature **Vol. 332,** No. 6162, pp. 332-334 (1988).

52. W. A. Rees, R. W. Keller, J. P. Vesenka, G. L. Yang and C. Bustamante, *EVIDENCE OF DNA BENDING IN TRANSCRIPTION COMPLEXES IMAGED BY SCANNING FORCE MICROSCOPY*, Science **Vol. 260,** No. 5114, pp. 1646-1649 (1993).

53. D. A. Erie, G. L. Yang, H. C. Schultz and C. Bustamante, *DNA BENDING BY CRO PROTEIN IN SPECIFIC AND NONSPECIFIC COMPLEXES - IMPLICATIONS FOR PROTEIN SITE RECOGNITION AND SPECIFICITY*, Science **Vol. 266,** No. 5190, pp. 1562-1566 (1994).

54. M. Radmacher, M. Fritz, H. G. Hansma and P. K. Hansma, *DIRECT OBSERVATION OF ENZYME-ACTIVITY WITH THE ATOMIC-FORCE MICROSCOPE*, Science **Vol. 265,** No. 5178, pp. 1577-1579 (1994).

55. M. Radmacher, R. W. Tillmann, M. Fritz and H. E. Gaub, *FROM MOLECULES TO CELLS - IMAGING SOFT SAMPLES WITH THE ATOMIC FORCE MICROSCOPE*, Science **Vol. 257,** No. 5078, pp. 1900-1905 (1992).

56. E. Henderson, *IMAGING OF LIVING CELLS BY ATOMIC-FORCE MICROSCOPY*, Progress In Surface Science **Vol. 46,** No. 1, pp. 39-60 (1994).

57. B. Drake, C. B. Prater, A. L. Weisenhorn, S. A. C. Gould, T. R. Albrecht, C. F. Quate, D. S. Cannell, H. G. Hansma and P. K. Hansma, *IMAGING CRYSTALS, POLYMERS, AND PROCESSES IN WATER WITH THE ATOMIC FORCE MICROSCOPE*, Science **Vol. 243,** No. 4898, pp. 1586-1589 (1989).

58. W. Haberle, J. K. H. Horber and G. Binnig, *FORCE MICROSCOPY ON LIVING CELLS*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 2, pp. 1210-1213 (1991).

59. A. Correia, N. Garcia, J. Massanell and J. L. Costakramer, *STRUCTURAL STUDY OF NANOCRYSTALLIZED CU83CO17 RIBBONS BY SCANNING FRICTION FORCE MICROSCOPY*, Applied Physics Letters **Vol. 68,** No. 3, pp. 340-342 (1996).

60. B. Bhushan, J. N. Israelachvili and U. Landman, *NANOTRIBOLOGY - FRICTION, WEAR AND LUBRICATION AT THE ATOMIC-SCALE*, Nature **Vol. 374,** No. 6523, pp. 607-616 (1995).

61. S. J. Oshea, M. E. Welland and T. Rayment, *ATOMIC FORCE MICROSCOPE STUDY OF BOUNDARY-LAYER LUBRICATION*, Applied Physics Letters **Vol. 61,** No. 18, pp. 2240-2242 (1992).

62. S. Manne and H. E. Gaub, *MOLECULAR-ORGANIZATION OF SURFACTANTS AT SOLID-LIQUID INTERFACES*, Science **Vol. 270,** No. 5241, pp. 1480-1482 (1995).

63. O. Marti, H. O. Ribi, B. Drake, T. R. Albrecht, C. F. Quate and P. K. Hansma, *ATOMIC FORCE MICROSCOPY OF AN ORGANIC MONOLAYER*, Science **Vol. 239,** No. 4835, pp. 50-52 (1988).

64. E. Meyer, L. Howald, R. M. Overney, H. Heinzelmann, J. Frommer, H. J. Guntherodt, T. Wagner, H. Schier and S. Roth, *MOLECULAR-RESOLUTION IMAGES OF LANGMUIR-BLODGETT-FILMS USING ATOMIC FORCE MICROSCOPY*, Nature **Vol. 349,** No. 6308, pp. 398-400 (1991).

65. L. F. Chi, M. Anders, H. Fuchs, R. R. Johnston and H. Ringsdorf, *DOMAIN-STRUCTURES IN LANGMUIR-BLODGETT-FILMS INVESTIGATED BY ATOMIC FORCE MICROSCOPY*, Science **Vol. 259,** No. 5092, pp. 213-216 (1993).

66. E. L. Florin, V. T. Moy and H. E. Gaub, *ADHESION FORCES BETWEEN INDIVIDUAL LIGAND-RECEPTOR PAIRS*, Science **Vol. 264,** No. 5157, pp. 415-417 (1994).

67. V. T. Moy, E. L. Florin and H. E. Gaub, *INTERMOLECULAR FORCES AND ENERGIES BETWEEN LIGANDS AND RECEPTORS*, Science **Vol. 266,** No. 5183, pp. 257-259 (1994).

68. E. H. Synghe, *SUGGESTED METHOD FOR EXTENDING MICROSCOPIC RESOLUTION INTO THE ULTRA-MICROSCOPIC REGION*, Philosophical Magazine **Vol. 6,** pp. 356-362 (1928).

69. E. A. Ash and G. Nicholls, *SUPER-RESOLUTION APERTURE MICROSCOPE*, Nature **Vol. 237,** pp. 510-512 (1972).

70. R. C. Reddick, R. J. Warmack and T. L. Ferrell, *NEW FORM OF SCANNING OPTICAL MICROSCOPY*, Physical Review B-Condensed Matter **Vol. 39,** No. 1, pp. 767-770 (1989).

71. R. C. Reddick, R. J. Warmack, D. W. Chilcott, S. L. Sharp and T. L. Ferrell, *PHOTON SCANNING TUNNELING MICROSCOPY*, Review Of Scientific Instruments **Vol. 61,** No. 12, pp. 3669-3677 (1990).

72. T. L. Ferrell, J. P. Goundonnet, R. C. Reddick, S. L. Sharp and R. J. Warmack, *THE PHOTON SCANNING TUNNELING MICROSCOPE*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 2, pp. 525-530 (1991).

73. E. Betzig, P. L. Finn and J. S. Weiner, *COMBINED SHEAR FORCE AND NEAR-FIELD SCANNING OPTICAL MICROSCOPY*, Applied Physics Letters **Vol. 60,** No. 20, pp. 2484-2486 (1992).

74. U. Durig, D. W. Pohl and F. Rohner, *NEAR-FIELD OPTICAL-SCANNING MICROSCOPY*, Journal Of Applied Physics **Vol. 59,** No. 10, pp. 3318-3327 (1986).

75. E. Betzig, M. Isaacson, H. Barshatzky, A. Lewis and K. Lin, *SUPER-RESOLUTION IMAGING WITH NEAR-FIELD SCANNING OPTICAL MICROSCOPY (NSOM)*, Ultramicroscopy **Vol. 25,** No. 2, pp. 155-163 (1988).

76. E. Betzig and J. K. Trautman, *NEAR-FIELD OPTICS - MICROSCOPY, SPECTROSCOPY, AND SURFACE MODIFICATION BEYOND THE DIFFRACTION LIMIT*, Science **Vol. 257,** No. 5067, pp. 189-195 (1992).

77. S. Shalom, K. Lieberman, A. Lewis and S. R. Cohen, *A MICROPIPETTE FORCE PROBE SUITABLE FOR NEAR-FIELD SCANNING OPTICAL MICROSCOPY*, Review Of Scientific Instruments **Vol. 63,** No. 9, pp. 4061-4065 (1992).

78. E. Betzig, S. G. Grubb, R. J. Chichester, D. J. Digiovanni and J. S. Weiner, *FIBER LASER PROBE FOR NEAR-FIELD SCANNING OPTICAL MICROSCOPY*, Applied Physics Letters **Vol. 63,** No. 26, pp. 3550-3552 (1993).

79. E. Betzig, J. K. Trautman, T. D. Harris, J. S. Weiner and R. L. Kostelak, *BREAKING THE DIFFRACTION BARRIER - OPTICAL MICROSCOPY ON A NANOMETRIC SCALE*, Science **Vol. 251,** No. 5000, pp. 1468-1470 (1991).

80. Y. Inouye and S. Kawata, *NEAR-FIELD SCANNING OPTICAL MICROSCOPE WITH A METALLIC PROBE TIP*, Optics Letters **Vol. 19,** No. 3, pp. 159-161 (1994).

81. F. Zenhausern, M. P. Oboyle and H. K. Wickramasinghe, *APERTURELESS NEAR-FIELD OPTICAL MICROSCOPE*, Applied Physics Letters **Vol. 65,** No. 13, pp. 1623-1625 (1994).

82. F. Zenhausern, Y. Martin and H. K. Wickramasinghe, *SCANNING INTERFEROMETRIC APERTURELESS MICROSCOPY - OPTICAL IMAGING AT 10 ANGSTROM RESOLUTION*, Science **Vol. 269,** No. 5227, pp. 1083-1085 (1995).

83. R. C. Davis, C. C. Williams and P. Neuzil, *OPTICAL-INTENSITY MAPPING ON THE NANOMETER-SCALE BY NEAR-FIELD PHOTODETECTION OPTICAL MICROSCOPY*, Optics Letters **Vol. 21,** No. 7, pp. 447-449 (1996).

84. H. Yamada, H. Tokumoto, S. Akamine, K. Fukuzawa and H. Kuwano, *IMAGING OF ORGANIC MOLECULAR FILMS USING A SCANNING NEAR-FIELD OPTICAL MICROSCOPE COMBINED WITH AN ATOMIC-FORCE MICROSCOPE*, Journal Of Vacuum Science & Technology B **Vol. 14,** No. 2, pp. 812-815 (1996).

85. A. Lewis, M. Isaacson, A. Harootunian and A. Muray, *DEVELOPMENT OF A 500-A SPATIAL-RESOLUTION LIGHT-MICROSCOPE .1. LIGHT IS EFFICIENTLY TRANSMITTED THROUGH GAMMA-16 DIAMETER APERTURES*, Ultramicroscopy **Vol. 13,** No. 3, pp. 227-231 (1984).

86. M. Tortonese, H. Yamada and C. F. Quate, *MICROFABRICATED PROBE FOR A NEAR FIELD SCANNING MICROSCOPE USING A FORCE DETECTION SCHEME*. presented at Transducers '91 - 6th International Conference on Solid-State Sensors and Actuators, San Francisco, California, USA (1991).

87. K. Lieberman, S. Harush, A. Lewis and R. Kopelman, *A LIGHT-SOURCE SMALLER THAN THE OPTICAL WAVELENGTH*, Science **Vol. 247,** No. 4938, pp. 59-61 (1990).

88. A. Lewis and K. Lieberman, *NEAR-FIELD OPTICAL IMAGING WITH A NON-EVANESCENTLY EXCITED HIGH- BRIGHTNESS LIGHT-SOURCE OF SUBWAVELENGTH DIMENSIONS*, Nature **Vol. 354,** No. 6350, pp. 214-216 (1991).

89. K. Lieberman and A. Lewis, *SUPERRESOLUTION OPTICAL IMAGING WITH A HIGH-BRIGHTNESS SUBWAVELENGTH LIGHT-SOURCE*, Ultramicroscopy **Vol. 42,** No. PtA, pp. 399-407 (1992).

90. G. Gottlich, W. Lang and W. M. Heckl, *A NOVEL PROBE FOR NEAR FIELD OPTICAL MICROSCOPY BASED ON LUMINESCENT SILICON*. presented at STM '95 - 8th International Conference on Scanning Tunneling Microscopy/Spectroscopy and Related Techniques, Snowmass Village, Colorado, USA (1995).

91. J. K. Trautman, E. Betzig, J. S. Weiner, D. J. Digiovanni, T. D. Harris, F. Hellman and E. M. Gyorgy, *IMAGE-CONTRAST IN NEAR-FIELD OPTICS*, Journal Of Applied Physics **Vol. 71,** No. 10, pp. 4659-4663 (1992).

92. E. Betzig, J. K. Trautman, R. Wolfe, E. M. Gyorgy, P. L. Finn, M. H. Kryder and C. H. Chang, *NEAR-FIELD MAGNETOOPTICS AND HIGH-DENSITY DATA-STORAGE*, Applied Physics Letters **Vol. 61,** No. 2, pp. 142-144 (1992).

93. R. E. Betzig, J. K. Trautman, R. Wolfe, E. M. Gyorgy, P. L. Finn, M. H. Kryder and C. H. Chang, *HIGH-DENSITY NEAR-FIELD OPTICAL-RECORDING*, Journal Of Applied Physics **Vol. 73,** No. 10 Pt2A, pp. 5791-5791 (1993).

94. B. D. Terris, H. J. Mamin, D. Rugar, W. R. Studenmund and G. S. Kino, *NEAR-FIELD OPTICAL-DATA STORAGE USING A SOLID IMMERSION LENS*, Applied Physics Letters **Vol. 65,** No. 4, pp. 388-390 (1994).

95. E. Betzig and R. J. Chichester, *SINGLE MOLECULES OBSERVED BY NEAR-FIELD SCANNING OPTICAL MICROSCOPY*, Science **Vol. 262,** No. 5138, pp. 1422-1425 (1993).

96. J. K. Trautman, J. J. Macklin, L. E. Brus and E. Betzig, *NEAR-FIELD SPECTROSCOPY OF SINGLE MOLECULES AT ROOM-TEMPERATURE*, Nature **Vol. 369,** No. 6475, pp. 40-42 (1994).

97. O. Hollricher, I. Hoersch, R. Kusche, O. Kirschenhofer and O. Marti, *INVESTIGATIONS OF SEMICONDUCTOR SURFACES WITH A NEAR FIELD OPTICAL MICROSCOPE*. presented at STM '95 - 8th International Conference on Scanning Tunneling Microscopy/Spectroscopy and Related Techniques, Snowmass Village, Colorado, USA (1995).

98. L. A. Nagahara and H. Tokumoto, *SCANNING NEAR-FIELD OPTICAL MICROSCOPY/SPECTROSCOPY OF THIN ORGANIC FILMS*, Journal Of Vacuum Science & Technology B **Vol. 14,** No. 2, pp. 800-803 (1996).

99. R. D. Grober, T. D. Harris, J. K. Trautman, E. Betzig, W. Wegscheider, L. Pfeiffer and K. West, *OPTICAL SPECTROSCOPY OF A GAAS/ALGAAS QUANTUM-WIRE STRUCTURE USING NEAR-FIELD SCANNING OPTICAL MICROSCOPY*, Applied Physics Letters **Vol. 64,** No. 11, pp. 1421-1423 (1994).

100. H. F. Hess, E. Betzig, T. D. Harris, L. N. Pfeiffer and K. W. West, *NEAR-FIELD SPECTROSCOPY OF THE QUANTUM CONSTITUENTS OF A LUMINESCENT SYSTEM*, Science **Vol. 264,** No. 5166, pp. 1740-1745 (1994).

101. H. M. Barlow, *Chapter 8: MICRO-WAVES AND WAVE-GUIDES* in Micro-wave Technique: Measurements and Applications, published by Constable and Company, London, pp. 97-113 (1947).

102. B. I. Bleaney and B. Bleaney, *Chapter 8: ELECTROMAGNETIC WAVES* in Electricity and Magnetism, 3rd edition, published by Oxford University Press, Oxford, pp. 225-257 (1976).

103. J. R. Reich, P. F. Ordung, H. L. Krauss and J. G. Skalnik, *Chapter 5: MICROWAVE THEORY AND TECHNIQUES* in Waveguides, published by Macmillan & Company, London, pp. 217-320 (1953).

104. R. A. Waldron, *Chapter IV: THEORY OF GUIDED ELECTROMAGNETIC WAVES* in Elementary Theory of Waveguides, published by Van Nostrand Reinhold, London, pp. 204-266 (1970).

105. P. F. Combes, *Chapter 16: MICROWAVE TRANSMISSION FOR TELECOMMUNICATIONS* in Metallic and Dielectric Waveguides, published by John Wiley & Sons, Chichester, pp. 243-263 (1991).

106. J. H. Weaver and H. P. R. Frederikse, *Section 12: OPTICAL PROPERTIES OF METALS AND SEMICONDUCTORS* in Handbook of Chemistry and Physics, 73rd edition, published by Chemical Rubber Publishing Co, Cleveland, Ohio, USA, pp. 109-132 (1996).

107. K. L. Saenger and J. Gupta, *LASER INTERFEROMETRIC THERMOMETRY FOR SUBSTRATE TEMPERATURE MEASUREMENT*, Applied Optics **Vol. 30,** No. 10, p. 1221 (1991).

108. W. B. Jackson, N. M. Amer, A. C. Boccara and D. Fournier, *PHOTOTHERMAL DEFLECTION SPECTROSCOPY AND DETECTION*, Applied Optics **Vol. 20,** No. 8, pp. 1333-1344 (1981).

109. D. Guidotti and H. M. Vandriel, *SPATIALLY RESOLVED DEFECT MAPPING IN SEMICONDUCTORS USING LASER-MODULATED THERMOREFLECTANCE*, Applied Physics Letters **Vol. 47,** No. 12, pp. 1336-1338 (1985).

110. B. Wunderlich and U. Gaur, *DIFFERENTIAL SCANNING CALORIMETRY OF FLEXIBLE, LINEAR MACROMOLECULES*, Advances In Chemistry Series No. 203, pp. 195-208 (1983).

111. D. A. C. Compton, D. J. Johnson and J. R. Powell, *A COMBINED DIFFERENTIAL SCANNING CALORIMETRY FOURIER-TRANSFORM INFRARED APPROACH FOR THE STUDY OF POLYMERIC MATERIALS*, Advances In Chemistry Series No. 236, pp. 661-675 (1993).

112. D. J. Johnson, P. J. Stout, S. L. Hill and K. Krishnan, *DIFFERENTIAL SCANNING CALORIMETRY FOURIER-TRANSFORM IR SPECTROSCOPY AND THERMOGRAVIMETRIC ANALYSIS FOURIER-TRANSFORM IR SPECTROSCOPY TO DIFFERENTIATE BETWEEN VERY SIMILAR POLYMER MATERIALS*, Acs Symposium Series **Vol. 581,** pp. 149-161 (1994).

113. C. C. Williams and H. K. Wickramasinghe, *SCANNING THERMAL PROFILER*, Applied Physics Letters **Vol. 49,** No. 23, pp. 1587-1589 (1986).

114. J. M. R. Weaver, L. M. Walpita and H. K. Wickramasinghe, *OPTICAL-ABSORPTION MICROSCOPY AND SPECTROSCOPY WITH NANOMETER RESOLUTION*, Nature **Vol. 342,** No. 6251, pp. 783-785 (1989).

115. E. Oesterschulze, M. Stopka, L. Ackermann, W. Scholz and S. Werner, *THERMAL IMAGING OF THIN FILMS BY SCANNING THERMAL MICROSCOPE*, Journal Of Vacuum Science & Technology **Vol. 14,** No. 2, pp. 832-837 (1996).

116. M. Nonnenmacher, M. P. Oboyle and H. K. Wickramasinghe, *KELVIN PROBE FORCE MICROSCOPY*, Applied Physics Letters **Vol. 58,** No. 25, pp. 2921-2923 (1991).

117. A. Majumdar, J. P. Carrejo and J. Lai, *THERMAL IMAGING USING THE ATOMIC FORCE MICROSCOPE*, Applied Physics Letters **Vol. 62,** No. 20, pp. 2501-2503 (1993).

118. J. Lai, M. Chandrachood, A. Majumdar and J. P. Carrejo, *THERMAL DETECTION OF DEVICE FAILURE BY ATOMIC-FORCE MICROSCOPY*, IEEE Electron Device Letters **Vol. 16,** No. 7, pp. 312-315 (1995).

119. J. R. Barnes, R. J. Stephenson, M. E. Welland, C. Gerber and J. K. Gimzewski, *PHOTOTHERMAL SPECTROSCOPY WITH FEMTOJOULE SENSITIVITY USING A MICROMECHANICAL DEVICE*, Nature **Vol. 372,** No. 6501, pp. 79-81 (1994).

120. J. R. Barnes, R. J. Stephenson, C. N. Woodburn, S. J. Oshea, M. E. Welland, T. Rayment, J. K. Gimzewski and C. Gerber, *A FEMTOJOULE CALORIMETER USING MICROMECHANICAL SENSORS*, Review Of Scientific Instruments **Vol. 65,** No. 12, pp. 3793-3798 (1994).

121. O. Nakabeppu, M. Chandrachood, Y. Wu, J. Lai and A. Majumdar, *SCANNING THERMAL IMAGING MICROSCOPY USING COMPOSITE CANTILEVER PROBES*, Applied Physics Letters **Vol. 66,** No. 6, pp. 694-696 (1995).

122. A. Majumdar, J. Lai, M. Chandrachood, O. Nakabeppu, Y. Wu and Z. Shi, *THERMAL IMAGING BY ATOMIC-FORCE MICROSCOPY USING THERMOCOUPLE CANTILEVER PROBES*, Review Of Scientific Instruments **Vol. 66,** No. 6, pp. 3584-3592 (1995).

123. Y. Martin, D. W. Abraham and H. K. Wickramasinghe, *HIGH-RESOLUTION CAPACITANCE MEASUREMENT AND POTENTIOMETRY BY FORCE MICROSCOPY*, Applied Physics Letters **Vol. 52,** No. 13, pp. 1103-1105 (1988).

124. C. C. Williams, W. P. Hough and S. A. Rishton, *SCANNING CAPACITANCE MICROSCOPY ON A 25 NM SCALE*, Applied Physics Letters **Vol. 55,** No. 2, pp. 203-205 (1989).

125. C. C. Williams, J. Slinkman, W. P. Hough and H. K. Wickramasinghe, *LATERAL DOPANT PROFILING ON A 100 NM SCALE BY SCANNING CAPACITANCE MICROSCOPY*, Journal Of Vacuum Science & Technology A-Vacuum Surfaces and Films **Vol. 8,** No. 2, pp. 895-898 (1990).

126. D. W. Abraham, C. Williams, J. Slinkman and H. K. Wickramasinghe, *LATERAL DOPANT PROFILING IN SEMICONDUCTORS BY FORCE MICROSCOPY USING CAPACITIVE*

DETECTION, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 2, pp. 703-706 (1991).

127. J. M. R. Weaver and D. W. Abraham, *HIGH-RESOLUTION ATOMIC FORCE MICROSCOPY POTENTIOMETRY*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 3, pp. 1559-1561 (1991).

128. J. M. R. Weaver and H. K. Wickramasinghe, *SEMICONDUCTOR CHARACTERIZATION BY SCANNING FORCE MICROSCOPE SURFACE PHOTOVOLTAGE MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 3, pp. 1562-1565 (1991).

129. K. Domansky, Y. Leng, C. C. Williams, J. Janata and D. Petelenz, *MAPPING OF MOBILE CHARGES ON INSULATOR SURFACES WITH THE ELECTROSTATIC FORCE MICROSCOPE*, Applied Physics Letters **Vol. 63,** No. 11, pp. 1513-1515 (1993).

130. Y. Huang and C. C. Williams, *CAPACITANCE-VOLTAGE MEASUREMENT AND MODELING ON A NANOMETER-SCALE BY SCANNING C-V MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 1, pp. 369-372 (1994).

131. S. Cunningham and J. M. R. Weaver, *CONTACT CHARGING DYNAMICS AND SURFACE-CHARGE BEHAVIOR AT SUBMICRON SPATIAL-RESOLUTION USING SCANNING FORCE MICROSCOPY*, Institute Of Physics Conference Series **Vol. 143,** pp. 109-112 (1995).

132. M. R. Scheinfein, J. Unguris, M. H. Kelley, D. T. Pierce and R. J. Celotta, *SCANNING ELECTRON-MICROSCOPY WITH POLARIZATION ANALYSIS (SEMPA)*, Review Of Scientific Instruments **Vol. 61,** No. 10, pp. 2501-2526 (1990).

133. S. A. Crooker, D. D. Awschalom and N. Samarth, *TIME-RESOLVED FARADAY ROTATION SPECTROSCOPY OF SPIN DYNAMICS IN DIGITAL MAGNETIC HETEROSTRUCTURES*, IEEE Journal Of Selected Topics In Quantum Electronics **Vol. 1,** No. 4, pp. 1082-1092 (1995).

134. B. G. Frost, N. F. Vanhulst, E. Lunedei, G. Matteucci and E. Rikkers, *STUDY OF THE LEAKAGE FIELD OF MAGNETIC FORCE MICROSCOPY THIN-FILM TIPS USING ELECTRON HOLOGRAPHY*, Applied Physics Letters **Vol. 68,** No. 13, pp. 1865-1867 (1996).

135. R. Wiesendanger, H. J. Guntherodt, G. Guntherodt, R. J. Gambino and R. Ruf, *OBSERVATION OF VACUUM TUNNELING OF SPIN-POLARIZED ELECTRONS WITH THE SCANNING TUNNELING MICROSCOPE*, Physical Review Letters **Vol. 65,** No. 2, pp. 247-250 (1990).

136. R. Wiesendanger, D. Burgler, G. Tarrach, A. Wadas, D. Brodbeck, H. J. Guntherodt, G. Guntherodt, R. J. Gambino and R. Ruf, *VACUUM TUNNELING OF SPIN-POLARIZED ELECTRONS DETECTED BY SCANNING TUNNELING MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 2, pp. 519-524 (1991).

137. A. Mathai, D. Song, Y. Gim and F. C. Wellstood, *ONE-DIMENSIONAL MAGNETIC-FLUX MICROSCOPE BASED ON THE DC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE*, Applied Physics Letters **Vol. 61,** No. 5, pp. 598-600 (1992).

138. J. R. Kirtley, M. B. Ketchen, K. G. Stawiasz, J. Z. Sun, W. J. Gallagher, S. H. Blanton and S. J. Wind, *HIGH-RESOLUTION SCANNING SQUID MICROSCOPE*, Applied Physics Letters **Vol. 66,** No. 9, pp. 1138-1140 (1995).

139. R. C. Black, F. C. Wellstood, E. Dantsker, A. H. Miklich, J. J. Kingston, D. T. Nemeth and J. Clarke, *EDDY-CURRENT MICROSCOPY USING A 77-K SUPERCONDUCTING SENSOR*, Applied Physics Letters **Vol. 64,** No. 1, pp. 100-102 (1994).

140. R. C. Black, F. C. Wellstood, E. Dantsker, A. H. Miklich, D. Koelle, F. Ludwig and J. Clarke, *IMAGING RADIOFREQUENCY FIELDS USING A SCANNING SQUID MICROSCOPE*, Applied Physics Letters **Vol. 66,** No. 10, pp. 1267-1269 (1995).

141. A. M. Chang, H. D. Hallen, L. Harriott, H. F. Hess, H. L. Kao, J. Kwo, R. E. Miller, R. Wolfe, J. Vanderziel and T. Y. Chang, *SCANNING HALL PROBE MICROSCOPY*, Applied Physics Letters **Vol. 61,** No. 16, pp. 1974-1976 (1992).

142. A. Oral, S. J. Bending and M. Henini, *SCANNING HALL PROBE MICROSCOPY OF SUPERCONDUCTORS AND MAGNETIC MATERIALS*, Journal Of Vacuum Science & Technology B **Vol. 14,** No. 2, pp. 1202-1205 (1996).

143. Y. Martin and H. K. Wickramasinghe, *MAGNETIC IMAGING BY FORCE MICROSCOPY WITH 1000-A RESOLUTION*, Applied Physics Letters **Vol. 50,** No. 20, pp. 1455-1457 (1987).

144. J. J. Saenz, N. Garcia, P. Grutter, E. Meyer, H. Heinzelmann, R. Wiesendanger, L. Rosenthaler, H. R. Hidber and H. J. Guntherodt, *OBSERVATION OF MAGNETIC FORCES BY THE ATOMIC FORCE MICROSCOPE*, Journal Of Applied Physics **Vol. 62,** No. 10, pp. 4293-4295 (1987).

145. P. C. D. Hobbs, D. W. Abraham and H. K. Wickramasinghe, *MAGNETIC FORCE MICROSCOPY WITH 25 NM RESOLUTION*, Applied Physics Letters **Vol. 55,** No. 22, pp. 2357-2359 (1989).

146. D. Rugar, H. J. Mamin, P. Guethner, S. E. Lambert, J. E. Stern, I. McFadyen and T. Yogi, *MAGNETIC FORCE MICROSCOPY - GENERAL-PRINCIPLES AND APPLICATION TO LONGITUDINAL RECORDING MEDIA*, Journal Of Applied Physics **Vol. 68,** No. 3, pp. 1169-1183 (1990).

147. H. J. Mamin, D. Rugar, J. E. Stern, R. E. Fontana and P. Kasiraj, *MAGNETIC FORCE MICROSCOPY OF THIN PERMALLOY-FILMS*, Applied Physics Letters **Vol. 55,** No. 3, pp. 318-320 (1989).

148. D. W. Abraham, C. C. Williams and H. K. Wickramasinghe, *MEASUREMENT OF INPLANE MAGNETIZATION BY FORCE MICROSCOPY*, Applied Physics Letters **Vol. 53,** No. 15, pp. 1446-1448 (1988).

149. D. W. Abraham and F. A. McDonald, *THEORY OF MAGNETIC FORCE MICROSCOPE IMAGES*, Applied Physics Letters **Vol. 56,** No. 12, pp. 1181-1183 (1990).

150. P. Grutter, D. Rugar, H. J. Mamin, G. Castillo, S. E. Lambert, C. J. Lin, R. M. Valletta, O. Wolter, T. Bayer and J. Greschner, *BATCH FABRICATED SENSORS FOR MAGNETIC FORCE MICROSCOPY*, Applied Physics Letters **Vol. 57,** No. 17, pp. 1820-1822 (1990).

151. K. Sueoka, K. Okuda, N. Matsubara and F. Sai, *STUDY OF TIP MAGNETIZATION BEHAVIOUR IN MAGNETIC FORCE MICROSCOPE*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 2, pp. 1313-1317 (1991).

152. J. R. Barnes, S. J. Oshea and M. E. Welland, *MAGNETIC FORCE MICROSCOPE STUDY OF LOCAL PINNING EFFECTS*, Journal Of Applied Physics **Vol. 76,** No. 1, pp. 418-423 (1994).

153. D. Rugar, C. S. Yannoni and J. A. Sidles, *MECHANICAL DETECTION OF MAGNETIC-RESONANCE*, Nature **Vol. 360,** No. 6404, pp. 563-566 (1992).

154. O. Zuger and D. Rugar, *1ST IMAGES FROM A MAGNETIC-RESONANCE FORCE MICROSCOPE*, Applied Physics Letters **Vol. 63,** No. 18, pp. 2496-2498 (1993).

155. O. Zuger and D. Rugar, *MAGNETIC-RESONANCE DETECTION AND IMAGING USING FORCE MICROSCOPE TECHNIQUES*, Journal Of Applied Physics **Vol. 75,** No. 10 Pt2A, pp. 6211-6216 (1994).

156. D. Rugar, O. Zuger, S. Hoen, C. S. Yannoni, H. M. Vieth and R. D. Kendrick, *FORCE DETECTION OF NUCLEAR-MAGNETIC-RESONANCE*, Science **Vol. 264,** No. 5165, pp. 1560-1563 (1994).

157. J. A. Sidles, J. L. Garbini, K. J. Bruland, D. Rugar, O. Zuger, S. Hoen and C. S. Yannoni, *MAGNETIC-RESONANCE FORCE MICROSCOPY*, Reviews Of Modern Physics **Vol. 67,** No. 1, pp. 249-265 (1995).

158. A. J. Denboef, *SCANNING FORCE MICROSCOPY USING A SIMPLE LOW-NOISE INTERFEROMETER*, Applied Physics Letters **Vol. 55,** No. 5, pp. 439-441 (1989).

159. C. Schonenberger and S. F. Alvarado, *A DIFFERENTIAL INTERFEROMETER FOR FORCE MICROSCOPY*, Review Of Scientific Instrumentation **Vol. 60,** No. 10, pp. 3131-3134 (1989).

160. S. Watanabe, K. Hane and T. Goto, *FORCE MICROSCOPE USING A TWIN-PATH INTERFEROMETER*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 1, pp. 1-5 (1992).

161. C. W. See, R. K. Appel and M. G. Somekh, *SCANNING DIFFERENTIAL OPTICAL PROFILOMETER FOR SIMULTANEOUS MEASUREMENT OF AMPLITUDE AND PHASE VARIATION*, Applied Physics Letters **Vol. 53,** No. 1, pp. 10-12 (1988).

162. M. J. Offside, M. G. Somekh and C. W. See, *COMMON PATH SCANNING HETERODYNE OPTICAL PROFILOMETER FOR ABSOLUTE PHASE MEASUREMENT*, Applied Physics Letters **Vol. 55,** No. 20, pp. 2051-2053 (1989).

163. D. Rugar, H. J. Mamin, R. Erlandsson, J. E. Stern and B. D. Terris, *FORCE MICROSCOPE USING A FIBER-OPTIC DISPLACEMENT SENSOR*, Review Of Scientific Instruments **Vol. 59,** No. 11, pp. 2337-2340 (1988).

164. D. Rugar, H. J. Mamin and P. Guethner, *IMPROVED FIBER-OPTIC INTERFEROMETER FOR ATOMIC FORCE MICROSCOPY*, Applied Physics Letters **Vol. 55,** No. 25, pp. 2588-2590 (1989).

165. M. Nonnenmacher, M. Vaeziravani and H. K. Wickramasinghe, *ATTRACTIVE MODE FORCE MICROSCOPY USING A FEEDBACK-CONTROLLED FIBER INTERFEROMETER*, Review Of Scientific Instruments **Vol. 63,** No. 11, pp. 5373-5376 (1992).

166. D. Sarid, D. Iams, V. Weissenberger and L. S. Bell, *COMPACT SCANNING-FORCE MICROSCOPE USING A LASER DIODE*, Optics Letters **Vol. 13,** No. 12, pp. 1057-1059 (1988).

167. C. A. J. Putman, B. G. Degrooth, N. F. Vanhulst and J. Greve, *A THEORETICAL COMPARISON BETWEEN INTERFEROMETRIC AND OPTICAL BEAM DEFLECTION TECHNIQUE FOR THE MEASUREMENT OF CANTILEVER DISPLACEMENT IN AFM*, Ultramicroscopy **Vol. 42,** No. PtB, pp. 1509-1513 (1992).

168. G. Meyer and N. M. Amer, *NOVEL OPTICAL APPROACH TO ATOMIC FORCE MICROSCOPY*, Applied Physics Letters **Vol. 53,** p. 1045 (1988).

169. S. Alexander, L. Hellemans, O. Marti, J. Schneir, V. Elings, P. K. Hansma, M. Longmire and J. Gurley, *AN ATOMIC-RESOLUTION ATOMIC-FORCE MICROSCOPE IMPLEMENTED USING AN OPTICAL-LEVER*, Journal Of Applied Physics **Vol. 65,** No. 1, pp. 164-167 (1989).

170. C. Ascoli, F. Dinelli, C. Frediani, D. Petracchi, M. Salerno, M. Labardi, M. Allegrini and F. Fuso, *NORMAL AND LATERAL FORCES IN SCANNING FORCE MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 3, pp. 1642-1645 (1994).

171. T. Goddenhenrich, H. Lemke, U. Hartmann and C. Heiden, *FORCE MICROSCOPE WITH CAPACITIVE DISPLACEMENT DETECTION*, Journal Of Vacuum Science & Technology A-Vacuum Surfaces and Films **Vol. 8,** No. 1, pp. 383-387 (1990).

172. G. Neubauer, S. R. Cohen, G. M. McClelland, D. Horne and C. M. Mate, *FORCE MICROSCOPY WITH A BIDIRECTIONAL CAPACITANCE SENSOR*, Review Of Scientific Instruments **Vol. 61,** No. 9, pp. 2296-2308 (1990).

173. J. Brugger, N. Blanc, P. Renaud and N. F. Derooij, *MICROLEVER WITH COMBINED INTEGRATED SENSOR ACTUATOR FUNCTIONS FOR SCANNING FORCE MICROSCOPY*, Sensors and Actuators A-Physical **Vol. 43,** No. 1-3, pp. 339-345 (1994).

174. M. Tortonese, R. C. Barrett and C. F. Quate, *ATOMIC RESOLUTION WITH AN ATOMIC FORCE MICROSCOPE USING PIEZORESISTIVE DETECTION*, Applied Physics Letters **Vol. 62,** No. 8, pp. 834-836 (1993).

175. J. Tansock and C. C. Williams, *FORCE MEASUREMENT WITH A PIEZOELECTRIC CANTILEVER IN A SCANNING FORCE MICROSCOPE*, Ultramicroscopy **Vol. 42,** No. PtB, pp. 1464-1469 (1992).

176. T. Itoh and T. Suga, *PIEZOELECTRIC FORCE SENSOR FOR SCANNING FORCE MICROSCOPY*, Sensors and Actuators A-Physical **Vol. 43,** No. 1-3, pp. 305-310 (1994).

CHAPTER 2

FABRICATION TECHNOLOGY

This chapter describes some of the major fabrication technologies which can be used to manufacture probes for applications in scanning probe microscopy. Since many of the fabrication technologies used are simply those used in the semiconductor industry, only a brief description is given here. From the capabilities of each technique, a choice of patterning method is made which will allow the objective of the work in this thesis to be fulfilled. This choice is based on the resources available, the cost of technique, the resolution of the technique and the flexibility of the technique.

2.1 SILICON SUBSTRATE TECHNOLOGY

Silicon has become one of the most useful and versatile materials of the twentieth century. Its many desirable qualities have made it the major player in the semiconductor industry and the now expanding microsensor industry. Some qualities which make silicon attractive to the former are low cost and abundance, ability to form a high quality insulating oxide, non-reactive nature, reasonably large bandgap and the ease with which it can be purified and doped. Qualities which make silicon attractive to the latter are low density and high strength, the prevalence of techniques for processing it, the ease with which it can be sculpted (micromachined), the availability of single crystal material offering a high mechanical quality factor, and the scope for integrating electronics and sensor on a single substrate. Some properties of silicon are given in Table 2.1.1.

Property	Value
Atomic Number	14
Relative Atomic Mass	28.0855
Crystal Structure	Diamond: FCC with basis (0, 0, 0), ($\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$)
Electronic Structure	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^2$ (4 valence electrons)
Photo-electric Work Function	4.2 eV
Isotopic Abundance	92.23% (28), 4.67% (29), 3.10% (30)
Ionisation Potentials	8.15, 16.3, 33.5, 45.1, 167, 205 eV
Density at 20°C	2340 kg m^{-3}
Young's Modulus	113 GPa (polycrystalline form)
Melting Point	1410°C
Boiling Point	2355°C
Specific Heat Capacity at 25°C	703 J kg^{-1} K^{-1}
Linear Expansion Coefficient 0-100°C	4.7-7.6 ppm K^{-1}
Thermal Conductivity 0-100°C	80-150 W m^{-1} K^{-1}

Table 2.1.1: Some properties of silicon from Goodfellow[1] (1993).

Low density, high strength and high mechanical quality factor are aspects of silicon which have made it particularly useful for the manufacture of pressure sensors and strain gauges where the forces acting on a silicon diaphragm are used to produce an

output signal [Petersen[2] (1982)]. As the pressure changes, air-induced viscous damping varies so that the resonant frequency and oscillation amplitude of a resonating silicon membrane also change. Silicon allows a high resonant frequency to be achieved with a relatively small amount of material and is strong enough to make sensors with 'fragile-like' dimensions practical. High strength and small dimensions mean that a large deflection amplitude, and hence increased signal level, is realisable without fracture of the sensor. This can be quantified using results from Appendix K (§K.1).

Since these qualities are exactly those desirable of AFM cantilever probes, and because of the wide variety of processing technologies now available, silicon has become the most popular choice of material for manufacturing probes for SPM applications. As silicon processing techniques have been covered in detail elsewhere [see for example, Sze[3] (1988)], this chapter only briefly describes the most common and relevant techniques for probe fabrication, a few of which are either too specialised or have not yet become mainstream methods.

Silicon wafers are usually manufactured by the Czochralski (Cz) process [Brice[4] (1986)] which involves growing a single crystal from a seed in a melt of silicon. The melt usually consists of silicon which has gone through several purification stages. The wafers are subsequently sliced up into industry standard thicknesses. Several diameters are readily available, including 1 inch, 2 inch, 3 inch, 4 inch, 6 inch, 8 inch and 12 inch. Economies of scale mean that the major consumers of silicon wafers prefer larger wafer sizes. Many mass fabrication plants now use at least 6 inch diameter wafers with 8 inches becoming increasingly common, although the use of 12 inch wafers is also rising. In general, larger wafers require more accurate processing techniques for the same yield. More care is also required in the growing process, since areas of the crystal with large defects, and wafers with more than a fixed number of particular kinds of defects are discarded as part of the wafer manufacturing process. The greater the volume of silicon in a wafer, the greater the number of defects per wafer.

However, for prototyping manufacturing methods, it is still common to use 3 or 4 inch wafers, since the equipment available for single, as opposed to batch processing, cannot be economically scaled to larger sizes due to the stricter tolerances required. It is also far easier for an individual to handle this size of wafer rather than the larger wafers which are better suited to machine processing. The increased volume of chemical solutions required for larger wafers also mean that it is more economical to process larger wafers in batches rather than individually.

Wafers may be unpolished, polished on both sides, or only polished on one side. In general, patterning must be carried out on a polished surface, although very coarse patterning with low tolerances can be done on an unpolished surface. The polished surface has a mirror finish with typically less than 2μm of topographic variation across the entire surface and no rapid topographic changes [Sze[3] (1988)]. The unpolished

surface has a matt finish with many 'sudden' topographic changes which are between 5 and 10μm in magnitude.

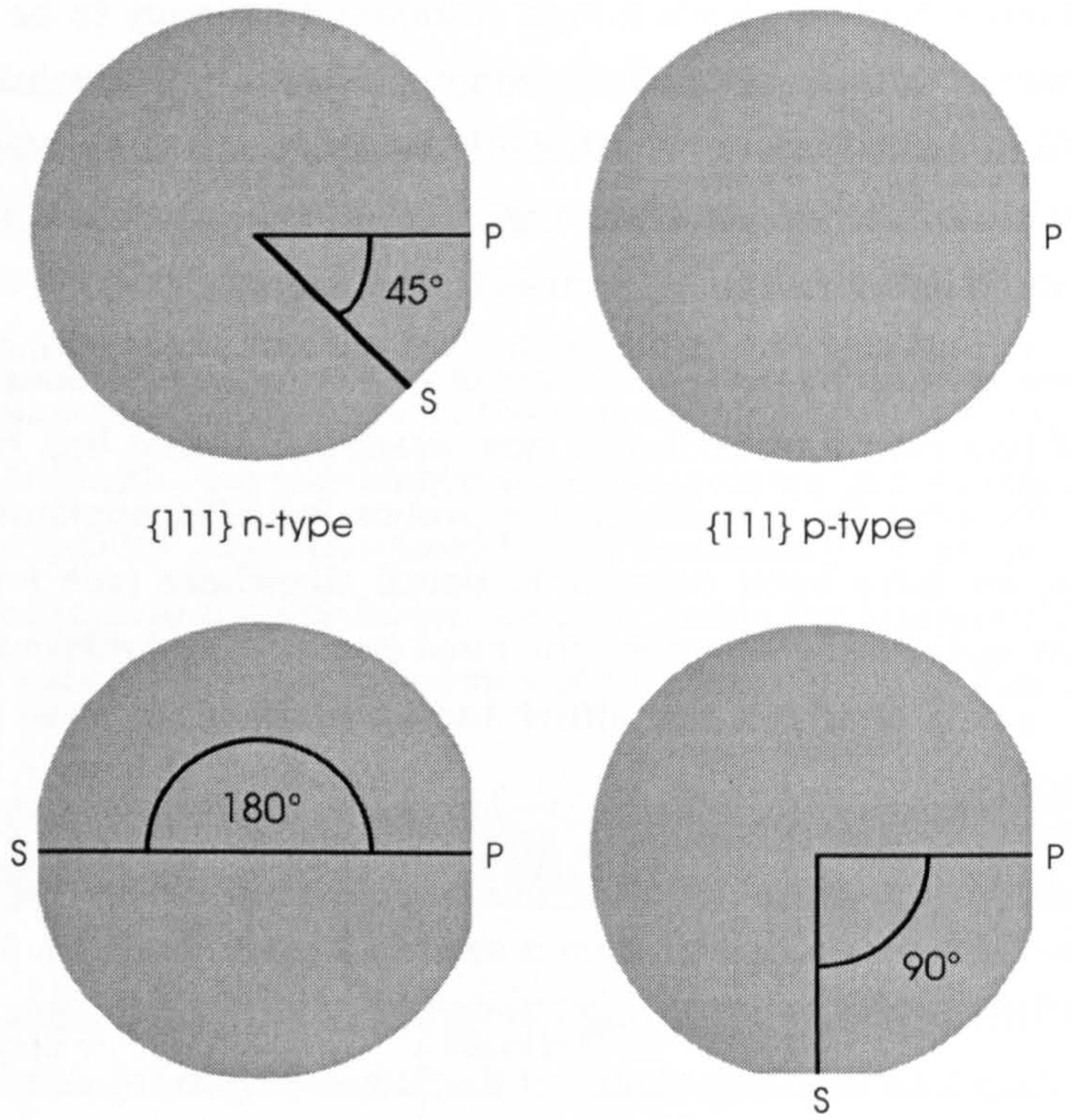


Figure 2.1.1: The primary and secondary flats which are used to identify wafers with different crystal orientations and doping types [SEMI Standard].

Larger wafers must be thicker to make wafer handling easier. The thickness roughly scales with diameter, and since 3 and 4 inch wafers were used exclusively for the work in this thesis, only these will be described here. A 3 inch wafer is usually about 390μm thick with 4 inch wafers being about 33% thicker at 520μm. However the thickness depends on the manufacturer and can vary by as much as 10μm in each direction due to the polishing processes used.

Wafers can be *n*-type, usually doped with arsenic or phosphorus, or *p*-type, usually doped with boron. The resistivities available typically range from 0.01 Ω-cm to 100 Ω-cm. Wafers with surfaces oriented along different crystal axes are available, the more common ones being <100>, <111> and <110> in decreasing order of popularity, although <110> wafers are rarely used. X-ray diffraction is usually used to determine the angle at which the crystal is sliced to make the wafers. The <111> wafers are normally deliberately sliced off-axis by 2° to 5° (to make them suitable for epitaxy) while the <100> wafers are usually sliced on-axis. Typical tolerances are about 0.5°. To allow easy identification of wafer types, flats are machined onto the wafer edges in different ways according to doping and crystal orientation. Figure 2.1.1 illustrates this for the two most common wafer orientations.

2.2 RESIST-BASED LITHOGRAPHIES

All silicon device processing involves one or more steps of patterning. These patterns are usually transferred to the silicon substrate, or to a metallic, insulating or dielectric layer deposited on top of the silicon substrate by means of an intermediate, easily patternable layer known as a *resist*. Since the pattern transfer process often involves subjecting the substrate to chemically active species in gaseous form or in solution, the intermediate layer must be able to survive in these reactive environments for the duration of the process, without serious degradation, hence the name, resist.

2.2.1 Resists

Resists can be divided into several categories: (i) photolithographic resists, which are organic polymers that can be patterned by exposure to light; (ii) electron-beam resists which are organic polymers that can be patterned by exposure to electrons, and; (iii) dielectric resists which are inorganic in nature but can also be patterned by exposure to electrons. Usually the device is coated with a resist layer for each patterning level. The various patterning processes are described in more detail later in this chapter.

Polymer resists can either be deposited on the substrate by spin-casting or by spray-coating. In both cases, the polymer is dissolved in a solvent prior to deposition. The deposition process involves draining the solvent to leave a polymer layer on the substrate.

In the case of spin-casting, the solvent-containing polymer is dripped over the substrate and the substrate is then spun at high speed, which causes the solvent to be thrown off by the centripetal forces, leaving the solute behind. In general, the higher the concentration of solute and the lower the spin speed, the thicker the resulting layer of polymer [Kuan, *et al.*[5] (1988)]. This is usually followed by a short bake to drain off any residual solvent. The main problems with this technique are that any particles located on the wafer prior to spinning result in the resist streaking from that point onwards, radially away from the centre of rotation; coating of non-circular substrates is not uniform at the edges due to the radial nature of the process, and; the resist thickness decreases slightly towards the edge of the wafer.

As the name suggests, the spray process involves spraying a uniform coating of the solvent-containing polymer over the substrate and heating the wafer to dry off the solvent. This process is much more suitable for larger wafers and requires a controlled environment due to the harmful solvents often used to dissolve resists. The equipment required to carry out this process is generally expensive and is therefore not practical for the low volumes used in device prototyping.

Dielectric resists are usually evaporated directly onto the wafer surface by means of an evaporator or sputtering machine [*e.g.* Singh, Beaumont, Bower and Wilkinson[6] (1982);

Broers[7] (1988)]. The advantages of using this dry deposition process are that the resist flow which can occur during the heating periods in the above processes is completely avoided, resulting in more uniform coatings, and; that it is able to coat non-uniform areas, such as step edges, conformally. The main disadvantage with these inorganic resists is that they usually require larger exposure doses compared to polymer based resists. This means that the sources of exposure must be capable of higher intensities or the resist must be exposed for a longer time. The former benefits both polymer and inorganic resists so that it is the latter which becomes the determining factor. Longer exposure times translate into lower throughput and exacerbates the problems of the substrate moving during the exposure process (due to thermal drift), which then results in a loss of resolution.

The patterning process consists of exposing selected areas of the resist layer and not others. This can be done by masking off certain areas if the exposure method is parallel in nature (such as flooding with a light source of a specific wavelength) or by directly avoiding exposure to certain areas if the exposure method is serial in nature (such as scanning a laser beam over the areas to be exposed). The exposure usually chemically modifies the appropriate areas [*e.g.* Pacansky and Lyerla[8] (1979); Choi, *et al.*[9] (1988)], with the result that the exposed and unexposed parts behave differently, so that a process which removes one, does not remove the other, thereby leaving a patterned layer behind. This patterned layer is then used as a mask in subsequent processing. The resist contrast determines the sharpness of the definition between exposed and unexposed areas and limits the smallest feature achievable (where the exposure mechanism itself is not the limiting factor).

In general, resists fall into two categories, positive resists and negative resists. In positive resists the area exposed is removed so that the patterned layer looks identical to the masking layer *i.e.* a positive copy of the mask (or written pattern) is formed. With negative resists this situation is reversed, so that the patterned layer looks like the negative of the masking layer or written pattern. Which is most suitable for any particular application is determined by several factors, some of which are discussed below.

At first it would seem obvious that if the area to be removed is greater than the area to be left behind a negative resist should be used and vice versa. However other considerations make this decision less clear-cut. In general negative resists tend to offer lower resolution than positive resists, as a consequence of the exposure mechanism making the exposed areas of a negative resist more difficult to remove (rather than easier to remove as is the case with positive resists). The mechanism by which this occurs is usually the linking up of short chains of polymer to form larger chains (which may contain cross-linked elements in some cases). The length of these chains combined with any swelling of the resist during development (absorption of developer) is sufficient to limit the resolution achievable. These resists also invariably have lower contrast than positive resists [Pease[10] (1981)]. The resolution of positive resists may also be limited by the length of the polymer

chains, but here the exposure breaks these chains into smaller ones, thereby increasing resolution [Wilkinson and Beaumont[11] (1980-1993); Hatzakis[12] (1988)]. It is here that the inorganic resists have the greatest advantage as they do not consist of polymer chains and so can offer very high resolution.

In addition, negative resists tend to require higher doses than positive resists and therefore longer exposure times are necessary. However, a resist layer made up of longer chain polymers tends to be much more resistant to the reactive patterning processes used in subsequent steps and so negative resists are generally more chemically resistant than positive resists. This can be a disadvantage when it comes to removing the resist layer after it has been used for patterning, as the treatment required may be quite involved and require the use of hazardous chemicals.

Reversing resists are also available, where the resist pattern is exposed as with a positive resist, but then undergoes an additional exposure stage, this time over the entire layer (without any masking). This causes broken chains in the areas exposed the first time round to reform as cross-linked chains, while breaking down the polymer chains in those areas not previously exposed, thereby achieving a reversal of the written pattern. This is one method by which a positive resist can be made to behave as a negative resist. The disadvantages of using these resists are similar to those of using direct negative resists.

In recent years, some of the distinctions between positive and negative resists have been becoming more blurred with the development of chemically amplified resists [Chiong, Wind and Seeger[13] (1990); Tritchkov, Jonckheere and Vandenhove[14] (1995)]. By making use of complex chemistry, resists can be made to contain chemically active elements in a non-active form [Namaste, Obendorf and Rodriguez[15] (1988)]. The exposing process then activates these elements which break down or cross-link the resist polymer chains much more effectively. These chemical amplifiers act as catalysts, enabling lower exposure doses to be used with the same effect, thus increasing throughput.

Another desirable quality of a resist is a short development time. The process by which either the exposed or unexposed areas are removed is known as *development*. Most development processes consist of a wet dissolution of the unwanted resist areas. In this case the longer the development time, the greater the quantity of developer absorbed by the resist and the more resist swelling occurs. This swelling can result in a loss of resolution and pattern quality as well as edge profile. This is one of the mechanisms by which the resolution of negative resists is limited, as longer development times mean such resists are more prone to swelling than positive resists [Pease[10] (1981); Namaste, *et al.*[15] (1988)]. Faster development also has the additional advantage of resulting in higher throughput.

The exposure dose required for a particular pattern depends on various factors including, the resist used, its thickness, the size and density of the features required, the age of the resist, the age of the developer and the performance of the exposure source. Usually the

correct dose is found by experimentation using a set of guidelines which are resist specific. It is therefore important to be able to distinguish between patterns which have been overexposed and those which have been underexposed. As might be expected, resists which require low exposure doses are usually more sensitive to small changes in dose compared to those which require large exposure doses.

If a resist pattern is overexposed, the exposed areas will be slightly larger than intended with sharp vertices becoming more rounded. In the case of positive resists, the edge profile will be concave as opposed to vertical, while with negative resists the edge profile will be more convex. If a resist pattern is underexposed, the exposed areas will be slightly smaller than intended with sharp vertices also becoming more rounded and narrow features developing more 'pointed' ends. In this case the development process often results in patches where the pattern is as required and patches where the pattern appears to be unexposed, *i.e.* underexposure typically results in a patchy rather than partial pattern appearing.

In general, several lithography steps are required to make any real device, and these layers must be aligned to each other. In this case it is necessary to perform *registration* of one pattern with another via identifying marks laid down in the first level patterning. These marks are known as *alignment* or *registration marks*.

2.2.2 Photolithography

The most commonly used lithographic technique in the world today is photolithography where the resist is sensitive to light at short wavelengths, usually in the ultraviolet part of the spectrum. For positive resists, this typically results in a breakdown of the exposed areas [Pacansky, *et al.*[8] (1979)] which can subsequently be dissolved using a special solution known as a *developer*.

There are generally two methods for parallel exposure using photolithography, contact printing and projection photolithography [Watts[16] (1988)]. Both involve the use of a mask with the pattern to be exposed laid out. Such masks are typically made using electron-beam lithography (discussed in §2.2.3) or using direct-write photolithography (discussed below). They usually consist of a quartz plate, a millimetre or more thick, upon which a thin (typically less than 100nm thick) layer of chromium has been fashioned into the desired pattern. The chromium acts so as to prevent light at the exposing wavelength getting through to the resist layer. The mask must therefore be placed between the light source and the substrate. The resist exposure is usually carried out using high brightness mercury arc lamps (with strong emission lines at 436nm, 405nm and 365nm) to decrease exposure time.

In the case of alignment of one patterned level to another, it is common to use a ferric oxide masking layer instead of chromium. This is opaque to the ultraviolet while being

transparent to longer wavelengths so that an optical microscope can be used to enable the registration of the mask with the previous pattern.

It is also sometimes necessary to align patterns on one side of the wafer with that on the other. In the case of semiconductor substrates, this can sometimes be achieved by using an infrared microscope since most semiconductors are reasonably transparent in this part of the spectrum. However, for accurate alignment, or where there is an absorbing layer impeding the light, it is usually necessary to either expose both sides of the wafer simultaneously, first having aligned the top and bottom masks to each other, or to make alignment marks which have been etched through from one side of the substrate to the other.

In contact photolithography, the mask is placed directly in contact with the substrate, with the chromium side actually in physical contact with the resist layer. The exposure pattern is thus a full-size replica of the mask pattern. The problems with contact photolithography are that the mask must make uniform contact over the surface of the substrate, accommodating any tilt in the wafer, and that any contaminating particles on the mask surface may be physically transferred to the resist. In addition, the mask and the substrate surfaces can form the two ends of a Fabry-Perot cavity, resulting in standing wave patterns forming in the resist perpendicular to the substrate surface. In thin coatings (less than half a wavelength) this can have a dramatic effect on exposure times as on average the interference may be more constructive than destructive [Azuma, Sato and Aochi[17] (1995)]. These effects can be minimised by making sure that the optical thickness of the resist is an odd number of quarter wavelengths so that the reflected light destructively interferes with the incident light. Alternatively, the light source may be made polychromatic, or anti-reflection coatings may be applied beneath the resist layer (this layer then has to be etched through after resist development) [Schattenburg, Aucoin and Fleming[18] (1995)]. Some other disadvantages are the sticking of the mask to the resist so that its removal can cause some damage or loss of pattern fidelity and the difficulty in using this technique to pattern non-uniform substrates.

In projection photolithography, the mask is displaced from the substrate and lenses or mirrors are used to focus an image of the mask onto the substrate. These lenses (or mirrors) are typically used to reduce the pattern by a factor of between two and ten. The main problems with projection photolithography are the cost of the equipment, and in particularly the lenses (or mirrors), which must be of very high quality and reduce aberrations to a minimum at the exposure wavelength. A means must be provided for making sure the mask is held parallel to the substrate surface and the lens is properly focused on the resist layer. However, its principal advantage is that the effect of ragged edges in the mask pattern (caused by the lithographic techniques used to produce the mask) is reduced by making the mask pattern bigger and de-magnifying the pattern using lenses, thus dividing the edge roughness by the same factor. This benefit comes at the cost of using bigger mask plates and larger lenses and resolving all the problems

associated with their manufacture. It should be noted that contaminating particles on the mask surface cannot be physically transferred to the resist, and that any particles which lie in an exposure region have a reduced effect on the resist as the mask pattern is demagnified. This technique is also afflicted by similar interference effects to contact photolithography, but to a lesser degree. Here the optical cavity is formed between the surfaces of the resist layer and substrate.

A hybrid technique also exists, known as proximity printing [Watts[16] (1988)], where the mask is held slightly away from the resist surface (typically less than 1μm) and exposure takes place as in contact printing. Here the main problem is that the light tends to diverge between the mask and the resist surface resulting in a loss of resolution and edge contrast.

In general, the resist layers used are between 0.25μm and 2μm thick. As the resist layer becomes thicker, loss of resolution and edge contrast caused by the divergence of light becomes a bigger problem. In addition the low rigidity of the polymer means that narrow features cannot support themselves against the capillary forces generated by the removal of the resist from the developer solution. Typically, it is difficult to produce features with an aspect ratio of greater than 5:1 (height:width) without resorting to more novel methods.

For both contact and projection photolithography, it is important that the illumination source be uniform over the whole pattern area otherwise the resulting pattern will appear to be overexposed and underexposed in parts. However, the main problem with both of these mask based techniques is that fine features at a high density can result in serious diffraction effects which can, at best, alter the exposure time, and at worst, result in moiré or interference patterns being formed in the resist. These interference effects can be used to advantage where periodic patterns are required, by designing the mask pattern appropriately, and are the mechanism by which phase-shifting masks produce higher resolution patterns [Ronse, Opdebeeck and Vandenhove[19] (1994)]. Its key advantage is that the exposure time is usually of the order of a few tens of seconds or less.

The alternative to this is to write the pattern directly using a laser. This has the advantage of allowing rapid turnaround in pattern designs (as a mask does not have to be made first) and not suffering from many of the interference effects common with mask-based exposure techniques. Good edge contrast and uniform exposure can be relatively easily achieved and the technique can pattern non-uniform substrates with fewer defects (as there is no mask). The disadvantage is that the process is serial so that exposure times may be on the scale of hours, the equipment required to accommodate such large exposure times must take account of thermal drift, must be able to maintain focus across the scan area, and if the pattern is larger than the scan area of the laser, a precise translation stage must be used (usually requiring accurate interferometers on two

axes and readout electronics which must be able to calibrate against temperature drift) to move the substrate so that each part of the written pattern is aligned properly with every other part. This last procedure is known as *stitching*. As a result of these requirements, the equipment for performing this technique is very expensive.

However, despite this, the method is very useful as a technique for prototyping device fabrication processes because the pattern to be written can be kept on a computer and modified easily before each development run in order to determine the best pattern layout. It is also one of the techniques by which masks can be made for use in the above parallel techniques. This means that once a development process has been fixed a mask can be made directly using the same pattern.

However, the smallest feature definable using photolithography is diffraction limited by the wavelength of light used according to [Jenkins and White[20] (1981)]:

$$\text{resolution} \approx 1.221\lambda F,\ F = \frac{f}{D} \approx \frac{1}{2NA},\ NA = \frac{D}{2\sqrt{0.25D^2 + f^2}} \approx \frac{D}{2f},\ D << f, \qquad (2.2.1)$$

where λ is the wavelength of coherent light used, F is the f-number or relative aperture of the optical system, f is the effective focal length of the optical system, D is the diameter of the exit pupil and NA is the numerical aperture of the optical system. [The resolution for incoherent light is a factor of two worse, see Broers[21] (1981).] To define smaller features, the semiconductor industry has been moving to shorter and shorter wavelengths (deep ultraviolet) in an effort to integrate more devices on each substrate. Currently state of the art chips contain features as small as 0.18μm (using a wavelength of 193nm from an ArF excimer laser) with 0.15μm also being feasible using this technique (by using light from an F_2 excimer laser at a wavelength of 157nm) [Owen, *et al.*[22] (1992); Rothschild, *et al.*[23] (1992); Grenville, Owen and Pease[24] (1994)].

However, as the wavelength approaches the soft *x*-ray region of the spectrum, materials become more transparent, so that refractive or reflective optics become more difficult to make, particularly since the aberration and manufacturing tolerances required scale linearly with wavelength. An alternative is to use diffractive masks, but this transfers the tolerance problem from the optics to the masks as these must be used without the benefits of demagnification and in a contact or proximity lithographic mode of operation. In addition, while the same aspect ratio limits apply, high resolution can only be obtained by using thin resist layers, so that pattern transfer becomes more difficult. Shorter wavelengths also mean that more accurate focusing is required to achieve accurate pattern transfer, because the depth of focus, d_f, scales according to [Watts[16] (1988)]:

$$d_f = 2\lambda F^2 \approx \frac{\lambda}{2NA^2},\ D << f. \qquad (2.2.2)$$

Direct-write is always possible, but its serial nature means that a throughput drop of several orders of magnitude per machine is incurred, making the process prohibitively expensive for use in mass production. One method for reducing this problem is to split

the laser beam into several smaller beams which are displaced from each other so that many substrates can be written at one time. The whole area of scaling photolithography to smaller dimensions is under active investigation, with the possibility that in the future, industry will be forced to make the expensive switch to one of the other techniques below, despite the high level of investment already made in photolithography.

2.2.3 Electron-Beam Lithography (EBL)

This technique makes use of resists which are sensitive to high energy electrons, usually in the range 1keV to 100keV. The principle by which this works is very similar to that of photolithography: exposure of the resist by electrons either breaks down the long polymer chains or induces cross-linking, thereby resulting in either a positive or negative copy of the written pattern. The motivation for using electron beams for lithography is identical to that for using electron beams for microscopy. Electron-beam lithography is not currently limited by diffraction and is capable of much higher resolution. Indeed, the de Broglie wavelength of a 1keV electron is of the order of 40pm.

In another analogue of the optical lithography techniques, two methods are possible: projection electron-beam lithography and direct-write electron-beam lithography [Watts[16] (1988)]. The former is less popular due to the problems associated with masking. Whilst it is relatively easy to stop electrons from passing through a mask, the gradual charging up of the mask causes electrons to be deflected away from their original trajectories and distorts the pattern transferred. It is also difficult to make masks with a resolution close to the limits of this technique. An alternative is to use the mask as a source of electrons. If the masking material is capable of generating photoelectrons when excited by ultraviolet illumination, then electric fields can be used to accelerate these electrons parallel to their emission direction. The main problem has been that the lifetime of the layer emitting the electrons is very limited.

Direct-write has therefore become the most dominant electron-beam lithography technique. Here the instrument is based on the principles of the scanning electron microscope, in that a cathode, either tungsten or lanthanum hexaboride (the latter has a lower work function but comparable Richardson constant to tungsten and is therefore a high brightness source of electrons [Broers[25] (1967); Hohn, *et al.*[26] (1982)]) or a thermionic field emitter [Tuggle, Swanson and Orloff[27] (1979)], is forced to emit electrons which are then accelerated by an anode and focused using a series of magnetic lenses. The spot size, which is an indicator of the smallest feature that can be defined, is then limited by aberrations in the lenses rather than diffraction. Spot sizes as small as 1-2nm are obtainable using converted scanning transmission electron microscopes (STEM), but these require a high numerical aperture and so result in a decreased depth of focus. Most electron-beam lithography machines are therefore based on scanning electron microscopes (SEM) which are designed to have a large depth of focus and can produce spot sizes as small as 5-15nm.

In general there are two categories of electron-beam lithography machine [Pease[10] (1981); Hohn[28] (1989)]: (i) raster scanned, and (ii) vector scanned. In the former, the pattern is written by scanning the whole imaging area and switching the beam on and off (blanking) according to the pattern to be written. In the latter, the pattern is broken up into objects which are then written one at a time. Vector scan machines can write faster than raster scan machines, but raster scan machines can place patterns more accurately relative to each other (provided they are in a single scan field).

The main problems with this technique are the same as those associated with direct-write photolithography: low throughput, expensive positioning equipment, complex drift and self-calibration mechanisms and active aberration minimisation. However, because the feature sizes achievable using electron-beam lithography are much smaller, the positioning, focusing and calibration systems need to be much more accurate than that for photolithography. In addition, stitching errors need to be much smaller than that permissible in photolithography and the need for a vacuum environment also slows the rate at which substrates can be put through the machine.

There are also some additional problems not present in photolithography: charging of the resist and substrate [Liu, Ingino and Pease[29] (1995)]; forward scattering of electrons, and; backscattering of electrons [Broers[7] (1988); Chang, *et al.*[30] (1988)]. The first of these can only be tackled by using conducting substrates or coating the resist in a thin sacrificial layer of metal prior to exposure (which can be removed prior to development). The second is caused by electrons being deflected as they pass through the resist. The angle of deflection can be reduced by using higher energy electrons, typically 50-100keV, or alternatively the effects can be minimised by using thinner resist layers so that there is less likelihood of scattering [Kuan, *et al.*[5] (1988)]. [It is interesting to note that forward scattering in thicker resists can help produce an undercut profile which is useful for some techniques such as lift-off – see §2.7.1.] The third is caused by electrons being scattered back from the substrate. The greater the atomic number of the substrate, the greater the backscattering effects will be. This effect can be reduced by patterning on thin membranes [Beaumont, Bower, Tamamura and Wilkinson[31] (1981)] or using very low energy electrons [Yau, Pease, Iranmanesh and Polasko[32] (1981); Lee, *et al.*[33] (1992); McCord and Newman[34] (1992); Peterson, Radzimski, Schwalm and Russell[35] (1992)].

Backscattering is primarily responsible for the proximity problems of electron-beam lithography [Chang[36] (1975)]. If several features are densely packed, or a large area is to be patterned, significant background exposure results due to backscattered electrons and so the dose used must be altered according to feature size, feature density and electron energy, or alternatively the features must be altered in size to compensate. This complicates pattern design for electron-beam lithography. Alignment to pre-defined patterns is also possible, by using etched or heavy metal markers which produce a strong backscatter and secondary electron image – one of the advantages of electron-beam lithography is that the machine can be used as an electron microscope for inspecting the

substrate [Pease[10] (1981)]. The edges of the marker are typically located by looking at the signal change as the spot passes over the edge [Rishton, Beaumont and Wilkinson[37] (1984)].

Some of the smallest lithographically defined features have been produced using this technique. It is relatively easy to obtain features in the 20-50nm range and features as small as 3nm have been produced with this technique using one of the highest resolution electron-beam resists, polymethylmethacrylate (PMMA) [Cumming, Thoms, Beaumont and Weaver[38] (1996)]. It has also become the most popular choice of technique for defining photolithography masks. However, it is worth noting that electron-beam resists are, generally, less resistant than photoresists to most of the chemical treatments used as part of the fabrication process.

Despite this, attempts are being made to increase throughput by making miniature electron-beam microcolumns which can be paralleled together and operate at electron energies around 1keV [Chang and Kern[39] (1992); Kratschmer, *et al.*[40] (1994); Kratschmer, *et al.*[41] (1995)]. It has also been suggested that, since most of the features required are at a scale for which photolithography can be used, in future, electron-beam lithography might be used to define a few high resolution levels with photolithography defining the other levels (mix and match lithography) [Koek[42] (1980-1993)]. This has already been done in the case of some specialised microwave devices.

In the final analysis, it is the same flexible patterning capabilities as direct-write photolithography combined with high resolution which makes electron-beam lithography a very useful tool for prototyping nanoscale devices [Rosenfield, *et al.*[43] (1993)].

2.2.4 *X*-Ray Lithography

As mentioned in §2.2.2, the wavelengths used in photolithography are gradually approaching the *x*-ray region of the spectrum. Historically, the ability of photolithography to scale down to smaller dimensions has been underestimated for many years. It is only now that further scaling is beginning to look less and less practical for reasons of cost of development rather than physics. However, the original predictions of the demise of photolithography spurred the development of *x*-ray lithography as a successor. This has proven to be much more difficult than originally expected and it is only now that it is beginning to look like a viable alternative [Smith[44] (1995)].

Originally the resolution limit of photolithography was believed to be at the level that photolithography routinely achieves today. However, as it became clear that photolithography was being pushed further and further, development of *x*-ray lithography concentrated on wavelengths much shorter than photolithography could ever reach. The distinction between the extreme ultraviolet and *x*-ray regions is simply the method by which the radiation is created. Ultraviolet light is created by exciting atoms to higher energy levels and allowing them to return to their normal state by the emission of

radiation, while *x*-rays are usually artificially created by the acceleration or deceleration of charged particles, the radiation emitted in this way being known as *bremsstrahlung radiation*. Work has concentrated on wavelengths in the soft *x*-ray region from 0.5nm to 10nm.

Everything about *x*-ray lithography has proven difficult. The difficulty in obtaining inexpensive high intensity sources of monochromatic *x*-rays has been one factor spurring the development of *x*-ray lasers [Warlaumont[45] (1989)]. Currently the most popular *x*-ray source is the synchrotron. In this (expensive) instrument, electrons are accelerated in a circular trajectory by the application of oscillating magnetic and electric fields until they reach the required energy. The fields separate out electrons with different starting velocities so that when the electrons are decelerated they emit at the same wavelength. This method achieves high intensities and recent advances have led to a decrease in the size of the instrument.

Another problem has been the manufacture of masks for this technique [McCord, Wagner and Donohue[46] (1993)]. *X*-ray lithography is intended to offer an initial resolution of 0.1μm with 50nm becoming available shortly afterwards and an ultimate resolution of about 30nm eventually being reached [Smith and Schattenburg[47] (1993)]. Since optical pattern reduction techniques cannot be used, the masks must be made with the same dimensions as the pattern required on the substrate. This means that contact and proximity printing are the only techniques possible. As *x*-rays are very penetrating, the thickness of the masking (absorbing) layer required to attenuate *x*-rays sufficiently for pattern transfer is in the range of hundreds of nanometres, so that the mask must consist of structures with a very high aspect ratio at very small linewidths to achieve the resolution required. These masks are also required to be free of defects and offer features as small as 30nm for a 300nm metal thickness. In general the masking material should have a high atomic number as the absorption of *x*-rays scales with the electron density of the material [Watts[16] (1988)]. The development of masks suitable for this have had to wait for the development of two other lithographic techniques capable of working at high resolution: electron-beam lithography and focused ion-beam lithography. The former is used to manufacture the mask while the latter is used to repair any defects or damage that occurs to the mask. The pattern transfer process required to produce these masks has been the motivation behind the development of techniques capable of producing high aspect ratio structures. Arguably, many of the developments in fabrication technology have been inspired by the problems of *x*-ray mask manufacture.

The final problem is that arising from the need to use a contact printing or proximity printing method. The problem with contact printing is that the pressure applied to the mask during the process may damage it, so that proximity printing is currently the most favoured method. In this case the resolution achievable is limited by diffraction blurring around the mask in the gap between mask and resist. The method by which these masks are aligned to previous patterns, the distortion of the mask due to the thickness of the

metal masking (absorbing) layer supported by a relatively weak mask substrate, and the thermal mismatch between the masking layer and supporting mask substrate are some of the other problems currently being addressed.

However, the advantages of this technique are also clear. It is parallel and therefore offers the same throughput advantages as mask-based photolithography. It is scaleable to dimensions small enough to mean that it is unlikely to be superseded in the near future. As *x*-rays are so penetrating, particulate contamination has little effect on the quality of pattern transfer. The high aspect ratio of the masking features mean that the radiation is well collimated allowing very tall aspect ratios with very vertical edge profiles at high feature densities to be obtained in the resist without the problems of forward scattering, backscatter and proximity associated with electron-beam lithography. There are no focusing problems so that vertical positioning need be less accurate and subsequent pattern transfer is made easier because thicker resist layers can be used. This also makes *x*-ray lithography a popular choice for surface micromachining where 'deep' patterning is often required. As a result, a process known as LIGA (a German acronym for Lithographie, Galvanoformung, Abformung) which works at much lower resolution, has been developed [Mastrangelo and Tang[48] (1994)].

Resists commonly used for x-ray lithography are the same as those used in electron-beam lithography (since the principal mechanism by which *x*-ray lithography affects the resist is by generating high energy photoelectrons), with polymethylmethacrylate (PMMA) being a popular choice. Features as small as 25nm have been achieved in resists as thick as 400nm [Haghiri-Gosnet, *et al.*[49] (1995)]. However, cost is still one of the main factors preventing widespread adoption of this technique, although the cost of photolithography at shorter wavelengths is fast approaching the point where this may change.

2.2.5 Imprint Lithography (Embossing)

This technique, possibly inspired by the method used to produce compact discs, is based on the physical transfer of a mask pattern to a resist layer by the application of pressure and heat. The idea is that a mask pattern can simply be used to stamp out patterns in the resist at low cost and with high throughput. Features as small as 15nm have already been demonstrated using the technique [Casey, Monaghan and Wilkinson[50] (1997); Chou and Krauss[51] (1997)].

One of the main problems is that the mask must be very flat so that the resist can be cleared out of the areas where it is not desired. Another problem is that the resist must be squeezed out from the solid areas of the mask, which means that there is a dependence on feature size, geometry and density. Yet more problems include sticking of the mask to the resist, thermal expansion of the mask relative to the resist, the lifetime of the mask due to contamination and damage, the lack of an obvious scheme for alignment of one pattern to another and the inability to operate on non-level substrates. However,

one of its main advantages is that a wet developer (or indeed any developer) is not necessary so that the problems with forming high aspect ratio features (associated with the capillary forces upon removal of the resist from the developer, and swelling of the resist by absorption of the developer) are eliminated and replaced by the problems associated with manufacturing the masks in the first place. It is interesting to note that much of the technology for making *x*-ray masks can be applied to making masks for embossing with the advantage that the mask substrate can be of an arbitrary thickness and material.

This technique is under active development and would seem to be attractive for stamping out devices which require a single layer at high resolution, such as diffractive optics (although optics are very sensitive to the magnification errors associated with some of the problems above).

2.2.6 STM Lithography

The scanning tunnelling microscope, as described in §1.2, can be used as an alternative source of electrons for use in electron-beam lithography. Here, the STM can either be made to operate in field emission or the resist can be exposed directly by the tunnelling electrons. The former produces larger currents and therefore faster exposure, but suffers from the lack of any focusing, so that the electrons spread out and result in resist patterns with poor edge profiles, unless the resist is kept thin and the tip scanned very close to the surface. The latter produces better edge profiles at the expense of exposure time. Both techniques are relatively new and features comparable to the resolution of electron-beam lithography machines have been produced [McCord and Pease[52] (1988); Zhang, *et al.*[53] (1989)]. The disadvantages are that the substrate must be conducting, the scanning technique is inherently slow and large step edges are a problem for the STM. Other problems include thermal drift, non-linear piezoelectric ceramic behaviour, alignment to previous levels and inability to expose substrates containing both insulating and conducting areas. Interestingly, it appears that STM lithography also suffers from the proximity problems associated with electron-beam lithography [McCord and Pease[54] (1987)]. This technique is still under development but at present is an inexpensive option for those who have such microscopes but do not have electron-beam lithography equipment [Chang, *et al.*[30] (1988)].

2.3 DIRECT LITHOGRAPHIC TECHNIQUES

The alternative to carrying out lithography using an intermediate resist layer to pattern the layer beneath, is to pattern that layer directly. Much progress has been made in this direction although the idea is not new: electron-beam lithography can be used in a direct lithographic mode by making use of silanised resists which are directly vaporised by the

incident beam (with obvious contamination consequences), or as discussed below in §2.3.2 and §2.3.3.

2.3.1 Focused Ion-Beam Lithography (FIB)

This is a technique closely related to sputtering [Gardner[55] (1994)]. Here ions are accelerated and focused just as electrons are in an electron-beam lithography machine. The advantage of ions is that the have masses large enough to physically knock off atoms from the substrate. The ions are produced by placing the neutral atoms in proximity to a sharp-pointed tungsten tip acting as an anode and source of intense electric field. The atomic species used are typically argon and gallium, although others are possible. In the case of argon, the tungsten tip is surrounded by argon gas, while in the case of gallium, a liquid metal reservoir is made to supply gallium to the tip [Watts[16] (1988)]. Ions with energies as large as 300keV are used to sculpt the substrate on the scale of nanometres [Sawaragi, Kasahara and Aihara[56] (1988)].

Extremely novel geometries have been produced in this way [Edenfeld, *et al.*[57] (1994); Vasile, Biddick and Schwalm[58] (1994)]. The electron optics for ion-beam systems tends to be more complex than that for electron-beam systems because the masses of the ions necessitate large magnetic fields. For this reason, electrostatic lenses are preferred, although these suffer from greater aberrations. However, the main problem with the technique is its speed. For this reason, the technique has been used for the production of low volumes of novel sculpted devices, particularly nanostructures and micromachines, and for the repair of both *x*-ray masks and photolithography masks.

2.3.2 Electron-Beam Deposition Lithography

This technique is related to electron-beam lithography in that a similar machine is required. However, the areas exposed by the beam are directly deposited with a metallic layer, dispensing with the need for a resist. This is done by introducing a gaseous organo-metallic compound, such as tungsten hexafluoride or dimethyl-gold, into the chamber [Koops, Weiel and Kern[59] (1988); Lee and Hatzakis[60] (1989)]. The electron-beam then decomposes the gas into solid metal and a gaseous reaction by-product (which is removed from the chamber). Decomposition occurs where the electron intensity is greatest, *i.e.* at the focal point of the electron-beam. By using optics with a high numerical aperture it is possible to decompose the metal in a very localised volume. Since the focal point can be moved vertically as well as laterally, it is possible to produce three-dimensional structures using this technique.

The main problems are the contamination of the chamber, the hazardous nature of the gases required, the limited number of metals that can be used and the slow writing speed. This is another technique more useful for specialised low volume applications and repair rather than mass production.

2.3.3 Contamination Lithography

In a similar manner to electron-beam deposition lithography, this technique relies on the formation of a deposit at the focal point of an electron-beam. In this case, however, the source of the deposit are the residual short chain hydrocarbons left in the sample chamber. It is common knowledge (amongst electron microscopists) that imaging under an electron-beam for a considerable time results in the contamination of the sample. This contamination is due to the decomposition of short chain hydrocarbons, and results in a carbon-rich deposit forming [Broers[7] (1988)]. The process can be accelerated by fixing the beam at a particular spot for a certain length of time. This technique has been used to grow 'supertips' on the end of AFM cantilevers for improved specimen access and image resolution (although the practical lifetime of such a structure is questionable) [*e.g.* Yamaki, Miwa, Yoshimura and Nagayama[61] (1992); Fischer, Wei and Chou[62] (1993)]. Unfortunately, not enough is known about the properties of these deposits for it to be used in other applications. The main disadvantage of the technique is the length of exposure time required for the formation of the deposit.

2.3.4 Nanoindentation and Nanodeposition

This technique is inspired by the numerous scanning probe microscopes now available. In nanoindentation, an STM or AFM tip is scraped across the substrate surface so as to produce very narrow lines, or simply crashed into the substrate in a gentle manner to produce nanometre scale indentations [Abraham, Mamin, Ganz and Clarke[63] (1986); McCord and Pease[64] (1987)].

In nanodeposition, an STM tip is pulsed with a high voltage so that material transfer takes place between the tip and substrate (the direction depends on the voltage used). This technique can be used to form small mounds or dents in surfaces, although, some structures appear to gradually disappear with time [Schimmel, Fuchs, Akari and Dransfeld[65] (1991); Schimmel, *et al.*[66] (1995)].

Currently these techniques have some novelty value and it is not yet clear what the future holds for them. However, the main problems are the limited lifetimes of the tip, extreme material dependence of the patterning process, slow scanning speed, non-linear scan fields, lack of an obvious alignment scheme and problems with substrates which are either non-conducting or have many step edges.

2.4 METAL DEPOSITION

All electrical devices require metallic regions, either as electrical contacts to the devices or for electrical connections to other devices. Both cases require a metal layer to be patterned. Normally a metallic layer is deposited onto the substrate and a resist pattern is formed over it to act as a mask for the selective removal of the metallic layer.

Alternatively, the metal can be deposited on the substrate by using a resist pattern as a stencil. Both of these techniques will be discussed in §2.7. The thickness of metal deposited may be limited by the residual stress in the film which can result in the film cracking. Such stresses, which may be induced by the thermal expansion mismatch between the substrate and film, are metal specific, and also usually dependent on factors such as deposition rate and temperature. Alloy films tend to be more stressed than pure metal films.

2.4.1 Evaporation

In this technique, the metal to be deposited is evaporated under vacuum conditions and allowed to condense on the sides of the chamber and the substrate [Murarka[67] (1988); Mastrangelo, *et al.*[48] (1994)]. Various heating techniques can be used such as resistive heating, inductive heating (radio frequency), laser heating or electron-beam heating. For some metals, particularly refractory metals, resistive heating can be difficult because of the high melting points required, so that electron-beam or laser heating is preferred.

One problem with resistive and inductive heating is that the amount of infra-red radiation generated can substantially heat the substrate. This is particularly undesirable when evaporating onto a resist layer. In addition, resistive and inductive heating cause all of the metal in the source to melt, so that any low boiling point contaminants present are deposited onto the substrate prior to the desired metal, causing problems with metal adhesion or changing the properties of the deposited film. In contrast, electron-beam and laser heating can be made to melt only a central area of the metal, thus avoiding these problems. However, electron-beam heating can generate ultra-violet radiation which may adversely affect resist layers, leaving laser heating as the best choice. Unfortunately, the lasers required are expensive so that resistive and electron-beam heating are the most commonly used techniques.

A characteristic of deposition by evaporation is that the metals are generally deposited in a slightly directional manner – this is particularly useful when evaporating onto a resist stencil, but may not be desirable if step edges are required to have a uniform coating of metal. The directionality arises because the metal source is usually confined to a small region and acts almost as a point source of metal atoms. However, if the distance from the source to the substrate is too short, metal atoms reaching the edges of the substrate may arrive at a very different angle to those at the centre.

One drawback of evaporation is that it is not generally suitable for evaporating alloys, unless each of the constituents is evaporated in parallel using independent heating sources. Another drawback is that because the metal atoms solidify on the substrate at a higher temperature than the substrate, metal films deposited by evaporation tend to be highly stressed and are usually limited to less than 1μm in thickness, or cracking of the film occurs. The thickness of metal depositions are usually monitored by measuring the

change in resonant frequency of a quartz crystal (caused by the change in the mass of the deposit on the crystal).

2.4.2 Physical Sputtering

In this technique, ions of an inert gas such as argon or helium are used to bombard a sputter target made of the metal to be deposited [Mastrangelo, *et al.*[48] (1994)]. These ions knock off the metal atoms which then condense on the substrate surface. The advantages of this technique are that, because the metal is not melted, but simply eroded, high temperatures are not required, minimising the amount of substrate heating and reducing the stress in the film so that much thicker metal layers can be deposited. In addition, sputtering allows alloys to be deposited with compositions very close to that of the target, even though the constituents may have quite different melting points. Sputtering also produces layers which are more conformal, as the metal target is of a larger diameter than the substrate so that metal atoms can be knocked off in all directions, producing uniform coverage across step edges. This means that this technique is not suitable for evaporating onto a resist stencil. It should be noted that the metal films produced by sputtering may contain small amounts of trapped gas atoms.

2.5 DEPOSITION OF NON-METALS

Insulating layers are often required to separate conducting or semiconducting regions from each other. For some devices, particularly sensors, deposition of semiconductors is also an important step in the fabrication procedure. These layers are generally patterned after deposition, rather than being deposited in the pattern required.

2.5.1 Chemical Vapour Deposition (CVD)

In this technique, the film is made by the dissociation of various gaseous components on the surface of a hot substrate by means of pyrolysis [Adams[68] (1988)]. Films such as silicon oxide, silicon nitride and polycrystalline silicon (polysilicon) can be deposited in this way. Several variants are possible such as low pressure chemical vapour deposition (LPCVD) which produces the highest quality films with good stoichiometry, but requires deposition at high temperatures (around 700°C), and plasma enhanced chemical vapour deposition (PECVD) where the gases are ionised to form a reactive plasma (using high d.c. or r.f. voltages) which results in faster deposition at lower temperatures (around 300°C), but compromises on film quality and density.

In general, LPCVD film characteristics can be better controlled than PECVD films, because the films formed by the latter process often contain trapped reaction products, such as hydrogen, which affects the density and film stress. PECVD films also often have pinholes, and may even contain cracks. PECVD is usually used only when high temperatures are not suitable, such as when devices already contain patterned metal

layers, while LPCVD is the preferred choice where this is not a concern. Both methods can result in highly stressed films, but this can be relieved by altering the deposition conditions or by subsequently annealing the films, with ion implantation into the film being yet another option [Mastrangelo, *et al.*[48] (1994)]. The main limitation of this deposition technique is that many dielectric films cannot be easily produced in this way.

2.5.2 Molecular-Beam Epitaxy (MBE)

This technique is used to grow single crystal semiconductors by the very slow evaporation of its constituent elements [Pearce[69] (1988)]. However, deposition must be carried out under ultra-high vacuum conditions. By evaporating the elements slowly, the atoms landing on the substrate are very mobile and have sufficient time to locate at the most favourable position (as determined by surface energy considerations) so that a single crystal is formed. The film growth is dependent on the substrate crystal orientation. The main advantage of this method is that dopants can be introduced during film formation in a very controlled manner, allowing precise doping profiles to be obtained. Dopants can also be limited to a much thinner layer (which may be buried) than is possible with other techniques. However, the process is very slow and is generally used to produce optically active materials and substrates for optical devices, such as quantum-well lasers.

2.5.3 Physical Sputtering

As with metals and alloys, semiconductors and dielectrics can be sputtered by using an appropriate target. However, the interface between substrate and sputtered material is not generally as good as that obtained by chemical vapour deposition, although a much wider range of materials can be deposited by this method.

2.5.4 Thermal Oxidation

This technique is not a deposition technique but a film growth technique. It is used to form silicon oxide films on raw silicon substrates [Katz[70] (1988)]. Typically the substrate is heated to a high temperature in a furnace in the presence of oxygen or steam, so that the silicon reacts thermally to produce silicon oxide. These films are of very good quality, with no pinholes, and are therefore particularly good as diffusion barriers. They trap less charge at the silicon-silicon oxide interface but are highly stressed, although they do not crack or peel.

The main limitation of this process is the very high temperatures required (around 1000°C) and slow growth rates (particularly for thicker layers where the rate of film growth is limited by the rate at which the reacting gases can diffuse to the substrate-film interface). Initially the thickness of the oxide film tends to increase linearly with time but this rate of increase eventually decays to the square root of time. In practice this means

that a whole day may be required to grow layers greater than 1μm to 1.5μm thick. The main requirements of this technique are a furnace which can maintain very good temperature uniformity across the silicon substrates.

2.6 MISCELLANEOUS PROCESSES

This section describes some other important processes not easily categorised under the above sections.

2.6.1 Silicon Doping

Doping is the technique by which impurity atoms are introduced into the substrate to change the sign of the majority carrier or alter the concentrations of these carriers [Pearce[69] (1988)]. Doping during wafer manufacture determines the majority carrier in the substrate, however, for the formation of abrupt junctions, as required in the fabrication of semiconductor transistors, diodes and capacitors, a method must be used to alter the doping type in a selected layer or region of the substrate. Three main methods exist: diffusion into the substrate from a gaseous source at high temperatures, diffusion from a heavily doped glass layer deposited on the substrate and ion implantation of dopant atoms into the substrate. For silicon, boron is the dopant used to generate *p*-type regions and phosphorus or arsenic is the dopant used to generate *n*-type regions.

In the case of diffusion doping [Tsai[71] (1988)], the initial incorporation of impurity atoms results in a highly doped, thin layer at the interface. The substrate is then heated to drive the dopant deeper into the substrate, thus reducing the density of dopant atoms and producing a more evenly doped layer. In the gaseous scheme, a furnace is used and an inert gas containing the dopant, such as arsenic hydride, phosphorus hydride or boron hydride, is passed over the substrate and becomes incorporated into the exposed areas of the silicon. In the glass-layer scheme, a glass containing the dopant, such as borosilicate glass or phosphosilicate glass, is deposited in liquid form using spin-casting or in solid form by sputtering and then annealed for a short time before the dopants are driven in using a furnace.

In ion implantation, the impurity atoms are ionised, and accelerated so as to bombard the substrate [Giles[72] (1988); Gardner[73] (1994)]. The dopant atoms penetrate the substrate to a certain depth determined by the acceleration energy and tend to occupy interstitial sites. Since this process is not carried out at high temperatures, the dopant needs to be activated, usually by a rapid thermal anneal (RTA), which forms the bonds between the silicon and dopant atoms and allows the silicon lattice to relax without much spreading of the dopant. This technique is particularly useful for producing very shallow doped layers without resorting to molecular-beam epitaxy. In addition, ion implantation can be used to implant other species, resulting in a layer buried within the silicon

substrate. Buried silicon oxide layers have been produced in this way, although the layers are located fairly close to the silicon surface.

2.6.2 Sintering and Annealing

To form proper contacts between metals and semiconductors, the metal often has to be made to diffuse into the semiconductor, so as to form a graded metal atom density profile. Other materials, such as polycrystalline semiconductors can have their properties altered by being made to recrystallise, thus forming bigger crystals. A similar procedure can be carried out on metal layers. All these processes require heating of the layer in question to a set temperature.

In the case of sintering, the substrate is raised to a temperature at which the deposited layer just begins to melt so that the atoms can diffuse into and fuse with the substrate. This technique is used to make electrical contacts and is usually carried out in an inert atmosphere such as argon.

More commonly, substrates are annealed using very high power tungsten-halogen lamps, which produce a rapid rise in temperature over a short period (a few minutes at most), during which recrystallisation, lattice relaxation or dopant activation can take place. The principal advantage of this method is that the short heating time minimises the diffusion distance and the creation and coalition of defects in the substrate. This process is usually carried out in an inert gas environment.

2.6.3 Wafer Bonding

This is a process by which a silicon wafer can be attached to another substrate, which may also be a silicon wafer, to produce a more complex, or solid, device geometry. This is particularly useful where devices must be encapsulated or sealed in a vacuum cavity, or where it is easier to fabricate the device on a thin substrate (as in many bulk micromachining applications) and then attach the substrate to a more bulky substrate for ease of handling or integration with the final instrument. There are a number of versions of this technique: anodic bonding, low-temperature glass bonding, fusion bonding and reactive metal bonding [Bengtsson[74] (1992); Gardner[73] (1994)].

In anodic bonding [Wallis and Pomerantz[75] (1969); Albaugh[76] (1991)], the substrates are held together by strong electrostatic forces. One substrate must be conducting while the other is a sodium-rich glass substrate. The substrates are held together and heated to a high temperature so that the sodium ions become mobile. The application of a large voltage then forces sodium ions to cross over to the silicon substrate, thus creating a depletion region (analogous to that in a *p-n* junction) across which strong electrostatic fields are present. These fields are typically very strong, so that the bond is often stronger than the substrates themselves. This technique is sometimes known as electrostatic

bonding. However, the high temperatures and electric discharge mean that this technique is not suitable for many devices which would be damaged by this procedure.

In low-temperature glass bonding, special low-temperature glasses are attached to the substrate using simple thermal annealing while applying pressure to hold the surfaces together. Bonds form between the glass and substrate and temperatures as low as 300°C can be used. The bonds are not as strong as in anodic bonding and the process is sensitive to the flatness and cleanliness of the bonding surfaces. Silicon wafers can be bonded together by coating both surfaces with such glasses, usually by spin casting.

Fusion bonding is normally used to bond together like surfaces, particularly single crystal substrates such as semiconductor wafers. Here, the surfaces are cleaned to remove any intermediate layers and must be very flat so that when placed in contact with each other, the surfaces stick together due to weak Van der Waal's forces. The surfaces are then annealed at high temperature so that proper chemical bonds form across the interface. This technique can produce results where the interface is almost invisible, but suffers from the problem that high temperatures are needed and the initial weak bonding, which is sensitive to flatness and cleanliness, determines the quality of the bond.

In reactive-metal bonding, a thin layer of metal is deposited onto both substrates. These are then held together and heated. Metals which alloy with silicon when heated are chosen so that the bond consists of a silicon alloy interface. The temperature used usually corresponds to the eutectic temperature of the metal-silicon system, allowing low temperatures to be used with the proper choice of metal. For example, the gold-silicon system requires a temperature of only 360°C to alloy. In this technique the native silicon oxide layer must be removed prior to deposition of the metal.

2.7 PATTERN TRANSFER TECHNIQUES

This section describes the processes by which metal, semiconductor and dielectric layers are patterned using an appropriate patterned masking layer such as a resist or dielectric. Many schemes have been developed in order to produce a faithful reproduction of the patterned mask in the layer beneath. These range from the use of reactive gases and reactive solutions to simple deposition.

2.7.1 Lift-Off

This is one technique for producing metal patterns on the substrate. Here the resist mask is used as a stencil upon which metal is deposited. Provided the resist profile is slightly undercut, and the metal layer is thinner than the resist, the metal will be discontinuous across the resist edges. Then the resist can be dissolved taking the metal deposited upon it away from the substrate and with only the metal that was deposited directly on the

substrate remaining [Beaumont, *et al.*[31] (1981)]. Adhesion of the metal to the substrate is an important factor in this process so that it is sometimes necessary to use a thin intermediate metal layer to act as 'glue' between the substrate and the required metal. These 'glue' metals are generally referred to as 'sticky layers'. Typical 'sticky layer' metals include titanium, chromium and nickel/chromium. The resulting metal pattern is an exact negative of the resist pattern

This technique is described as an additive pattern transfer process because material is added after the resist pattern has been defined. Additive processes invariably result in an inversion of the mask pattern. The requirements for this process to work are an undercut profile and a reasonably vertical metal deposition scheme which is non-conformal across step edges. This rules out sputtering and means that evaporators in which the metal source is separated from the substrate by some distance must be used (so that the solid angle subtended by the substrate at the metal source is small).

In electron-beam lithography, thick resists naturally have an undercut profile due to the effects of forward scattering. However, the use of thin resists in electron-beam lithography require alternative mechanisms for the production of a resist 'overhang'. This is typically done by using two resist layers with slightly differing sensitivities [Hatzakis[12] (1988)]. A thinner layer of the lower sensitivity resist is deposited on top of a thicker layer of the higher sensitivity resist so that when the bottom layer is exposed adequately, the top layer is slightly underexposed, resulting in an overhanging resist profile. With some photoresists, treatment with a chemical can affect the sensitivity of the top layer of the resist to produce the same end effect. Due to the use of an underexposed resist layer, this method usually cannot produce crisp vertices as the features of the top layer of resist are slightly rounded.

In general, the metal pattern is slightly smaller than the resist features, although this is easily corrected by adjusting the pattern opening size. The thickness of the metallisation is limited by the need for a discontinuity in the metal layer. However, for tall aspect ratio structures, another effect can be the dominant factor. As the metallisation proceeds, metal deposits on the edge of the overhanging resist, gradually closing up the resist opening. This results in the metal which is deposited in the opening gradually tapering with increasing layer thickness. The angle of the sidewalls is therefore not vertical but typically at an angle of between 70° and 80° to the substrate. This effect has been used to advantage in the manufacture of sharp tips for STM or field emitters.

When lift-off fails, usually due to the absence of discontinuities between the metal upon the resist and that directly on the substrate, it is the vertices which tend to fail first. Lift-off can produce patchy results, and is not generally regarded as a reliable enough process for high volume production, but its ability to produce very small features with reasonable accuracy mean that it is commonly used in the definition of ultra-small structures and where other techniques are not suitable. This direct replication ability means that rough

edges in the resist masking layer are also reproduced in the metal layer which can be a problem for some kinds of devices. In addition, the edges of metal structures patterned in this way often exhibit the presence of 'flags' where the metal on the substrate has been ripped from the metal upon the mask during the dissolution of the resist because the metal regions were not completely discontinuous across the resist step edge. Unfortunately, the lack of vertical sidewalls means that this technique is unsuitable for the manufacture of *x*-ray masks or masters for use in imprint lithography. The latter also requires the top surface of the metal layer to be extremely flat.

2.7.2 Plating

This is an alternative additive process for producing metal structures [Hatzakis[12] (1988)]. Plating can be divided into two categories: electroplating and electro-less plating. In electroplating, the substrate, which must be electrically conductive, is used as one electrode in an electrolytic cell containing the metal to be deposited in solution, with another electrode, usually graphite or platinum, completing the cell. Deposition of the metal ions in the solution on the exposed parts of the substrate then takes place by electrolysis. A patterned resist mask can be used as a stencil for the plating. Very high aspect ratio structures (exceeding 10:1) at small dimensions (50nm) have been produced in this way [Haghiri-Gosnet, *et al.*[49] (1995)].

The main problem with this technique is that certain geometries, in particular those containing sharp vertices, can result in an enhanced localised electric field density and hence non-uniform deposition rates. Thick layers may also be stressed, although this can be controlled by adjusting the plating conditions appropriately. Several plating conditions must be controlled in order to obtain a uniform coating including the current density, temperature, concentration and composition. The quality of the plating surface is yet another factor affecting the end result, and the formation of bubbles in the solution can result in porous films being deposited. This technique is one of the possible methods by which masters for embossing and *x*-ray masks may be produced, and is one of the principal techniques used in the process known as LIGA for fabricating micromachines. It is also useful for the fabrication of hollow metal structures, such as air bridges, because the plating can be carried out in steps with a change of resist pattern between steps.

Electro-less plating does not require a conducting substrate and is a technique often used to coat plastics and other surfaces which metals often have difficulty adhering to [Parameswaran, Xie and Glavina[77] (1993)]. The first step is to chemically alter the exposed surfaces so that they are covered in chemical 'hooks' which are bonded to the substrate and contain a metal atom suitable for use as a seed for the plating process. The whole substrate is then immersed in a special solution containing the metal to be deposited in ionic form. An enzyme or catalyst may also be present in the solution. Heating of the solution then induces the condensation of the metal from the solution onto the chemical hooks.

This technique can produce results similar to electroplating, although the deposition rate can only be controlled by the temperature so that the properties of the resulting film are in general less well controlled than with electroplating. In addition, the adhesion of electro-less plated layers is often not as good as those of electroplated layers, and the deposition rates are usually more feature size dependent, making it difficult to obtain uniform metal layers. The main difficulty with this technique is that the metal also coats the resist, and in particular the sidewalls, so that lift-off cannot easily be carried out afterwards. Alternatively, if a thin layer of the metal to be plated is evaporated and patterned using lift-off, immersion in the solution, without using the chemical 'hooks', causes the condensation of the metal ions onto the pre-defined metal pattern. Unfortunately, the use of a resist cannot be eliminated when a vertical structure is required, as the plating will in general grow outwards as well as upwards unless constrained by the walls of a stencil.

2.7.3 Wet Etching

This is a subtractive process used to pattern metals, dielectrics and semiconductors. Material is removed from the substrate after the resist layer has been patterned. In this process, the substrate is immersed in a chemically active solution, where the exposed parts of the layer being patterned react to produce volatile by-products which diffuse into the solution. The resist acts as a mask preventing the unexposed parts of the layer being patterned from being etched. Many wet etching solutions are highly acidic or alkaline so that the etching is often limited by the ability of the resist to withstand the etch. Thicker resists survive better than thinner resists and photoresists generally withstand the etching solution better than electron-beam resists. Some etching solutions are so corrosive that a solid masking layer must be used to mask the layer being patterned, this layer having itself been patterned using a resist mask by another, more compatible process.

Wet etches can be classified as isotropic or anisotropic [Gardner[73] (1994); Mastrangelo, *et al.*[48] (1994)]. In an isotropic etch, the film is etched at a uniform rate in all directions so that the resist mask is undercut and the edges of the patterned layer have curved profiles. This makes isotropic wet etching unsuitable for the definition of very small features, as the feature increases its width by an amount equal to the depth of the etch.

Anisotropic etches can be subdivided into two further categories: crystallographic and vertical etches. Vertical wet etches result in sharply defined edge profiles but make use of novel surface activation techniques as discussed below. These etches are often used in the surface micromachining of silicon where very high aspect ratios are required. Crystallographic wet etches are mainly used for etching single crystal materials, usually the semiconductor substrate itself. These etches have different etch rates for different crystallographic planes and can produce novel geometries which may look very different from the original masking pattern. The final shape is determined, not only by the

masking layer pattern, but also the exact crystallographic orientation of the layer being etched. These etches are often used in the bulk micromachining of silicon.

At its simplest a wet etch contains a reducing agent and an oxidising agent [Kern and Deckert[78] (1978)]. The oxidising agent oxidises the material to be removed while the reducing agent attacks this oxidised layer, producing volatile by-products. The behaviour of this etch is determined by whether the reaction is limited by the rate at which the chemical reaction (the slowest of the reducing or oxidising stages) takes place, or by the rate at which the by-products diffuse into the solution. In the case of the former, the etch is usually crystallographic in nature as different planes have different dangling bond orientations and bond energies so that they are oxidised and reduced at different rates. In the case of the latter, the etch will behave isotropically. In some wet etches, an oxidising agent is not necessary as the material to be removed is already oxidised, but the behaviour is essentially the same. If the material being etched is amorphous or polycrystalline on a very small scale, crystallographic effects will not have a noticeable effect on the end result.

In the case where the reaction is rate limited, the dependence on the ability of the etch to attack chemical bonds can be used to advantage. The energy required to break bonds can be altered by the application of an electric field or by illuminating the area with light. The former is known as electrochemical etching while the latter is known as photochemical etching. By applying laser light to a small area of the material to be etched, the rate of etching of the illuminated area can be enhanced or impeded resulting in very vertical etch profiles (although the feature size is clearly limited by diffraction) [Khare, Hu, Brown and Melendes[79] (1993); Mlcak, *et al.*[80] (1994)]. The simplest electrochemical etching occurs by making the substrate one electrode of an electrolytic cell with a graphite or platinum electrode used as the other. Etching by electrolysis can then take place by maintaining an appropriate potential difference across the electrodes. This means that the substrate must be conducting. Another variation of electrochemical etching is seen when silicon is etched in certain solutions. Regions with different dopants etch at different rates with *p*-type regions typically etching slower than *n*-type regions on the same sample. This is caused by the in-built *p*-*n* junction bias. The relative etch rates can be altered by applying a bias across the junction (without forward biasing the junction) [Mastrangelo, *et al.*[48] (1994)]. One junction can even be etched preferentially with respect to the other by applying a positive bias between it and an additional electrode in the solution (again without forward biasing the junction) [Ozdemir and Smith[81] (1992); Soderbarg and Smith[82] (1993)].

One of the most important considerations in etching is the control of the etch depth. For long etches, as used in the bulk micromachining of silicon, it is desirable to have an etch-stop layer. This layer, when reached, is not etched by the solution so that small changes to the etch rate do not result in over-etching [Gardner[55] (1994)]. Even for short etches,

particularly isotropic etches, over-etching is undesirable as feature sizes may increase laterally (and vertically if the substrate beneath is not chemically resistant to the etch).

The other main requirements for accurate wet etching are the control of the etch solution concentration and composition, and the control of the uniformity and stability of temperature across the exposed substrate areas, with calibration of the etching solution often necessary for precise dimensional control. The main disadvantages of the wet etching technique are that it is usually isotropic, resulting in the undercutting of the mask; the corrosive nature of most etching solutions means that compatibility between the various materials on the substrate and the etching solution complicates the fabrication process, and; a special masking layer may be required if resist is not a suitable mask. However, wet etching is still the dominant metal patterning technique for high volume production.

2.7.4 Dry Etching

This is a subtractive process used to pattern dielectrics and semiconductors, as well as some metals. As with wet etching, material is removed from the substrate after the resist layer has been patterned. In this process, the substrate is placed in a chamber with a chemically reactive gas so that the exposed parts of the substrate react to produce volatile (gaseous) by-products which are removed from the etching chamber [Schutz[83] (1988)]. The resist acts as a mask preventing the unexposed parts of the substrate from being etched. Many dry etching gases are highly corrosive so that the etching is often limited by the ability of the resist to withstand the etch. As with wet etching, thicker resists survive better than thinner resists and photoresists generally withstand the etching gases better than electron-beam resists. For some dry etches a solid masking layer must be used to mask the layer being patterned, this layer having itself been patterned using a resist mask by another, more compatible process.

Whilst it is possible to chemically etch with pure reactive gases such as chlorine and fluorine, such etches are usually isotropic and do not offer much of an advantage over wet etching. In order to provide directionality and increase the rate and selectivity of the etch, most dry etching is carried out by reactive ions in the form of a plasma. The plasma is usually created using a high voltage radio-frequency source, typically at 13.56MHz. At this frequency, the difference in mobility between the ionised molecular fragments and the electrons results in an average negative potential on the radio-frequency electrode and the sidewalls of the etching chamber (which are usually grounded). The potential on the electrode is prevented from discharging by the presence of a blocking capacitor between the high voltage source and the electrode. The substrate is placed on this electrode and the bias it experiences is known as the self bias, which can range from a few tens to a few hundreds of volts (negative) depending on the plasma density. This bias attracts the positively charged fragments causing them to bombard the substrate almost

vertically. The result of this is that the exposed parts of the substrate are etched anisotropically. This technique is known as reactive-ion etching (RIE).

The exact etch profile is determined by many factors including the self bias, the plasma density, the flow of gases into and out of the reaction chamber, the ratios and composition of the reactive species, the masking material and the material being etched, and the pattern geometry (charging effects and electric field enhancement at particular geometries can dramatically deflect the reacting ions resulting in very different etching profiles [Ingram[84] (1989)]). The plasma density is controlled by the voltage and frequency of the exciting source, while the self bias is affected by both of these and the gas flow rate. In general, less dense plasmas produce more vertical etch profiles due to the reduction in intermolecular scattering, but at the expense of reduced self-bias.

Other factors which can affect the results are polymerisation of reaction by-products, reflection of ions from the substrate, and sputtering and erosion of the masking layer. If some of the reaction products are non-volatile, they can be polymerised by the action of the ions in the plasma resulting in uneven etching or even preventing the etch from proceeding further. With some combinations of etch gases, the gases themselves are polymerised along with the reaction by-products. This process can lead to the phenomenon known as sidewall passivation (deposition), where polymer builds up on the edge of the resist walls, having been sputtered off the bottom of the etched feature by the action of incident ions. This can help to prevent lateral feature growth. The addition of small quantities of oxygen to the plasma can often prevent polymerisation by reacting with the short polymer chains (or monomers) to produce carbon dioxide and water. However, too much oxygen results in resist masks being attacked at an accelerated rate.

Another problem is that the high self bias can result in the ions having enough energy to severely etch the resist or substrate by physical sputtering. If this is the dominant mechanism by which etching proceeds, the technique is often given the name sputter etching. When this is the required etching mechanism, it is common to use heavier plasma gases such as argon and xenon. In general, sputtering reduces the lifetime of the masking layer. To increase the lifetime of the mask, a lower self-bias is required without compromising on the plasma density. This can only be achieved by decoupling the plasma density from the self bias.

In one method, known as electron-cyclotron resonance reactive-ion etching (ECR-RIE), the plasma is generated by using microwave radiation, usually at 1.3GHz or 2.6GHz. The use of high frequencies reduces the effect of the disparity in the mobilities of the electrons and positive ions and leads to a lower self-bias. This would result in isotropic etching, but if the plasma is separated from the substrate by a reasonable distance (the substrate is not placed on an electrode in this technique), some collimation of the ions occur and an additional d.c. bias can be used to encourage anisotropic etching. By using a magnetic field gradient, the density of the plasma can also be altered independently so that much

more control over etching behaviour is obtained and the lifetime of the masking layer is increased. High etch rates can be obtained at lower pressures using this technique [Juan and Pang[85] (1995)], although if the magnetic fields used vary appreciably across the substrate, non-uniform etching will occur.

Fluorine and chlorine play an important role in most dry etching chemistries, due to their reactive nature, and can be introduced to the plasma in many different forms for different effects, such as in pure molecular form (diatomic molecules), as part of short chain hydrocarbons or attached to other light atoms such as silicon and boron. Oxygen and hydrogen are other gases normally used to subtly alter etching behaviour.

Proper optimisation of etching conditions can lead to accurate replication of the masking layer at very high resolution with very high aspect ratios being achievable [Rangelow and Loschner[86] (1995)]. This makes dry etching the most popular choice of pattern transfer technique. Unfortunately most metals cannot be etched with high selectivity over resist and wet etching must be used in these cases. However, dry etching is currently the most commonly used technique for the patterning of dielectric layers.

2.8 CONCLUSIONS

It is clear that many of the fabrication processes described in this chapter are compatible with all lithographic techniques. Therefore the most important choice is that of the basic lithographic technique used. Since the resolution achieved by sensors for scanning probe microscopy are usually limited by the size of the sensor, a high resolution technique is required for sensor fabrication. Unfortunately, all the high resolution schemes described here are expensive from a manufacturing perspective. However, the active area of the sensor is only a small part of the whole cantilever and tip so that the best option is to use a mix and match lithographic scheme, with photolithography used to define the parts of the fabrication process not requiring ultra-small dimensions. As the work presented here is of a development nature, it makes no sense to use *x*-ray lithography for the high resolution patterning as this requires a new mask to be made every time the process is changed. Thus the choice of high resolution patterning technique is limited to electron-beam lithography or focused ion-beam lithography. Both are low volume techniques, but this is not a problem as a single wafer is capable of holding many hundreds of sensors. This means that even if the fabrication process for a sensor was a month in duration, the use of these techniques would still represent a two order of magnitude increase in throughput over the individual crafting of such sensors, and the market for these sensors is likely to be low volume in any case.

In an ideal world, the arguments used above would allow one to pick whichever process was convenient for each type of sensor, however, every institution has a limited number of resources it can make available. In the case of the University of Glasgow, focused ion-beam lithography is a technique which is currently unavailable and the expected return

from these sensors, due to their low volumes, cannot justify the cost of acquiring this technique in any case. Fortunately, the University has its own direct-write electron-beam lithography system so that this technique was used for all the work presented here.

Normally, once the high resolution patterns are fixed, one might consider converting to a high throughput, low flexibility technique such as *x*-ray lithography. However, the University does not have such a facility and its acquisition cannot be justified on commercial grounds. In addition, the only photolithography technique available in the University is that of contact printing. Thus the choices have all been made, and it is clear from the above arguments, that the absence of the other facilities will not affect the achievement of the objective of this work, but merely the options for the future. [Note that the above arguments applied at the time the work in this thesis was carried out.]

The rest of this thesis is therefore concerned with the detailed sensor design and choice of fabrication processes in the context of contact photolithography and direct-write electron-beam lithography.

2.9 REFERENCES

1. Goodfellow, *METAL DATA TABLE* in The Goodfellow Catalogue, 1993/94 edition, published by Goodfellow, Cambridge, England, pp. 17-31 (1993).
2. K. E. Petersen, *SILICON AS A MECHANICAL MATERIAL*, Proceedings of the IEEE **Vol. 70**, No. 5, pp. 421-457 (1982).
3. S. M. Sze, *VLSI TECHNOLOGY*, published by McGraw-Hill Book Company, New Jersey (1988).
4. J. C. Brice, *CRYSTAL GROWTH PROCESSES*, published by Wiley, New York (1986).
5. S. W. Kuan, C. W. Frank, C. C. Fu, D. R. Allee, P. Maccagno and R. F. W. Pease, *ULTRATHIN POLYMER FILMS FOR MICROLITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2274-2279 (1988).
6. B. Singh, S. P. Beaumont, P. G. Bower and C. D. W. Wilkinson, *NEW INORGANIC ELECTRON RESIST SYSTEM FOR HIGH RESOLUTION LITHOGRAPHY*, Applied Physics Letters **Vol. 41,** No. 9, pp. 889-891 (1982).
7. A. N. Broers, *RESOLUTION LIMITS FOR ELECTRON-BEAM LITHOGRAPHY*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 502-513 (1988).
8. J. Pacansky and J. R. Lyerla, *PHOTOCHEMICAL DECOMPOSITION MECHANISMS FOR AZ-TYPE PHOTORESISTS*, IBM Journal of Research and Development **Vol. 23,** No. 1, pp. 42-55 (1979).
9. J. O. Choi, J. A. Moore, J. C. Corelli, J. P. Silverman and H. Bakhru, *DEGRADATION OF POLY(METHYLMETHACRYLATE) BY DEEP ULTRAVIOLET, X-RAY, ELECTRON BEAM, AND PROTON BEAM IRRADIATIONS*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2286-2289 (1988).
10. R. F. W. Pease, *ELECTRON BEAM LITHOGRAPHY*, Contemporary Physics **Vol. 22,** No. 3, pp. 265-290 (1981).
11. C. D. W. Wilkinson and S. P. Beaumont, *ELECTRON BEAM NANOLITHOGRAPHY*, published by Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1980-1993).
12. M. Hatzakis, *MATERIALS AND PROCESSES FOR MICROSTRUCTURE FABRICATION*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 441-453 (1988).

13. K. G. Chiong, S. Wind and D. Seeger, *EXPOSURE CHARACTERISTICS OF HIGH-RESOLUTION NEGATIVE RESISTS*, Journal Of Vacuum Science & Technology B **Vol. 8,** No. 6, pp. 1447-1453 (1990).

14. A. Tritchkov, R. Jonckheere and L. Vandenhove, *USE OF POSITIVE AND NEGATIVE CHEMICALLY AMPLIFIED RESISTS IN ELECTRON-BEAM DIRECT-WRITE LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 2986-2993 (1995).

15. Y. M. N. Namaste, S. K. Obendorf and F. Rodriguez, *REACTIVE PLASTICIZERS IN NEGATIVE ELECTRON BEAM RESISTS*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2245-2249 (1988).

16. R. K. Watts, *Chapter 4: LITHOGRAPHY* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 141-183 (1988).

17. T. Azuma, T. Sato and H. Aochi, *EFFECT OF THIN FILM INTERFERENCE ON PROCESS LATITUDE IN DEEP ULTRAVIOLET LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 2928-2933 (1995).

18. M. L. Schattenburg, R. J. Aucoin and R. C. Fleming, *OPTICALLY MATCHED TRILEVEL RESIST PROCESS FOR NANOSTRUCTURE FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 3007-3011 (1995).

19. K. Ronse, M. Opdebeeck and L. Vandenhove, *FUNDAMENTAL PRINCIPLES OF PHASE SHIFTING MASKS BY FOURIER OPTICS: THEORY AND EXPERIMENTAL VERIFICATION*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 2, pp. 589-600 (1994).

20. F. A. Jenkins and H. E. White, *Section 15.10: RESOLVING POWER OF A MICROSCOPE* in Fundamentals of Optics, 4th edition, published by McGraw-Hill, Singapore, pp. 332-334 (1981).

21. A. N. Broers, *RESOLUTION, OVERLAY AND FIELD SIZE FOR LITHOGRAPHY SYSTEMS*, IEEE Transactions On Electron Devices **Vol. ED-28,** No. 11, pp. 1268-1278 (1981).

22. G. Owen, R. F. W. Pease, D. A. Markle, A. Grenville, R. L. Hsieh, R. Vonbunau and N. I. Maluf, *1/8 UM OPTICAL LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 3032-3036 (1992).

23. M. Rothschild, R. B. Goodman, M. A. Hartney, M. W. Horn, R. R. Kunz, J. H. C. Sedlacek and D. C. Shaver, *PHOTOLITHOGRAPHY AT 193 NM*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 2989-2996 (1992).

24. A. Grenville, G. Owen and R. F. W. Pease, *CHARECTERIZATION OF A 193 NM OPTICAL LITHOGRAPHY SYSTEM FOR 0.18 UM AND BELOW*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 6, pp. 3814-3819 (1994).

25. A. N. Broers, *LONG LIFE LAB_6 CATHODES*, Journal Of Applied Physics **Vol. 38,** p. 1991 (1967).

26. F. J. Hohn, T. H. P. Chang, A. N. Broers, G. S. Frankel, E. T. Peters and D. W. Lee, *FABRICATION AND TESTING OF SINGLE-CRYSTAL LANTHANUM HEXABORIDE ROD CATHODES*, Journal Of Applied Physics **Vol. 53,** No. 3, pp. 1283-1296 (1982).

27. D. Tuggle, L. W. Swanson and J. Orloff, *APPLICATION OF A THERMAL FIELD EMISSION SOURCE FOR HIGH RESOLUTION HIGH CURRENT E-BEAM MICROPROBES*, Journal Of Vacuum Science & Technology **Vol. 16,** p. 1699 (1979).

28. F. J. Hohn, *ELECTRON-BEAM LITHOGRAPHY: ITS APPLICATIONS*, Journal Of Vacuum Science & Technology B **Vol. 7,** No. 6, pp. 1405-1411 (1989).

29. W. Liu, J. Ingino and R. F. Pease, *RESIST CHARGING IN ELECTRON BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 5, pp. 1979-1983 (1995).

30. T. H. P. Chang, D. P. Kern, E. Kratschmer, K. Y. Lee, H. E. Luhn, M. A. McCord, S. A. Rishton and Y. Vladimirsky, *NANOSTRUCTURE TECHNOLOGY*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 462-493 (1988).

31. S. P. Beaumont, P. G. Bower, T. Tamamura and C. D. W. Wilkinson, *SUB-20-NM-WIDE METAL LINES BY ELECTRON-BEAM EXPOSURE OF THIN*

POLY(METHYLMETHACRYLATE) FILMS AND LIFTOFF, Applied Physics Letters **Vol. 38,** No. 6, pp. 436-439 (1981).

32. Y. W. Yau, R. F. W. Pease, A. A. Iranmanesh and K. J. Polasko, *GENERATION AND APPLICATIONS OF FINELY FOCUSED BEAMS OF LOW-ENERGY ELECTRONS*, Journal Of Vacuum Science & Technology **Vol. 19,** No. 4, pp. 1048-1052 (1981).

33. Y. H. Lee, R. Browning, N. Maluf, G. Owen and R. F. W. Pease, *LOW VOLTAGE ALTERNATIVE FOR ELECTRON BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 3094-3098 (1992).

34. M. A. McCord and T. H. Newman, *LOW VOLTAGE, HIGH RESOLUTION STUDIES OF ELECTRON BEAM RESIST EXPOSURE AND PROXIMITY EFFECT*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 3083-3087 (1992).

35. P. A. Peterson, Z. J. Radzimski, S. A. Schwalm and P. E. Russell, *LOW-VOLTAGE ELECTRON BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 3088-3093 (1992).

36. T. H. P. Chang, *PROXIMITY EFFECTS IN ELECTRON-BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology **Vol. 12,** p. 1271 (1975).

37. S. A. Rishton, S. P. Beaumont and C. D. W. Wilkinson, *MEASUREMENT OF THE PROFILE OF FINELY FOCUSED ELECTRON BEAMS IN A SCANNING ELECTRON MICROSCOPE*, Journal of Physics E: Scientific Instrumentation **Vol. 17,** p. 296 (1984).

38. D. R. S. Cumming, S. Thoms, S. P. Beaumont and J. M. R. Weaver, *FABRICATION OF 3 NM WIRES USING 100 KEV ELECTRON BEAM LITHOGRAPHY AND POLY(METHYLMETHACRYLATE) RESIST*, Applied Physics Letters **Vol. 68,** No. 3, pp. 322-324 (1996).

39. T. H. P. Chang and D. P. Kern, *ARRAYED MINIATURE ELECTRON BEAM MICROCOLUMNS FOR HIGH THROUGHPUT SUB-100 NM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 2743-2748 (1992).

40. E. Kratschmer, H. S. Kim, M. G. R. Thomson, K. Y. Lee, S. A. Rishton, M. L. Yu and T. H. P. Chang, *SUB-40 NM RESOLUTION 1 KEV SCANNING TUNNELING MICROSCOPE FIELD-EMISSION MICROCOLUMN*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 6, pp. 3503-3507 (1994).

41. E. Kratschmer, H. S. Kim, M. G. R. Thomson, K. Y. Lee, S. A. Rishton, M. L. Yu and T. H. P. Chang, *AN ELECTRON-BEAM MICROCOLUMN WITH IMPROVED RESOLUTION, BEAM CURRENT, AND STABILITY*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 2498-2503 (1995).

42. B. H. Koek, *'MIX AND MATCH' LITHOGRAPHY*, published by Philips E-beam Lithography Department, Eindhoven, The Netherlands (1980-1993).

43. M. G. Rosenfield, M. G. R. Thomson, P. J. Coane, K. T. Kwietniak, J. Keller, D. P. Klaus, R. P. Volant, C. R. Blair, K. S. Tremaine, T. H. Newman and F. J. Hohn, *ELECTRON-BEAM LITHOGRAPHY FOR ADVANCED DEVICE PROTOTYPING: PROCESS TOOL METROLOGY*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2615-2620 (1993).

44. H. I. Smith, *100 YEARS OF X RAYS: IMPACT ON MICRO- AND NANOFABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 2323-2328 (1995).

45. J. Warlaumont, *X-RAY LITHOGRAPHY: ON THE PATH TO MANUFACTURING*, Journal Of Vacuum Science & Technology B **Vol. 7,** No. 6, pp. 1634-1641 (1989).

46. M. A. McCord, A. Wagner and T. Donohue, *RESOLUTION LIMITS AND PROCESS LATITUDE OF X-RAY MASK FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2958-2963 (1993).

47. H. I. Smith and M. L. Schattenburg, *X-RAY LITHOGRAPHY FROM 500 TO 30 NM: X-RAY NANOLITHOGRAPHY*, IBM Journal of Research and Development **Vol. 37,** No. 3, pp. 319-328 (1993).

48. C. H. Mastrangelo and W. C. Tang, *Chapter 2: SEMICONDUCTOR SENSOR TECHNOLOGIES* in Semiconductor Sensors, edited by S. M. Sze, published by John Wiley & Sons, New York, pp. 17-95 (1994).

49. A. M. Haghiri-Gosnet, C. Vieu, G. Simon, F. Carcenac, A. Madouri, Y. Chen, F. Rousseaux and H. Launois, *FABRICATION OF SUB-30 NM MASKS FOR X-RAY LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 3066-3069 (1995).

50. B. G. Casey, W. Monaghan and C. D. W. Wilkinson, *EMBOSSING OF NANOSCALE FEATURES AND ENVIRONMENTS*, Microelectronic Engineering **Vol. 35,** pp. 393-396 (1997).

51. S. Y. Chou and P. R. Krauss, *IMPRINT LITHOGRAPHY WITH SUB-10 NM FEATURE SIZE AND HIGH THROUGHPUT*, Microelectronic Engineering **Vol. 35,** pp. 237-240 (1997).

52. M. A. McCord and R. F. W. Pease, *LIFT-OFF METALIZATION USING POLY(METHYLMETHACRYLATE) EXPOSED WITH A SCANNING TUNNELING MICROSCOPE*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 1, pp. 293-296 (1988).

53. H. Zhang, L. S. Hordon, S. W. J. Kuan, P. Maccagno and R. F. W. Pease, *EXPOSURE OF ULTRATHIN POLYMER RESISTS WITH THE SCANNING TUNNELING MICROSCOPE*, Journal Of Vacuum Science & Technology B **Vol. 7,** No. 6, pp. 1717-1722 (1989).

54. M. A. McCord and R. F. W. Pease, *THE EFFECT OF REFLECTED AND SECONDARY ELECTRONS ON LITHOGRAPHY WITH THE SCANNING TUNNELING MICROSCOPE*, Surface Science **Vol. 181,** pp. 278-284 (1987).

55. J. W. Gardner, *Chapter 4: SPECIALISED MATERIALS PROCESSING* in Microsensors Principles and Applications, published by John Wiley & Sons, Chichester, pp. 59-78 (1994).

56. H. Sawaragi, H. Kasahara and R. Aihara, *DEVELOPMENT OF A FOCUSED ION BEAM SYSTEM: CURRENCT STATUS AND FUTURE PROSPECTS*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 3, pp. 962-965 (1988).

57. K. M. Edenfeld, K. F. Jarausch, T. J. Stark, D. P. Griffis and P. E. Russell, *FORCE PROBE CHARACTERIZATION USING SILICON THREE-DIMENSIONAL STRUCTURES FORMED BY FOCUSED ION BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 6, pp. 3571-3575 (1994).

58. M. J. Vasile, C. Biddick and S. A. Schwalm, *MICROFABRICATION BY ION MILLING: THE LATHE TECHNIQUE*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 4, pp. 2388-2393 (1994).

59. H. W. P. Koops, R. Weiel and D. P. Kern, *HIGH-RESOLUTION ELECTRON-BEAM INDUCED DEPOSITION*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 1, pp. 477-481 (1988).

60. K. L. Lee and M. Hatzakis, *DIRECT ELECTRON-BEAM PATTERNING FOR NANOLITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 7,** No. 6, pp. 1941-1946 (1989).

61. M. Yamaki, T. Miwa, H. Yoshimura and K. Nagayama, *EFFICIENT MICROTIP FABRICATION WITH CARBON COATING AND ELECTRON BEAM DEPOSITION FOR ATOMIC FORCE MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 2447-2450 (1992).

62. P. B. Fischer, M. S. Wei and S. Y. Chou, *ULTRAHIGH-RESOLUTION MAGNETIC FORCE MICROSCOPE TIP FABRICATED USING ELECTRON-BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2570-2573 (1993).

63. D. W. Abraham, H. J. Mamin, E. Ganz and J. Clarke, *SURFACE MODIFICATION WITH THE SCANNING TUNNELING MICROSCOPE*, IBM Journal of Research and Development **Vol. 30,** No. 5, pp. 492-499 (1986).

64. M. A. McCord and R. F. W. Pease, *SCANNING TUNNELING MICROSCOPE AS A MICROMECHANICAL TOOL*, Applied Physics Letters **Vol. 50,** No. 10, pp. 569-570 (1987).

65. T. Schimmel, H. Fuchs, S. Akari and K. Dransfeld, *NANOMETER-SIZE SURFACE MODIFICATIONS WITH PRESERVED ATOMIC ORDER GENERATED BY VOLTAGE PULSING*, Applied Physics Letters **Vol. 58,** No. 10, pp. 1039-1041 (1991).

66. T. Schimmel, R. Kemnitzer, J. Kuppers, H. Fuchs and M. Luxsteiner, *WRITING AND DELETING ATOMIC-SCALE STRUCTURES WITH THE SCANNING TUNNELING MICROSCOPE*, Thin Solid Films **Vol. 254,** No. 1-2, pp. 147-152 (1995).

67. S. P. Murarka, *Chapter 9: METALLIZATION* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 375-421 (1988).

68. A. C. Adams, *Chapter 6: DIELECTRIC AND POLYSILICON FILM DEPOSITION* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 233-271 (1988).

69. C. W. Pearce, *Chapter 2: EPITAXY* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 55-97 (1988).

70. L. E. Katz, *Chapter 3: OXIDATION* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 99-140 (1988).

71. J. C. C. Tsai, *Chapter 7: DIFFUSION* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 272-326 (1988).

72. M. D. Giles, *Chapter 8: ION IMPLANTATION* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 327-374 (1988).

73. J. W. Gardner, *Chapter 3: CONVENTIONAL SILICON PROCESSING* in Microsensors Principles and Applications, published by John Wiley & Sons, Chichester, pp. 36-57 (1994).

74. S. Bengtsson, *SEMICONDUCTOR WAFER BONDING: A REVIEW OF INTERFACIAL PROPERTIES AND APPLICATIONS*, Journal of Electronic Materials **Vol. 21,** No. 8, pp. 841-862 (1992).

75. G. Wallis and D. I. Pomerantz, *FIELD ASSISTED GLASS-METAL SEALING*, Journal Of Applied Physics **Vol. 40,** No. 10, pp. 3946-3950 (1969).

76. K. B. Albaugh, *ELECTRODE PHENOMENA DURING ANODIC BONDING OF SILICON TO SODIUM BOROSILICATE GLASS*, Journal of the Electrochemical Society **Vol. 138,** No. 10, pp. 3089-3094 (1991).

77. M. Parameswaran, D. Xie and P. G. Glavina, *FABRICATION OF NICKEL MICROMECHANICAL STRUCTURES USING A SIMPLE LOW-TEMPERATURE ELECTROLESS PLATING PROCESS*, Journal of the Electrochemical Society **Vol. 140,** No. 7, pp. L111-L113 (1993).

78. W. Kern and C. A. Deckert, *CHEMICAL ETCHING* in <Unknown>, published by Academic Press, pp. 401-496 (1978).

79. R. Khare, E. L. Hu, J. J. Brown and M. A. Melendes, *MICROMACHINING IN III-V SEMICONDUCTORS USING WET PHOTOELECTROCHEMICAL ETCHING*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2497-2501 (1993).

80. R. Mlcak, H. L. Tuller, P. Greiff, J. Sohn and L. Niles, *PHOTOASSISTED ELECTROCHEMICAL MICROMACHINING OF SILICON IN HF ELECTROLYTES*, Sensors and Actuators A-Physical **Vol. 40,** pp. 49-55 (1994).

81. C. H. Ozdemir and J. G. Smith, *NEW PHENOMENA OBSERVED IN ELECTROCHEMICAL MICROMACHINING OF SILICON*, Sensors and Actuators A-Physical **Vol. 34,** pp. 87-93 (1992).

82. A. Soderbarg and L. Smith, *ELECTROCHEMICAL ETCH STOP OBTAINED BY ACCUMULATION OF FREE CARRIERS WITHOUT P-N JUNCTION*, Journal of the Electrochemical Society **Vol. 140,** No. 1, pp. 271-275 (1993).

83. R. J. Schutz, *Chapter 5: REACTIVE PLASMA ETCHING* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 184-232 (1988).

84. S. G. Ingram, *THE INFLUENCE OF SUBSTRATE TOPOGRAPHY ON ION BOMBARDMENT IN PLASMA ETCHING*, Journal Of Applied Physics **Vol. 68,** No. 2, pp. 500-504 (1989).

85. W. H. Juan and S. W. Pang, *HIGH-ASPECT-RATIO SI ETCHING FOR MICROSENSOR FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 3, pp. 834-838 (1995).

86. I. W. Rangelow and H. Loschner, *REACTIVE ION ETCHING FOR MICROELECTRICAL MECHANICAL SYSTEM FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 2394-2399 (1995).

CHAPTER 3

PROJECT OVERVIEW

This chapter explains the objective of the work of this thesis in more detail and presents a strategy for meeting these objectives. The basic choices are explained and the chosen methods are justified. The compromises necessary to go from the ideal sensor to a realistically manufacturable sensor, which is the aim of this work, are explained and justified along with the obvious limitations of the end result. The results of an initial feasibility study, carried out by a previous student as part of a final year project, are briefly summarised, and the implications of this on the work of this thesis are discussed. The first faltering steps towards the achievement of the objective are described and the effect of these initial results on shaping the program of work required in order to satisfy the project goals are presented. This chapter concludes with a list of the problems which required attention before the primary goals of this work could be achieved.

3.1 THE BESPOKE MICROSCOPY CONCEPT

The basic idea behind the project is based on the observation that most new scanning probe microscopy techniques consist of a high resolution 'nanosensor' scanned in close proximity to the specimen, yielding high resolution information by making use of the enhanced distance dependence of most interactions in the near-field. Ideally, the sensor-specimen distance should be kept constant during scanning. The best distance-regulation technique for this purpose is the one with the greatest flexibility and general applicability. This is atomic force microscopy, due to its ability to handle insulating and conducting materials, its reduced spectroscopic dependence and the availability of two regimes of regulation (contact and non-contact) which allow most samples to be viewed.

Hence it would seem that if a general technique was available for fabricating arbitrary nanosensors on the end of an AFM probe, whole new microscopy techniques would be available based on the scaling down of many of the microsensors and macrosensors used today. If this technique was sufficiently reliable, much more complex sensors based on measuring the near-field response of the specimen to an active probe in the near-field could be realised in addition to simply measuring the far-field response to an active probe in the near-field (for example scanning near-field optical microscopy) or the near-field response to an active stimulus in the far field. This idea gives rise to the concept of microscopy sensors tailor-made for specific applications, *i.e.* bespoke microscopy.

The choice of high resolution technique has been discussed in Chapter 2. The application of electron-beam lithography and focused ion-beam lithography to sensors for scanning probe microscopy is not new. Focused ion-beam lithography has been used to produce atomic force microscope probes with very high aspect ratio tips and extremely small radii of curvature [Vasile, Biddick and Huggins[1] (1994); Vasile, Biddick and Schwalm[2] (1994)] while electron-beam lithography has been used, in the technique known as contamination lithography, to produce very high aspect ratio 'supertips' on the end of conventional atomic force microscope tips [*e.g.* Yamaki, Miwa, Yoshimura and Nagayama[3] (1992)]. The latter are even commercially available and have even been customised for magnetic force microscopy by evaporating magnetic materials, such as nickel and chromium, onto these supertips at an oblique angle while the probes are rotated [Fischer, Wei and Chou[4] (1993)].

However, all the probes made using these techniques have relied on an operator locating the apex of the atomic force microscope probe, using low intensity electron or ion-beam exposure, and then carrying out the lithography required. This is a manual alignment process, clearly unsuitable for batch fabrication because of its time consuming nature. To develop a truly volume scaleable process requires the use of automatic alignment schemes, so that the apices can be automatically exposed in a reproducible and controllable manner. This necessitates the use of a substrate with alignment markers

which are capable of sufficient accuracy. Often the manual alignment techniques have used a commercially available substrate of atomic force microscope probes as their starting point [*e.g.* Gortz, Kempf and Kretz[5] (1995)]. Unfortunately these are devoid of the necessary alignment markers. In addition, the use of electron-beam lithography or focused ion-beam lithography in a production process requires the equipment to be capable of patterning over wide areas with automatic calibration and correction – requirements which are fulfilled by commercially available direct-write and mask-writing machines, as opposed to the converted microscopes employed by most of those using such lithographic techniques for SPM probe fabrication. The availability of such an expensive and capable machine at the University of Glasgow, along with the necessary ancillary facilities and experience, are amongst the most important motivations for carrying out the development of a flexible probe manufacturing technique here.

3.2 THE BASIC IDEA

The basic idea underlying this project is to use standard atomic-force microscope probes as a substrate upon which electron-beam lithography can be carried out to place high resolution nanosensors on the apices of tips for bespoke microscopy applications. To achieve this requires accurate automatic alignment of the electron-beam lithography machine to these apices. Since commercially available probes supplied on a whole wafer basis do not have the necessary high resolution alignment markers, the in-house manufacture of suitable substrates becomes a prerequisite for this work.

Commercially, atomic force microscope probes are usually batch fabricated using a combination of photolithography, wet etching, dry etching and wafer bonding [*e.g.* Akamine, Barrett and Quate[6] (1990); Albrecht, Akamine, Carver and Quate[7] (1990); Wolter, Bayer and Greschner[8] (1991)] – this fabrication process is one of a set of related techniques known as *micromachining*. The probes consist of thin silicon or silicon nitride cantilevers with sharp, high aspect ratio tips at their ends, and are available with a range of spring constants from 0.01N/m to 100N/m and resonant frequencies from 1kHz to 500kHz, making them suitable for the various atomic force microscopy techniques and applications. Typical cantilever dimensions are of the order of tens of microns in width, up to a few microns thick and several hundred microns in length. The tips can range from a few microns to a few tens of microns tall, can have cone angles of 10° or less and a tip radius of curvature of less than 100Å. They can be coated with magnetic materials for use in AFM, or with metals for use in electric AFM and combined AFM/STM techniques. The probes are typically available in monolithic form (*i.e.* machined out of a single piece of substrate) or as a thinner machined layer wafer bonded to a plastic such as perspex or a more rigid material such as silicon or glass. The dimensions of the substrate holding the cantilever is usually a few millimetres in width and a few millimetres more in length with a thickness of some hundreds of microns. These probes may be sold individually or

in batches of up to a whole wafer. The whole wafer is usually 3 to 4 inches in diameter and can hold as many as 500 probes. These probes are designed to be compatible with many of the commercially available atomic force microscopes.

Initially, the process developed in this work was to be used to produce three specific kinds of probes based on a generic fabrication technology: (i) a contact mode AFM probe coated with a metal film in which a sub-micron aperture has been defined at the apex of the tip for use in scanning near-field optical microscopy; (ii) a contact mode AFM probe upon which a thin film thermocouple has been defined, with its sub-micron crossover at the apex of the tip, and; (iii) a non-contact mode AFM probe upon which a thin film Hall-probe has been defined, with its sub-micron cross at the apex of the tip. Although, many other, and more complex, probes ought to be possible, these were perceived to be some of the simplest and would serve to define the constraints and capabilities required of the final process.

It was already clear that certain compromises had to be made. It is not practical to fabricate a thin film thermocouple or Hall probe at the apex of a tip with a cone angle of only 10° due to the technical difficulties involved in patterning electrical connections on such a steep surface – a cone angle of at least 60° is probably desirable. In addition, repeatable automatic alignment to the apex of a tip with a radius of curvature of 100Å or less is not within the capabilities of current commercially available electron-beam lithography machines, even assuming the presence of ideal alignment markers. In any case, the errors in processing mean that alignment accuracies of the order of a few hundred nanometres is more reasonable, at least in the first instance.

Thus the resolution of the simultaneously acquired force image may have to be reduced by the necessity of using blunt atomic force microscopy tips, and access to the specimen will certainly be reduced for samples which have rapidly varying topography. However, as the force image is used as an indicator of the approximate lateral position of the sensor upon the sample, and mainly used as the regulator of the probe-sample separation, with the principal measurement being that of the sensor, this does not seriously detract from the capabilities of the probe. The need for blunt tips with large cone angles meant that a custom designed AFM probe substrate was necessary as these attributes are not available, or even desirable, in commercial AFM probes due to their intended application: high resolution AFM with good access to the specimen. The substrate required for the work in this thesis is made by using micromachining techniques similar to those employed commercially as well as in the references at the beginning of this section, but with different design constraints and criteria.

3.3 MICROMACHINING

Micromachining is the name given to the processing used to create geometrical devices on a micron, and increasingly, sub-micron scale. These range from pressure sensors and

strain-gauges to accelerometers, resonators and micromachines and microactuators [Wise and Najafi[9] (1991); Bryzek, Petersen and McCulley[10] (1994)]. Most micromachining is carried out on silicon although some is also carried out on other semiconductor substrates, such as gallium arsenide and silicon-germanium. Micromachining is usually split into two broad categories: surface micromachining and bulk micromachining.

In bulk micromachining, the required device is formed from a large single crystal substrate, usually silicon. Techniques such as wet etching, and in particular, crystallographic wet etching are used to shape the substrate to produce the required geometry [Wise, *et al.*[9] (1991); Gardner[11] (1994); Gardner[12] (1994); Mastrangelo and Tang[13] (1994)]. These are often integrated with electronics by additional fabrication on layers deposited on the substrate. One of the key advantages of this method is that it is inexpensive, well-developed and very reliable due to its dependence on single crystal substrates. Its main disadvantage is that the crystallographic etching requires patterns much bigger than the resultant devices due to the dependence on crystal planes. This leads to a lower device density and bulkier devices. In addition, the complexity of devices made using this technique is limited to single or double layered devices because of the use of etchants which are corrosive to most other commonly used materials.

In surface micromachining [Howe[14] (1988)], the required device is formed by patterning thin layers on a solid substrate, usually silicon but not necessarily so, using mainly dry etching techniques along with some wet etching [Wise, *et al.*[9] (1991); Gardner[11] (1994); Gardner[12] (1994); Mastrangelo, *et al.*[13] (1994)]. The substrate is merely used as a flat surface and as a convenient way of handling the resulting devices. One of the key aspects of this technique are the use of *sacrificial layers* to produce suspended structures *i.e.* a deposited layer which can be removed by etching through undercutting, thereby releasing a patterned layer deposited above, which is resistant to the etch, rather than having to etch through the whole substrate as is often necessary to achieve a similar result with bulk micromachining. Its main advantages are that dry etching allows greater device densities to be achieved and the surface micromachining scheme can be used to make very complex multi-layered devices by repeating deposition and patterning steps *ad infinitum*. In addition, the avoidance of the use of corrosive etching techniques allows a greater degree of process compatibility with the best materials for each function. Its key disadvantages are that it is generally more expensive than bulk micromachining, often requires more steps and relies on the patterning of amorphous or polycrystalline layers which have less well defined properties and are more sensitive to deposition conditions and post-processing than the single crystal substrates used in bulk micromachining.

As always, there are hybrid processes available, with one of the more significant techniques being that of LIGA (a German acronym for Lithographie, Galvanoformung, Abformung) [Mastrangelo, *et al.*[13] (1994)] in which very deep patterns in resist are produced using *x*-ray lithography and are used as masking for various fabrication steps such as plating and dry etching. The result of this is that single crystal substrates, as

used in bulk micro-machining, can be used to produce structures using fabrication techniques normally associated with surface micromachining, such as dry etching and sacrificial layers. Micromachines and microactuators are often produced in this way [*e.g.* Borner, *et al.*[15] (1996)].

Industrially, bulk micromachining is still the most commonly used technique, although elements of surface micromachining are beginning to creep into many processes. A case in point is that of the micromachining of silicon for producing atomic force microscope probes. To shorten the time required to release the cantilever (this normally requires etching all the way through the substrate), it is common to use relatively thin silicon substrates which are wafer-bonded onto other substrates, such as glass or perspex, for handling reasons. For simplicity and due to the lack of experience in the department with wafer bonding techniques at the time this work was carried out (although Brendan Casey has since been experimenting with wafer bonding techniques in other, unrelated work), the generic substrate was designed using traditional bulk micromachining techniques, employing heavy use of silicon wet etching techniques.

3.4 SILICON WET ETCHING

As discussed in §2.7.3, wet etching can be isotropic or anisotropic. This section is concerned specifically with the wet etching of silicon. This technique has been developed over the last forty years with rapid progress having been made in the last twenty years or so [Bean[16] (1978); Petersen[17] (1982)].

3.4.1 Isotropic Silicon Wet Etching

The isotropic etching of silicon is typically carried out with a solution consisting of a mixture of oxidising and deoxidising agents in a suitable solvent. The most common solutions are made up of combinations of nitric acid (oxidising agent), hydrofluoric acid (deoxidising agent) and acetic acid (complexing agent). The nitric acid oxidises the silicon surface, while the hydrofluoric acid removes the oxidised layer.

The behaviour of this etch varies dramatically with the exact ratios of each of these, with some combinations, such as 2 HNO_3:1 HF:1 CH_3COOH being completely isotropic [Kern[18] (1978); Petersen[17] (1982)], another combination, 3 HNO_3:1 HF:8 CH_3COOH, selectively etches n^+ or p^+ doped areas rather than lightly doped regions [Robbins and Schwarz[19] (1959); Robbins and Schwarz[20] (1960)], and yet another combination, 6 HNO_3:3 HF:8 CH_3COOH, etches more crystallographically [Muraoka, Ohhashi and Sumitomo[21] (1973)]. This variation in behaviour is determined by whether the reaction is diffusion limited or rate limited [Bogenschutz, Krusemark, Locherer and Mussinger[22] (1967)]. With relatively low hydrofluoric acid concentrations the silicon surface is completely isotropically oxidised as the oxide is attacked relatively slowly. With relatively low nitric acid concentrations the crystallographic dependence of oxide formation

combined with a relatively rapid attack of the oxide produces more crystallographic results. These etches are very fast, etching at rates up to 50 μm/min in certain compositions, and are strongly dependent on temperature. This makes them very difficult to control accurately at high etching rates so that the etching is often carried out at low temperatures to slow the reaction to a more practical rate [Theunissen, Appels and Verkuylen[23] (1970)].

The other issue is that of the masking material used. Silicon oxide can only resist this etch for a short time with the edges of the mask becoming seriously eroded so that silicon nitride or gold masks are often necessary, although masking can also be provided by some, usually negative, photoresists, for a short time Thus this etch is normally only used for polishing and planarising surfaces, where no masking is required, rather than micromachining. However, the isotropic nature of these etches have been used to produce high aspect ratio silicon tips for atomic force microscopy [Brugger, Buser and Derooij[24] (1992); Indermuehle, *et al.*[25] (1994)], but as a high aspect ratio is not desirable for the purposes of this work, these etches were avoided in favour of crystallographic silicon etches.

3.4.2 Crystallographic Silicon Wet Etching

Crystallographic wet etching is one of the most commonly used bulk micromachining processes. The geometries achieved with this method are based on the fact that these etches attack the {111} planes much more slowly than the {110} and {100} planes [Bean[16] (1978); Petersen[17] (1982)]. In silicon the {111} planes are the most densely populated planes and act as a reasonably effective etch stop with most crystallographic etches. Heavily boron doped silicon layers can also act as a very effective etch stop, and this can be used to advantage in the fabrication of certain devices. All these etches tend to be alkaline in nature and can be divided into organic and inorganic etches.

The most commonly used organic etch is a mixture of ethylene diamine ($NH_2(CH_2)_2NH_2$), pyracatechol ($C_6H_4(OH)_2$) and water, often known as EDP [Finne and Klein[26] (1967); Reisman, *et al.*[27] (1979)]. As with the isotropic etches described above, different mixture ratios can produce different results. This etch can also contain a small quantity of pyrazine ($C_4H_4N_2$) to alter the etching characteristics. Two particular mixtures of note are EDP type-S and EDP type-F [Reisman, *et al.*[27] (1979)], the former is a slower etch which works well over a temperature range of 50°C to 115°C, while the latter only works well at 115°C. Other organic etches include hydrazine ($N_2H_4 \cdot H_2O$), ammonium hydroxide (NH_4OH), choline ($(CH_3)_3N(CH_2OH)OH$) and, more recently, tetramethylammonium hydroxide or TMAH ($(CH_3)_4NOH$) [Declercq, Gerzberg and Meindl[28] (1975); Asano, Cho and Muraoka[29] (1976); Mehregany and Senturia[30] (1988); Schnakenberg, Benecke and Lochel[31] (1990); Schnakenberg, Benecke and Lange[32] (1991); Schnakenberg, *et al.*[33] (1991); Tabata, *et al.*[34] (1992); Gajda, Ahmed, Shaw and Putnis[35] (1994)].

These etches have various advantages and disadvantages and the choice of etchant is usually dependent on the specific application. However, it is generally the case that organic crystallographic etches are less corrosive and are therefore more likely to be compatible with other fabrication levels. They usually attack popular masking materials, such as silicon oxide, at a much slower rate in comparison with inorganic etches and result in less contamination of the wafer with metal ion impurities [Mastrangelo, *et al.*[13] (1994)].

The most commonly used inorganic etches are aqueous solutions of potassium hydroxide (KOH) and sodium hydroxide (NaOH) [Pugacz-Muraszkiewicz[36] (1972)], sometimes with the addition of isopropanol ($CH_3CH(OH)CH_3$) to alter the etching characteristics [Price[37] (1973)]. Their main advantages are that they can often achieve faster etching rates, are safer and easier to dispose of and often etch the {110} planes faster than the {100} planes, making them attractive in the machining of (110) oriented wafers [Kendall[38] (1975)]. The main disadvantages are that the etch rate of many masking materials is usually higher with these etches than with organic etches, and, particularly with potassium and sodium hydroxide solutions, the silicon can become contaminated with mobile metal ions. Other group I metal hydroxides which can be used to etch silicon include lithium hydroxide (LiOH), caesium hydroxide (CsOH) and rubidium hydroxide (RbOH).

The exact mechanism by which these solutions etch silicon is not well understood although several mechanisms have been proposed [see Petersen[17] (1982) for a discussion]. However, detailed discussions of these are beyond the scope of this thesis. For most of the etches mentioned above, the etch rate, R (in units of distance/time), generally follows the Arrhenius law [Robbins, *et al.*[19] (1959); Bogenschutz, *et al.*[22] (1967); Lee[39] (1969)]:

$$R(T) = R_o \exp\left(-\frac{E_a}{k_B T}\right), \tag{3.4.1}$$

where R_o is a crystal plane dependent rate constant, k_B is the Boltzmann constant, T is the temperature and E_a is a crystal plane dependent activation energy. This exponential dependence on temperature means that at higher temperatures, the level of etch depth control decreases. Whilst much etch rate data has been published in the literature, in practice, the exact etch rate is better found by experiment due to the sensitivity of the etch rate to environmental and equipment parameters, such as temperature calibration, stirring rate, wafer orientation in the solution and wafer doping type and concentration.

3.4.3 Bulk Micromachined Structures

In general, two orientations of wafer are commonly used in bulk micromachining applications: <110> and <100> oriented wafers. Since the {111} planes act as etch stop planes, the orientation of wafer required for a particular structure is determined by the positions of these 4 planes with respect to the wafer surface.

In the case of a <110> oriented wafer, two of the {111} planes, the (1-11) and (1-1-1) planes, are located at right angles to the surface while the remaining planes, the (111) and (11-1) planes are located at an angle of arctan(1/√2), or about 35.26°, to the surface. The intersections of these planes with the wafer surface are asymmetrically arranged with respect to each other. The vertically oriented planes mean that (110) oriented wafers are useful for structures which require deep vertical trenches [Kendall[38] (1975)].

In the case of a <100> oriented wafer, all four {111} planes are located at an angle of arctan(√2), or about 54.74°, to the surface. The intersection of these planes with the wafer surface are symmetrically arranged at right angles with respect to each other, forming a square on the wafer surface. This means that the {111} planes would form a pyramidal structure with a square base and slopes at an angle of about 54.74°. This is ideal for the fabrication of low aspect ratio tips suitable for the purposes of the work in this thesis. This geometry also makes <100> wafers ideal for the formation of square membranes, such as those used in pressure sensors and inkjet nozzles [Bassous[40] (1978); Bassous and Baran[41] (1978)].

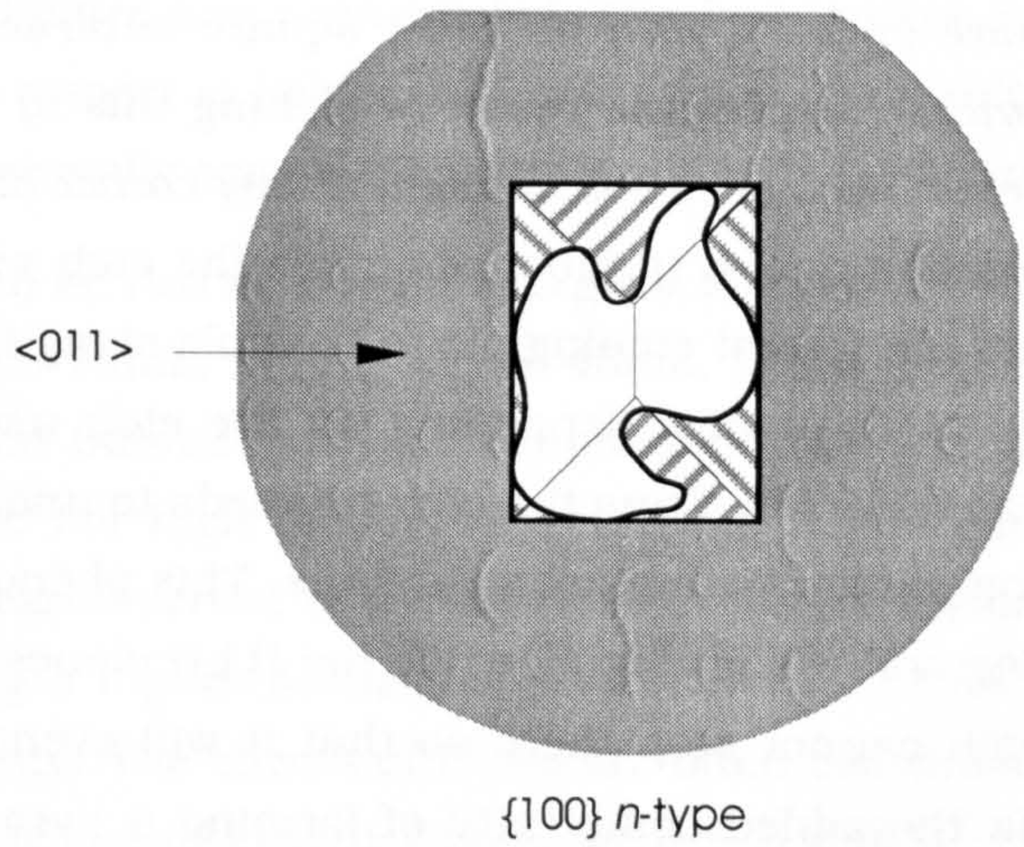


Figure 3.4.1: Illustration of the final result of anisotropically wet etching concave corners in silicon. The opening here is the white region. The etch undercuts the opening in the region shown by grey stripes. Any arbitrary opening in a {100} silicon wafer ends up as a rectangular-based inverted pyramid bounded by its [011] direction extremities.

An interesting aspect of the crystallographic etching of silicon is the difference between the structures obtained when etch masks containing convex corners are used in comparison with those when concave corners are used. As <100> oriented wafers are exclusively used for the work in this thesis, only this orientation will be considered here. Figure 3.4.1 illustrates how an arbitrarily shaped opening in a masking material eventually forms a rectangular-based pyramidal etch pit. The etch virtually stops because the etching of {111} planes, *i.e.* the breaking of Si-Si bonds on the {111} planes, is much more difficult than the other planes [Hesketh, *et al.*[42] (1993)]. An opening consisting of concave corners aligned along the intersections of the {111} planes with the (100) surface, *i.e.* along the [110] directions, can thus be easily made to produce this behaviour.

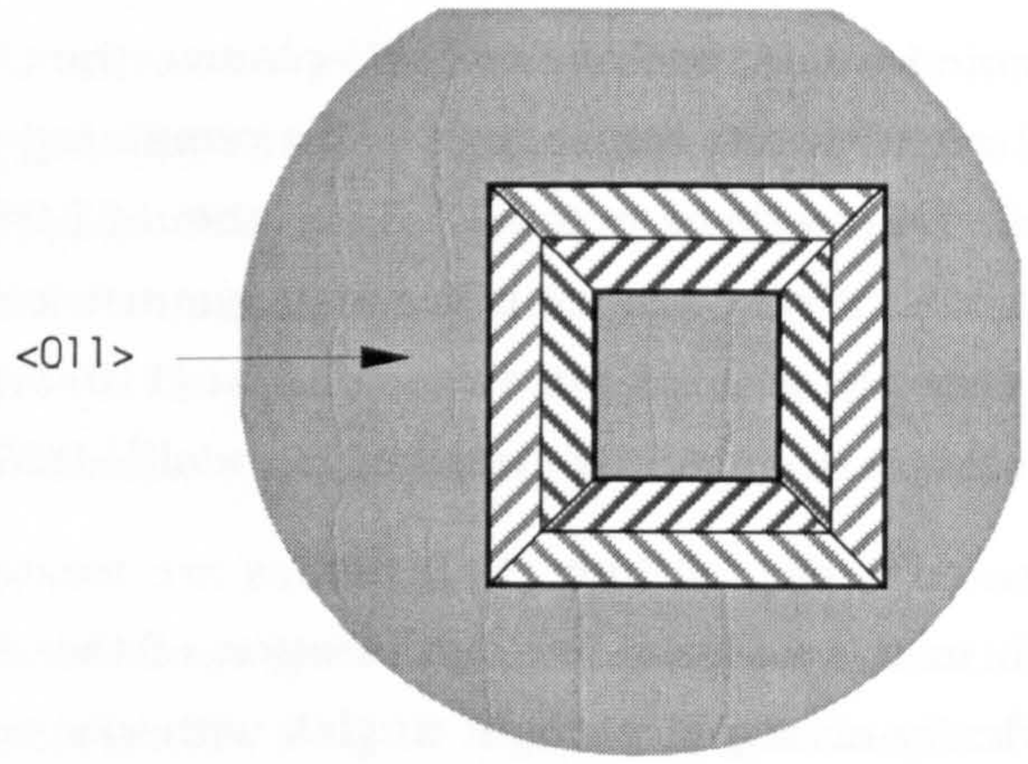


Figure 3.4.2: Illustration of the mistakenly envisaged result of anisotropically wet etching convex corners in silicon. The opening is square with an inset square subtracted from its centre. One would expect the structure above to result from this with the etch terminating in v-grooves formed by {111} planes. Unfortunately, convex corner undercutting means that this is not the case. The final result would be the complete undercutting of the inset mesa. The sloped edges are shown as stripes above .

Now consider the situation where convex corners are exposed along these very same directions. Consider a square opening with an inset square subtracted out from its centre. Then one would normally expect the result of etching this to be the structure shown in Figure 3.4.2. However, because a silicon atom at any corner of the inset square can be considered to be part of any of a number of planes the etch will behave as if a corner silicon atom is part of the fastest etching plane possible along these axes. These are apparently the {212} or {313} planes, depending on the etch used [Lee[39] (1969); Declercq, *et al.*[28] (1975); Bean[16] (1978)]. Thus the etch proceeds to undercut the corners of the mask, resulting in non-square pyramidal structures. This phenomenon is known as *convex corner undercutting* and will gradually erode the {111} planes at the surface to extinction. However, the etch cannot stop there so that it will eventually completely undermine the mask. Thus the added complexity of forming a pyramid on a silicon surface is that an etch stop is not possible and a timed etch must be used. As there are eight faster etching planes intersecting the (100) surface, along the [210] or [310] directions, the result of a timed etch will be a dodecagon-based pyramidal structure which is quite suitable for the tips required as part of the generic substrate. This will be discussed in more detail in §4.3.1 (also see Appendix J, §J.1).

3.5 EBL ON THE EBPG-HR5 (BEAMWRITER)

The EBPG-HR5 is the principal electron-beam lithography machine used for the work in this thesis. The EBPG-HR5 was bought from Phillips originally, and is one of their BeamWriter™ series of electron-beam lithography machines, but the electron-beam division was subsequently acquired by Leica-Cambridge. It offers higher specifications than its predecessor, the EBPG-4 [Koek[43] (1980-1993)].

3.5.1 Description of the EBPG-HR5 (BeamWriter)

The EBPG-HR5, which goes by the full name of electron-beam pattern generator - high resolution 5, is a commercial mask-writing machine with optional extras giving it high resolution, direct-write capability. The *beamwriter*, as it is normally dubbed, is a vector scanned machine with a gaussian shaped beam. It features automatic focus, alignment, rotation correction, gain correction, spot-size measurement, beam-current regulation and measurement, astigmatism correction, thermal drift measurement and many other compensation mechanisms. It is located in a Class 10 cleanroom in a temperature controlled environment and features multiple wafer writing capability via sample holders which are inserted into a cartridge as part of a load-lock (fast-entry) mechanism, so that many runs can be carried out between loads. The maximum substrate size it can handle is a 6 inch wafer with a 5.5 inch diameter writeable area. Two mask holders allow the writing of 2.5 inch and 4 inch masks, the latter being ideally suited for contact photolithography on 3 inch wafers. The thermal drift is specified to be better than 100nm/hour but 40nm/hour can be achieved if two to four hours of settling time is allowed after the holder has been loaded. The beamwriter periodically recalibrates for thermal drift (with respect to holder markers), which by default is every fifteen minutes. A full recalibration normally occurs every thirty minutes by default.

The beamwriter can be run at a beam energy of 20kV, 50kV and 100kV. It offers a range of spot sizes (12nm, 15nm, 20nm, 28nm, 40nm, 56nm, 80nm, 112nm, 160nm, 224nm, 300nm, 400nm and 800nm) with a beam-current which is dependent on the spot size and aperture area. For high resolution operation, it is normal to use a 12nm spot at 100kV giving a beam current of about 100pA using a 200μm diameter aperture. The advantage of using a higher beam energy is the reduced forward scattering and lower backscatter intensity. The effects of these are often conveniently represented using the following equation [Chang[44] (1975); Chang, *et al.*[45] (1988)]:

$$\varepsilon(r) = \frac{1}{\pi(1+\eta)}\left[\frac{1}{\beta_f^{\,2}}\exp\left(-\frac{r^2}{\beta_f^{\,2}}\right) + \frac{\eta}{\beta_b^{\,2}}\exp\left(-\frac{r^2}{\beta_b^{\,2}}\right)\right], \tag{3.5.1}$$

where $\varepsilon(r)$ is the energy density at a distance r from the centre of a very narrow electron beam, β_f and β_b are the characteristic half widths of the forward and backscattered beams, respectively, and η is the ratio of the backscattered to forward scattered beam energies. Some typical values for 25kV and 50kV on silicon and gallium arsenide [from Chang, *et al.*[45] (1988)] are plotted in Figure 3.5.1. From this it can be seen that the backscatter beam intensity decreases upon increasing the beam energy from 25kV to 50kV in both silicon and gallium arsenide. The forward scatter width decreases with thinner resists while the backscatter width remains unaffected, but the backscatter width increases with beam energy, since electrons penetrate deeper on average before being scattered. The backscatter dose acts like a general background exposure level and reduces the effective contrast of fine resist features. Due to the reducing backscatter

intensity at higher beam energies, it is found that, in general, the dose required to expose a feature is proportional to the beam energy and that finer features are possible with higher beam energies Chang, *et al.*[45] (1988).

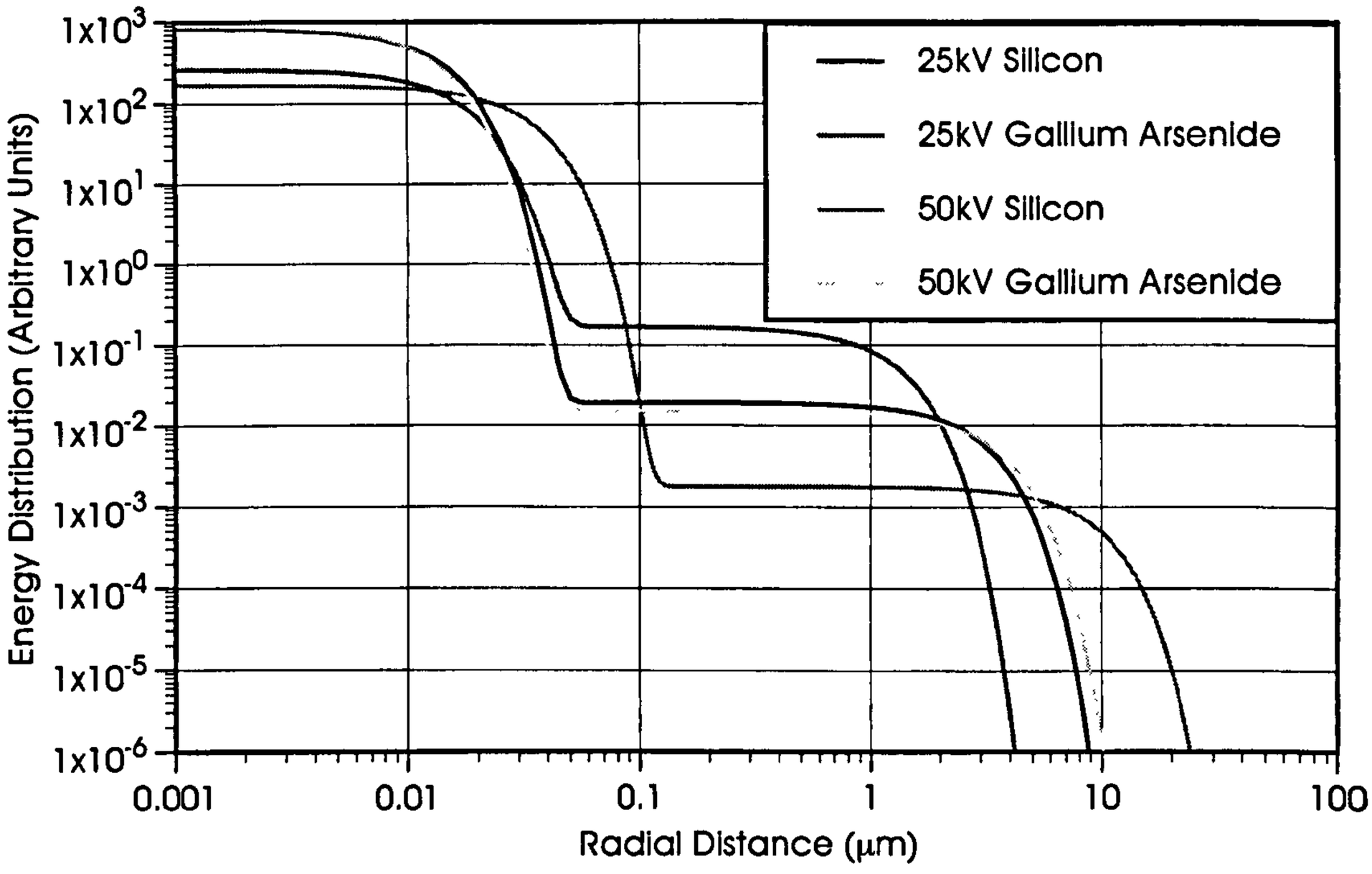


Figure 3.5.1: Comparison of the proximity effect between Silicon and Gallium Arsenide for 25kV and 50kV exposure. Data from Chang, *et al.*[45] (1988).

The current is proportional to the area of the aperture. The high writing rates are achieved by using a lanthanum hexaboride filament [Broers[46] (1967); Hohn, *et al.*[47] (1982)] and beam-deflection coils capable at operating over the range 500Hz to 10MHz pixels, each of which has a minimum size of 5nm. Bigger spot sizes are useful for writing large areas as they can be operated with a 400μm or 800μm aperture (the former being more commonly used) giving a beam current of up to 500nA at 50kV.

The spot shape is circular (gaussian beam) and is effectively an image of the filament limited by the aperture which reduces the spread of electrons allowed to go through the lenses. The aperture is located 40mm above the wafer holder, giving an effective *f*-number of 200 and numerical aperture of $2.5\text{x}10^{-3}$ with a 200μm aperture (from Equation 2.2.1). The de Broglie wavelength of electrons with kinetic energies which are a significant fraction of their rest mass energy can be calculated from [Sandin[48] (1989)]:

$$\lambda_d = \frac{hc}{pc} = \frac{hc}{\sqrt{\left(E_o + E_k\right)^2 - E_o^{\,2}}} = \frac{hc}{\sqrt{2E_oE_k + E_k^{\,2}}} = \frac{hc}{\sqrt{2m_ec^2eV_e + e^2V_e^{\,2}}}, \qquad (3.5.2)$$

where λ_d is the de Broglie wavelength of the electron, h is Planck's constant, c is the speed of light in vacuo, p is the relativistic momentum of the electron, E_o is the rest mass energy of the electron, E_k is the classical kinetic energy of the electron, m_e is the mass of the electron, e is the electron charge and V_e is the accelerating electric potential.

Then for electrons at 100kV, the de Broglie wavelength is 3.70pm so that the diffraction limited spot size, as given by Equation 2.2.1, is 1.81nm, which must be doubled to 3.62nm because the electron radiation is incoherent [Broers[49] (1981); Watts[50] (1988)]. The depth of focus as given by Equation 2.2.2 is then 296nm. This large value, due to the very large *f*-number, makes the written pattern more immune to topographic changes in the substrate than would be the case with a conventional scanning electron or transmission microscope. At 50kV, the electron wavelength increases to 5.36pm and with a 400μm aperture, the *f*-number is halved so that the minimum spot size increases to 5.23nm, with the depth of focus decreasing to 107nm. [This treatment ignores the fact that the beamwriter is a gaussian beam system and gaussian optic theory should be applied – this will be done in §5.1.]

In practice the minimum spot size is limited by aberrations in the lenses [Chang, *et al.*[45] (1988)], so that at 100kV the smallest spot size achievable with a 200μm aperture is 12nm. This corresponds to an effective *f*-number of 664 and an effective depth of focus of 3.26μm. This improvement in depth of focus comes at the cost of reduced resolution and is a result of the system not being able to work at the diffraction limit, where the spot size would be expected to diverge from its nominal value at its fastest rate with respect to distance from the focal point. At 50kV and with a 400μm aperture, assuming the minimum realisable spot size is 12nm, the effective *f*-number becomes 459 and the effective depth of focus is 2.25μm. [Although it appears that the depth of focus improves by working at larger spot sizes, it should be noted that the depth of focus is defined in terms of the ratio of increased spot size to the nominal spot size – see §5.1.]

The beamwriter features an interferometric stage capable of being located to an accuracy of 5nm (lambda/120) in each direction [Philips[51] (1980-1993)]. The wafer heights are measured using a solid state laser and the beam focus is automatically adjusted. The spot size is measured using the contrast changes at the edges of holder markers which are located in the same way as substrate markers are for local wafer alignment [Rishton, Beaumont and Wilkinson[52] (1984); Chisholm[53] (1988)]. The stitching accuracy is specified to be 80nm (mean plus two sigma) [Philips[51] (1980-1993)] although wafer processing introduces other errors which mean that the overlay accuracy typically varies from 50nm to 150nm in practice. The beamwriter can align to rectangular heavy metal markers which have positive backscatter electron contrast with respect to the surrounding substrate or to vertically etched rectangular pits which have negative contrast with respect to the substrate and also to square pyramidal silicon etch pits (known as topological markers) which have negative contrast. Positive contrast topological markers cannot be handled as, for reasons discussed in §3.5.1, the edges of such markers do not form a square when projected onto a flat surface.

The patterns to be written are generated using a computer aided drawing package, with most work being carried out in a package called WaveMaker, which allows multiple level patterns (using different layers) to be placed into hierarchical structures, known as cells.

These can themselves contain cells or arrays of cells. The files generated by this package are usually saved in an industrial format, known as the gdsii format. These patterns are fractionated (divided up into blocks) using another software package called CATS (Computer Aided Transcription Software) to produce a file suitable for the beamwriter (the Internal WorkFiLe format or IWFL). The block size is dependent on the pattern resolution used as a block may not contain more than 32000 pixels (although the deflection coils are operated by 15 bit DACs, the maximum number of pixels is not 32768). However, the block size is also limited by aberrations in the lenses to a maximum of 1000μm at 20kV, 800μm at 50kV and 560μm at 100kV. The blocks are divided up into pixels (beam deflection step size) which can range from 5nm to 312.5nm in size. An additional piece of software, BWL (BeamWriter Layout), written in-house by Dr. S. Thoms, processes information on pattern placement, pattern alignment, dose assignment, spot size assignment and various other details, to produce a command file for the beamwriter to execute.

The machine is run by dedicated operators who unload samples every morning and load samples during the day, setting them up as required, so that all the samples can be run during the night unattended, with problematic or short and urgent jobs sometimes being run during the day. This allows the most efficient use of the machine by the many dozens of users. One disadvantage of the beamwriter system is that changing the beam energy from 50kV to 100kV requires several hours of stabilisation, so that one or two days a week are usually dedicated to high resolution operation at 100kV with a 200μm aperture, the remaining days being at 50kV with a 400μm aperture. Much of the work of the beamwriter in this last mode of operation is used to write photolithography masks for various groups within the department as well as external educational institutions and commercial organisations.

3.5.2 Alignment on the BeamWriter

When writing an aligned pattern, the beamwriter distorts its co-ordinate system to match that of the alignment markers on the substrate [Koek[43] (1980-1993)]. This ensures the closest match between one pattern level and the next. The beamwriter expects the alignment markers to be arranged in a rectangular array with the pattern to be written located entirely within the alignment mark perimeter. The beamwriter can align, in turn, to each of an array of cells which each have their own alignment markers, in turn, but is limited to using a fixed step and repeat distance between alignment markers as shown in Figure 3.5.2. This means that alignment markers which are part of different cells (where the cell is defined as the repeating pattern) must be used to align to an individual cell. The beamwriter locally aligns to each cell and can compensate for non-orthogonality of the wafer axes, axis length errors and non-parallelism between the axes on one side of a cell with those on the opposite side. The first of these corrections are compensated using

the x- and y-rotation settings, the second using the x- and y-gain settings while the last is corrected using the x- and y-keystone settings.

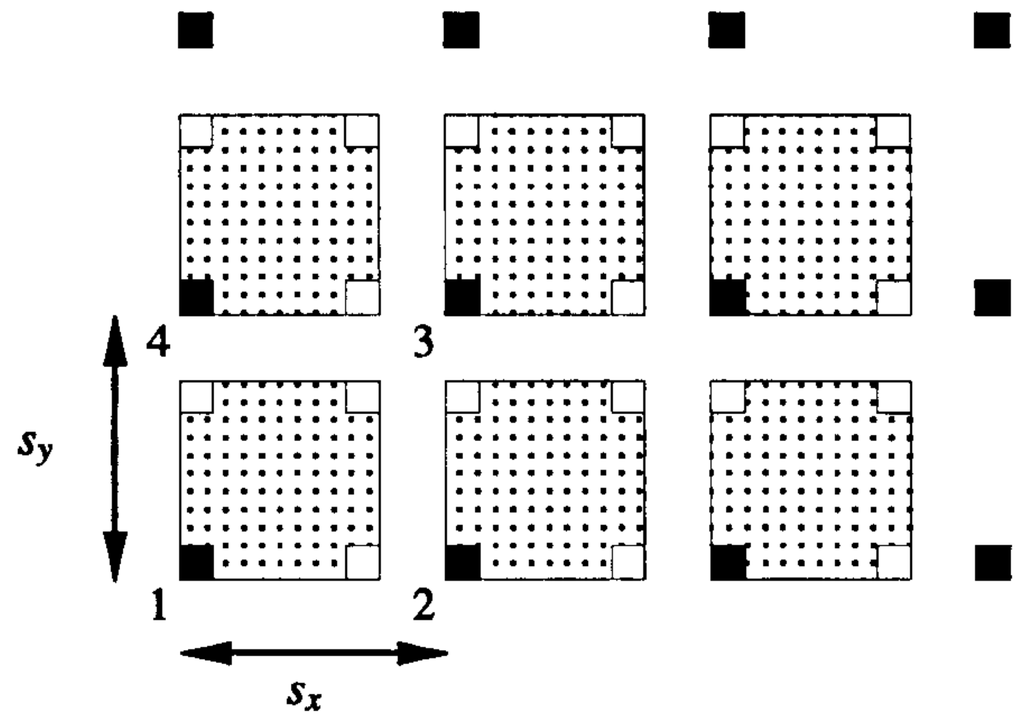


Figure 3.5.2: Illustration of the layout of patterns for exposure on the Beamwriter. The dotted areas are the cells to be stepped out. The alignment markers are shown as squares at the corners of each cell. However, as only one step and repeat distance can be specified for each axis, only the dark markers will be used for alignment. This means an extra row and column of markers must be provided if the patterns of the topmost row and rightmost column are to be placed accurately.

Corrections for alignment markers which have been fabricated on a conically (elliptic, hyperbolic or parabolic) distorted surface can also be carried out. If a large area wafer with a significant amount of wafer bow (*i.e.* the thickness of the wafer at the centre is greater than that at the edge, the surface profile being described by a second order polynomial in the two orthogonal axes of a Euclidean plane imposed on the surface), has markers defined on it using photolithography, the marker spacing at the centre will normally be smaller than that at the edges. In general the beamwriter calculates the actual wafer location, (x_a, y_a), using the following formula:

$$\begin{pmatrix} x_a \\ y_a \end{pmatrix} = \begin{pmatrix} c_x{}^{x}x + c_x{}^{y}y + c_x{}^{xx}x^2 + c_x{}^{xy}xy + c_x{}^{yy}y^2 \\ c_y{}^{x}x + c_y{}^{y}y + c_y{}^{xx}x^2 + c_y{}^{xy}xy + c_y{}^{yy}y^2 \end{pmatrix}, \tag{3.5.3}$$

where (x,y) is the required wafer location and the $c_a{}^{b}$ are the coefficients of $b \in \{x, y, x^2, xy, y^2\}$ along the $a \in \{x, y\}$ axis. Many of these parameters can be found from the positions of four markers used for alignment. The four pairs of equations for the positions of the markers, labelled i, where $i \in \{1,2,3,4\}$, as displayed in Figure 3.5.2, are given by:

$$\begin{pmatrix} x_a{}^{i} \\ y_a{}^{i} \end{pmatrix} = \begin{pmatrix} c_x{}^{x}x_i + c_x{}^{y}y_i + c_x{}^{xx}x_i{}^{2} + c_x{}^{xy}x_iy_i + c_x{}^{yy}y_i{}^{2} \\ c_y{}^{x}x_i + c_y{}^{y}y_i + c_y{}^{xx}x_i{}^{2} + c_y{}^{xy}x_iy_i + c_y{}^{yy}y_i{}^{2} \end{pmatrix}. \tag{3.5.4}$$

From this a set of six pairs of equations (counting only ij combinations with $i < j$ since ii equations are identically zero and ji equations are simply the negations of ij equations) can be obtained for the displacements between the markers:

$$\begin{pmatrix} \Delta x_a{}^{ij} \\ \Delta y_a{}^{ij} \end{pmatrix} = \begin{pmatrix} c_x{}^{x}\Delta x_{ij} + c_x{}^{y}\Delta y_{ij} + c_x{}^{xx}x_{Tij}\Delta x_{ij} + c_x{}^{xy}\Delta xy_{ij} + c_x{}^{yy}y_{Tij}\Delta y_{ij} \\ c_y{}^{x}\Delta x_{ij} + c_y{}^{y}\Delta y_{ij} + c_y{}^{xx}x_{Tij}\Delta x_{ij} + c_y{}^{xy}\Delta xy_{ij} + c_y{}^{yy}y_{Tij}\Delta y_{ij} \end{pmatrix}, \tag{3.5.5}$$

where:

$$\begin{aligned}&\Delta x_a^{\,ij} = x_a^{\,j} - x_a^{\,i},\ \Delta x_{ij} = x_j - x_i,\ x_{Tij} = x_i + x_j\\&\Delta y_a^{\,ij} = y_a^{\,j} - y_a^{\,i},\ \Delta y_{ij} = y_j - y_i,\ y_{Tij} = y_i + y_j,\ \text{and}\ \Delta xy_{ij} = x_j y_j - x_i y_i\end{aligned}. \tag{3.5.6}$$

However two of these equations are degenerate since:

$$\begin{pmatrix}\Delta x_a^{\,13}\\ \Delta y_a^{\,13}\end{pmatrix} = \begin{pmatrix}\Delta x_a^{\,12}\\ \Delta y_a^{\,12}\end{pmatrix} + \begin{pmatrix}\Delta x_a^{\,23}\\ \Delta y_a^{\,23}\end{pmatrix},\ \begin{pmatrix}\Delta x_a^{\,24}\\ \Delta y_a^{\,24}\end{pmatrix} = \begin{pmatrix}\Delta x_a^{\,23}\\ \Delta y_a^{\,23}\end{pmatrix} + \begin{pmatrix}\Delta x_a^{\,34}\\ \Delta y_a^{\,34}\end{pmatrix}, \tag{3.5.7}$$

so that if the following variables are defined:

$$x_1 = x_4 = x_l,\ x_2 = x_3 = x_r,\quad y_1 = y_2 = y_d,\ y_3 = y_4 = y_u,\quad s_x = x_r - x_l,\ s_y = y_u - y_d, \tag{3.5.8}$$

where s_x and s_y are the step and repeat distances in x and y, respectively, then Equation 3.5.5 becomes:

$$\begin{aligned}\begin{pmatrix}\Delta x_a^{\,12}\\ \Delta y_a^{\,12}\end{pmatrix} &= +\begin{pmatrix}c_x^{\,x}x_s + c_x^{\,xx}(x_l + x_r)x_s + c_x^{\,xy}x_s y_d\\ c_y^{\,x}x_s + c_y^{\,xx}(x_l + x_r)x_s + c_y^{\,xy}x_s y_d\end{pmatrix}\\ \begin{pmatrix}\Delta x_a^{\,34}\\ \Delta y_a^{\,34}\end{pmatrix} &= -\begin{pmatrix}c_x^{\,x}x_s + c_x^{\,xx}(x_l + x_r)x_s + c_x^{\,xy}x_s y_u\\ c_y^{\,x}x_s + c_y^{\,xx}(x_l + x_r)x_s + c_y^{\,xy}x_s y_u\end{pmatrix}\\ \begin{pmatrix}\Delta x_a^{\,14}\\ \Delta y_a^{\,14}\end{pmatrix} &= +\begin{pmatrix}c_x^{\,y}y_s + c_x^{\,xy}y_s x_l + c_x^{\,yy}(y_d + y_u)y_s\\ c_y^{\,y}y_s + c_y^{\,xy}y_s x_l + c_y^{\,yy}(y_d + y_u)y_s\end{pmatrix}\\ \begin{pmatrix}\Delta x_a^{\,23}\\ \Delta y_a^{\,23}\end{pmatrix} &= +\begin{pmatrix}c_x^{\,y}y_s + c_x^{\,xy}y_s x_r + c_x^{\,yy}(y_d + y_u)y_s\\ c_y^{\,y}y_s + c_y^{\,xy}y_s x_r + c_y^{\,yy}(y_d + y_u)y_s\end{pmatrix}\end{aligned}. \tag{3.5.9}$$

By combining the first two and last two of these, the following coefficients can be found:

$$\begin{aligned}\begin{pmatrix}\Delta x_a^{\,12}\\ \Delta y_a^{\,12}\end{pmatrix} + \begin{pmatrix}\Delta x_a^{\,34}\\ \Delta y_a^{\,34}\end{pmatrix} &= +s_x\begin{pmatrix}c_x^{\,xy}y_d\\ c_y^{\,xy}y_d\end{pmatrix} - s_x\begin{pmatrix}c_x^{\,xy}y_u\\ c_y^{\,xy}y_u\end{pmatrix} = -s_x s_y\begin{pmatrix}c_x^{\,xy}\\ c_y^{\,xy}\end{pmatrix}\\ \begin{pmatrix}\Delta x_a^{\,14}\\ \Delta y_a^{\,14}\end{pmatrix} - \begin{pmatrix}\Delta x_a^{\,23}\\ \Delta y_a^{\,23}\end{pmatrix} &= +s_y\begin{pmatrix}c_x^{\,xy}x_l\\ c_y^{\,xy}x_l\end{pmatrix} - s_y\begin{pmatrix}c_x^{\,xy}x_r\\ c_y^{\,xy}x_r\end{pmatrix} = -s_x s_y\begin{pmatrix}c_x^{\,xy}\\ c_y^{\,xy}\end{pmatrix}\end{aligned}. \tag{3.5.10}$$

However, after substituting these coefficients into Equation 3.5.9, no further coefficients can be found as the equations are degenerate. This can be seen to be the case since the coefficients are those of a parabola which can be reduced to a quadratic in each direction. To solve these quadratics it is necessary to have three points, or two displacements in each direction. This can be done by, for example, assuming that the second order terms are constant over the substrate, so that the markers at the extrema of the substrate are also used. However, as these second order corrections are small, they are not normally used by the beamwriter software so that the other coefficients can be found from:

$$\begin{aligned}\begin{pmatrix}\Delta x_a^{\,12} - s_x c_x^{\,xy}y_d\\ \Delta y_a^{\,12} - s_x c_y^{\,xy}y_d\end{pmatrix} &= -\begin{pmatrix}\Delta x_a^{\,34} + s_x c_x^{\,xy}y_u\\ \Delta y_a^{\,34} + s_x c_y^{\,xy}y_u\end{pmatrix} = +s_x\begin{pmatrix}c_x^{\,x}\\ c_y^{\,x}\end{pmatrix}\\ \begin{pmatrix}\Delta x_a^{\,14} - s_y c_x^{\,xy}x_l\\ \Delta y_a^{\,14} - s_y c_y^{\,xy}x_l\end{pmatrix} &= \begin{pmatrix}\Delta x_a^{\,23} - s_y c_x^{\,xy}x_r\\ \Delta y_a^{\,23} - s_y c_y^{\,xy}x_r\end{pmatrix} = +s_y\begin{pmatrix}c_x^{\,y}\\ c_y^{\,y}\end{pmatrix}\end{aligned}. \tag{3.5.11}$$

These coefficients are given names: c_x^x and c_y^y are the x- and y-gains respectively; c_x^y and c_y^x are the x- and y-axis rotations respectively, and; c_x^{xy} and c_y^{xy} are the x- and y-keystone coefficients respectively.

Even determining these coefficients requires all four markers of each cell to be found. This is a strong requirement, as it is fairly likely that one marker from each cell may not be in good condition after many patterning levels. If the keystone corrections are assumed to be small in comparison to other errors, only three markers need be found in each cell. This has been found to reduce alignment time without compromising alignment accuracy significantly [Thoms[54] (1993-1996)].

If the beamwriter does not find three markers, it uses information from another cell nearby instead. The algorithm used to decide which cell is used is not known precisely and, as would be expected, this is not a desirable situation. [Using the keystone corrections causes this situation to arise more often as four markers are required. Ideally the algorithm should use keystone corrections when it can find four markers and ignore them when only three are found. However, this software is commercially produced and substituting in-house algorithms requires a significant amount of development work.]

In order to obtain global wafer distortion corrections, the beamwriter looks for several markers before calculating a global wafer rotation angle and global stage translation settings (this is the displacement of the wafer origin due to the beamwriter compensating for the wafer rotation angle). The beamwriter can typically only cope with wafer rotation angles of 0.5° or less (0.2° or less recommended). The bottom left marker location is used as the origin for the global stage translation settings [Thoms[54] (1993-1996)].

For each cell on the wafer, the beamwriter then tries to locate as many markers as required. In general the beamwriter can detect markers with a size ranging from 0.1μm to 100μm, with the range from 3μm to 80μm yielding good results. A marker size of 20μm is normally recommended and is a necessity for the detection of topological markers (the reason for this is not known – it may only be a limitation of the software). When looking for a marker the beamwriter gradually spirals outwards from a minimum search area to a maximum area. If it does not find the marker, the software modifications automatically adjust various settings, such as contrast, and focus and the search is repeated several times with different parameters. If the marker is still not found the beamwriter looks for another marker in the cell. If it finds the required number of markers then alignment and exposure of the pattern takes place. If it cannot find a sufficient number of markers, then the situation mentioned above arises and values from a different cell are used to extrapolate to the pattern centre position for the current cell.

3.5.3 Practical Electron-Beam Lithography

In practice, there are many design rules or rules of thumb that are used in electron-beam lithography. Many of these are associated with the proximity problems inherent in

electron-beam lithography. These problems are a result of the background exposure due to backscatter electrons. To form a 100nm width line which is part of a densely patterned level may require a dose up to five times lower than the same size feature in a less dense pattern. In general, two proximity effects are normally identified: the intraproximity effect and the interproximity effect [Thoms[55] (1993-1996)].

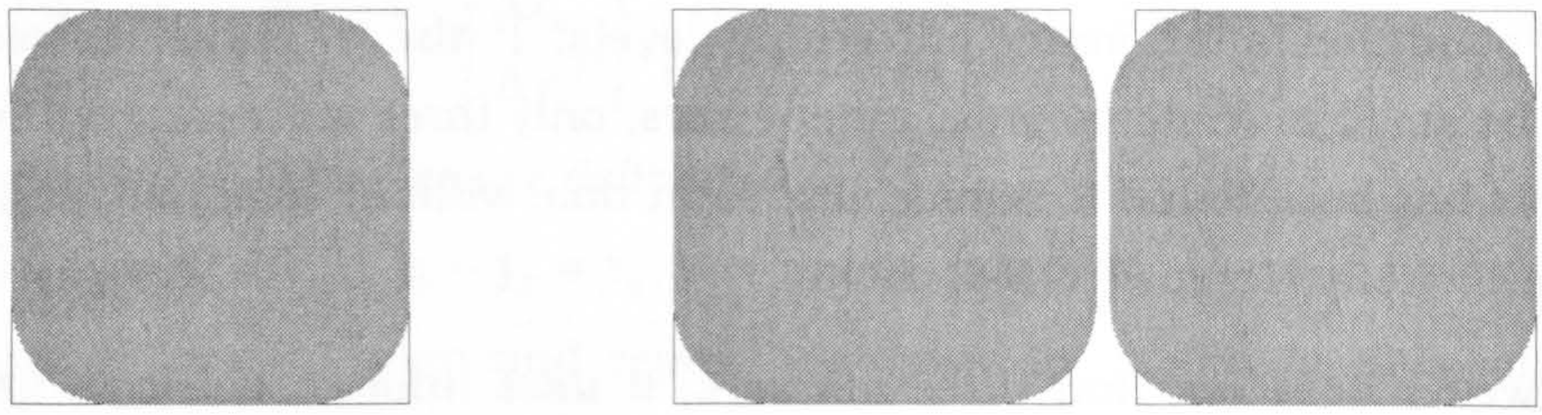


Figure 3.5.3: Illustration of the intra and inter proximity effect in electron-beam lithography. The square area is the exposed pattern, the grey areas are the resulting developed patterns.

The intraproximity effect particularly affects exposed concave vertices. The background exposure at the concave vertex of an exposed shape is much lower than that between vertices so that for a uniform dose the concave vertex is rounded off as shown in the left half of Figure 3.5.3. The interproximity effect is associated with the increased backscatter exposure of the resist between two exposed shapes separated by a narrow region. This causes the separation between the two shapes to be reduced as shown in the right half of Figure 3.5.3. In order to deal with this, the pattern is often altered to compensate for these effects, usually by another piece of software known as a proximity correction package. However, due to the extensive electron-beam lithography experience within the department, proximity correction is usually carried out by better assignment of exposure doses, based on exposure tests, together with pattern designs which take account of feature growth.

Some more general design rules are:

(a) For features larger than 100nm the pixel resolution of the pattern should be 10% of the smallest feature to be defined, unless edge rounding is acceptable [Thoms[55] (1993-1996)].

(b) The spot size should always be larger than the pixel resolution, usually 100% or more larger (unless a dot grating is to be generated where the resolution can be used to determine the pitch of the grating [Thoms[56] (1993-1996)]).

(c) For lift-off patterns the spot size at feature edges should be smaller than the resist thickness [Nicol[57] (1993-1996)].

(d) Exposure doses at 100kV should be approximately double that at 50kV, unless patterning on a surface with reduced backscatter (*e.g.* a thin membrane) [Thoms[55] (1993-1996)].

(e) For features of ten times the spot size or smaller exposed areas should be one to two spot diameters smaller than their physical size to keep dimensional errors below 10%.

(f) For features smaller than 50nm, the pattern should be one to two pixels wide and the dose should be used to achieve the dimension required.

(g) For features of 40nm or smaller, 100kV electrons produce better results.

(h) For lift-off patterns, rounded vertices lift-off easier than sharp vertices.

(i) The thickness of resist should not exceed five times the feature width.

(j) The alignment marker pattern should be written in as short a time as possible, and certainly in less time than the period between beamwriter calibrations [Thoms[54] (1993-1996)].

Design rule (b) is based on the separation between gaussian beams required to minimise the dose ripple effect. This is illustrated in Figure 3.5.4. The result is analogous to the Helmholtz condition for a uniform magnetic field on the central axis between two electromagnet coils [Duffin[58] (1990)]. In general, these rules are rules of thumb and may vary somewhat with substrate, resist, and pattern.

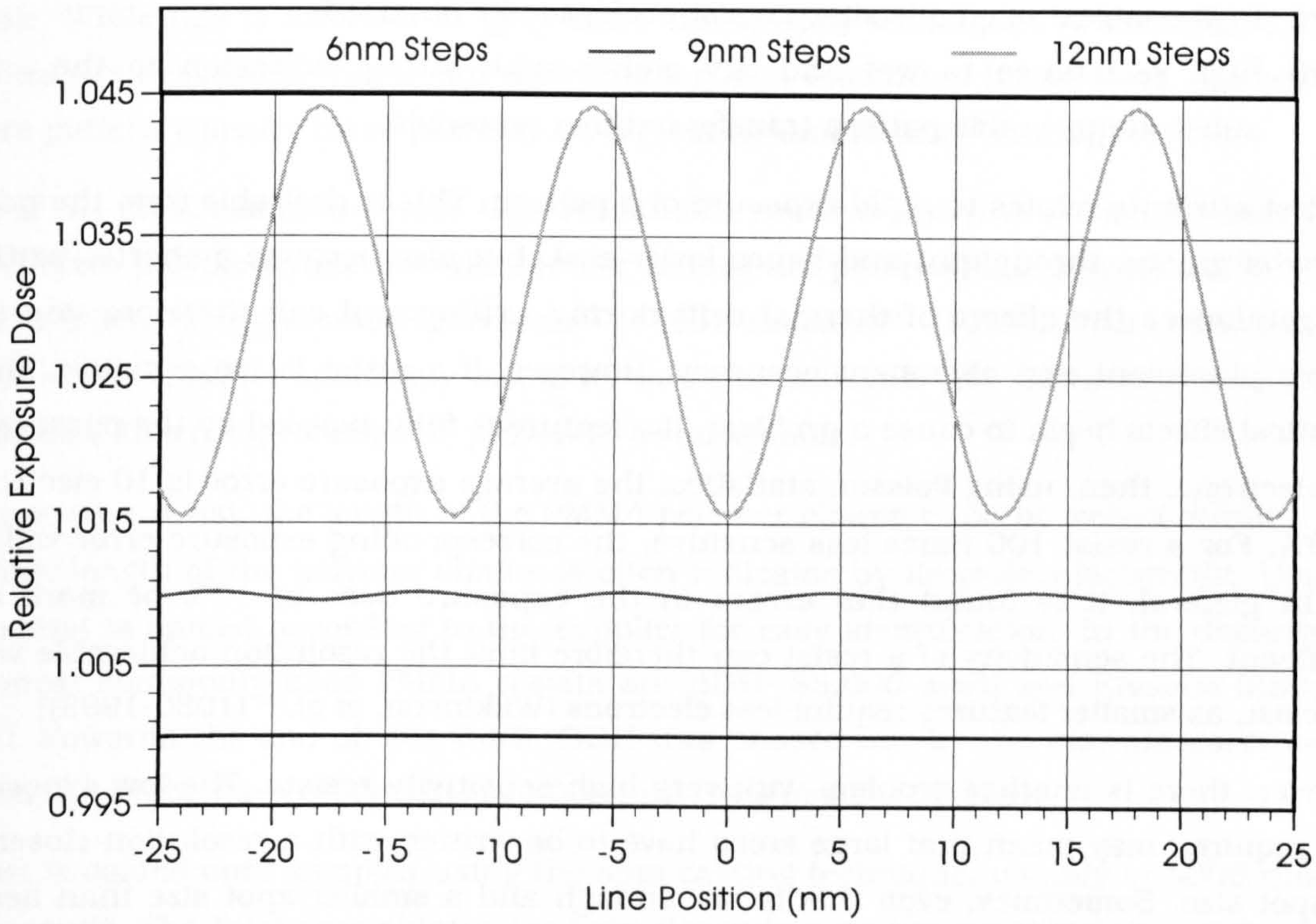


Figure 3.5.4: Illustration of the dose ripple effect for a 12nm spot diameter electron beam exposing a single pixel line in steps of 6nm, 9nm and 12nm. The doses have been normalised with respect to their average value. The '9nm Steps' line has been shifted up by 0.01 and the '12nm Steps' line has been shifted up by 0.03 for clarity.

To implement design rules (c) and (j), and where large features must be placed very accurately with respect to each other (for example where the gap between two one micron structures must be controlled to 50nm or less), a technique known as 'sleeving' can be used [Thoms[54] (1993-1996)]. This allows a pattern to be split into two patterns which

overlap by an amount greater than any expected misalignment, usually one to two microns, one of which defines the edges of the feature and is written with a small spot size, while the other is used to fill out the rest of the feature quickly by using a large spot size. This is particularly useful for defining alignment markers, as a smaller spot size allows more accurate marker edge definition (the edge is the most significant part of the marker for the search and measurement algorithms used by the beamwriter), and because of the reduced writing area when using sleeving, all the marker edges can be written in the time between beamwriter calibrations (which is one of the more likely sources of placement error, particularly when the marker level is being written without alignment to a previous level). A larger spot size can then be used to fill out the body of the marker. A quicker alignment marker level minimises placement errors caused by the thermal drift of the substrate holder during writing [Thoms[59] (1993-1996)].

Many resists can be used for electron-beam lithography. In general the desirable qualities of an electron-beam resist are [Wilkinson and Beaumont[60] (1980-1993); Pease[61] (1981)]:

(a) high sensitivity to reduce exposure time;
(b) high resolution to enable smaller features at higher densities;
(c) high contrast to enable sharp resist profiles to be obtained;
(d) high resistance to wet and dry etches with strong adhesion to the substrate to enable pattern transfer to other materials.

The first attribute relates to rapid exposure of a pattern. This is desirable from the point of view of greater throughput and hence lower cost, but also because a shorter writing time minimises the effects of thermal drift during writing and can therefore improve pattern placement and alignment accuracy. However, if a resist is too sensitive, then statistical effects begin to cause a problem. If a feature is fully exposed by the passage of 100 electrons, then, using Poisson statistics, the average exposure error is 10 electrons or 10%. For a resist 100 times less sensitive, the corresponding exposure error will be 1%. In general, it is found that errors in the exposure dose of 10% or more are significant. The sensitivity of a resist can therefore limit the resolution achievable with the resist, as smaller features require less electrons [Wilkinson, *et al.*[60] (1980-1993)].

However, there is another problem with very high sensitivity resists. The low exposure dose required may mean that large areas have to be written with a resolution closer to the spot size. Sometimes, even this is not enough and a smaller spot size than usual must be used. This is because the beamwriter can only deflect the beam at a maximum rate of 10MHz. Thus if the dose is so low that a higher frequency than this is required, the only option is to reduce the spot size to reduce the current, which has the effect of increasing the exposure time and lowering the deflection frequency required. The required frequency depends on the resolution, dose and beam current according to:

$$\text{beam frequency} = \frac{\text{beam current}}{\text{dose} \times \text{resolution}^2}. \qquad (3.5.12)$$

Thus a dose of $10\mu C/cm^2$ with a resolution of 80nm and a 160nm spot at 50kV, corresponding to a beam current of about 27nA, means a beam deflection frequency of about 42.19MHz. With a resolution of 160nm, the frequency has dropped to around 10.55MHz. The total writing time (excluding alignment, stage movement and calibration times) can then be calculated from:

$$\text{writing time} = \frac{\text{writing area}}{\text{beam frequency} \times \text{resolution}^2} = \frac{\text{dose} \times \text{writing area}}{\text{beam current}}. \quad (3.5.13)$$

Attribute (c) also requires further explanation. The concept of resist contrast is directly analogous to that of film contrast in photography, where a film is exposed and then developed to produce the picture. A high contrast film will go from black to white with very low light exposure. A low contrast film will produce an intermediate level of grey with the same exposure. Similarly, a high contrast resist will go from being unexposed to fully exposed over a very small exposure dose range. This is desirable, since the intensity profile at the edge of a gaussian electron beam is not sharp, but sloped. In order that such a beam can produce a sharp resist profile, the contrast must be high. A low contrast resist will produce an edge profile approaching the electron-beam intensity profile. While this is an asset in 'grey-scale' lithography techniques (where sloped resist profiles are desirable), this is not desirable for high resolution lithography, particularly where pattern transfer takes place by lift-off [Maker and Muller[62] (1992)].

One of the most commonly used high resolution resists is also one of the first ones to be discovered: polymethylmethacrylate (other names are perspex and acrylic) or PMMA for short [Haller, Hatzakis and Srinivasan[63] (1968); Hatzakis[64] (1969)]. This resist satisfies attributes (b) and (c) admirably but is let down by attribute (d). However, for very high resolution, this compromise has proven to be acceptable.

For use as a resist, the length of the PMMA polymer chains must be well-controlled. The average length of the polymer chains is often indicated by its molecular weight. Usually the resist is named according to the supplier for easy identification. In the department the most commonly used PMMA resists are BDH (85,000 mwt) and Elvacite (350,000 mwt). Towards the end of this work, BDH was phased out by the manufacturer and a substitute in the form of Aldrich (120,000 mwt) was used in its place.

Resist is coated onto samples using the spin casting technique, usually at 5000 rpm for 60 seconds. The thickness obtained is then dependent on the concentration of the resist in the solvent [Daughton and Givens[65] (1982); Kuan and Frank[66] (1988)]. Two solvents are commonly used for PMMA: orthoxylene and chlorobenzene. The latter is more hazardous, but a better solvent, and is normally used for resists of higher concentration as orthoxylene can only cope with resist concentrations up to about 12% by weight. Some typical resist thicknesses are given in Table 3.5.1 [Thoms[55] (1993-1996)].

Resist	Spin Speed	Solvent	Thickness	Metallisation
2.5% BDH	3000 rpm	orthoxylene	42 nm	30 nm
2.5% BDH	5000 rpm	orthoxylene	24 nm	20 nm
2.5% Elvacite	5000 rpm	orthoxylene	40 nm	not applicable
4% BDH	5000 rpm	orthoxylene	69 nm	50 nm
4% BDH	8000 rpm	orthoxylene	44 nm	35 nm
4% Elvacite	5000 rpm	orthoxylene	100 nm	not applicable
8% BDH	5000 rpm	orthoxylene	208 nm	150 nm
10% BDH	3000 rpm	orthoxylene	580 nm	450 nm
10% BDH	5000 rpm	orthoxylene	350 nm	250 nm
12% BDH	5000 rpm	chlorobenzene	700 nm	500 nm
15% BDH	5000 rpm	chlorobenzene	1420 nm	1000 nm
18% BDH	5000 rpm	chlorobenzene	1700 nm	1250 nm

Table 3.5.1: The typical thickness of PMMA resist layers obtained by spinning for 60 seconds and baking at 180°C for two hours or more. The last column represents the maximum metallisation thickness that can be used for the corresponding bottom layer resist. Data collated by Dr. S. Thoms [Thoms[55] (1993-1996)].

The dose required to expose the resist is higher for heavier molecular weight polymer chains so that an undercut profile suitable for lift-off can be obtained by using a bi-layer resist system consisting of a bottom layer of BDH or Aldrich and a top layer of Elvacite [Beaumont, Tamamura and Wilkinson[67] (1981)]. These are normally spun onto the sample one after the other with a 180°C bake following each layer. The first bake is typically one to two hours while the second bake is usually four hours or more. These bakes dry out the solvent and compact the PMMA layer for good results at the highest resolutions. Since the glass transition temperature for PMMA is about 120°C [Hatzakis[64] (1969)], the PMMA flows slightly during the bake, resulting in a flatter resist layer.

The typical 50kV exposure dose for bi-layers of BDH and Elvacite are around 300μC/cm^2 for features larger than 1μm, 400μC/cm^2 for features of around 1μm, 600μC/cm^2 for features of around 0.1μm and anywhere from 1000μC/cm^2 to 2000μC/cm^2 for isolated features smaller than this, depending on feature size and density [Thoms[55] (1993-1996)]. Doses at 100kV are approximately double this value.

The solution used to develop PMMA patterns is a combination of a fast PMMA dissolving component, methylisobutylkeytone (MIBK), and a slow dissolving component, isopropanol (isopropyl alcohol or IPA) [Greeneich[68] (1975)]. The fastest developer used is normally a 1:1 IPA:MIBK combination, typically used to develop thicker PMMA layers, where residue

at the feature bottom is more of a problem. The residue is due to a small proportion of the resist chains becoming cross-linked upon exposure, resulting in less soluble chains [Wilkinson, *et al.*[60] (1980-1993)]. The thickness of the residue layer can therefore be expected to be proportional to the original resist thickness.

In general, the faster the development, the lower the selectivity between exposed and unexposed resist areas, which results in a sharper feature and a cleaner exposed surface. However, this has the unfortunate side-effect of also reducing the level of overhang for lift-off applications. For thicker layers, forward scattering will produce a natural undercut profile, which when combined with the larger discontinuity between the metal deposited upon the resist surface and that on the substrate surface, gives good lift-off results, even if the higher molecular weight second layer is not used [Hatzakis[69] (1988)].

For higher resolution features where the overhang is more critical, slower developers, 2:1 IPA:MIBK, 2.5:1 IPA:MIBK and 3:1 IPA:MBK, are normally used, with the last of these offering the highest resolution and slowest development time. However, there is a trade-off here as the longer the development time, the more time is available for the resist to absorb solvent and swell up. This swelling can reduce the gaps between fine features at best, and cause features to merge together at worst [Chang, *et al.*[45] (1988)]. To reduce this effect, the development times used with the slower developers are reduced in comparison with the faster developers by using increased exposure doses. Typical development times used are 60 seconds for 1:1 IPA:MIBK, 45 seconds for 2.5:1 IPA:MIBK and 2:1 IPA:MIBK and 30 seconds for 3:1 IPA:MIBK, all at a fixed temperature of 23±1°C [Thoms[54] (1993-1996)].

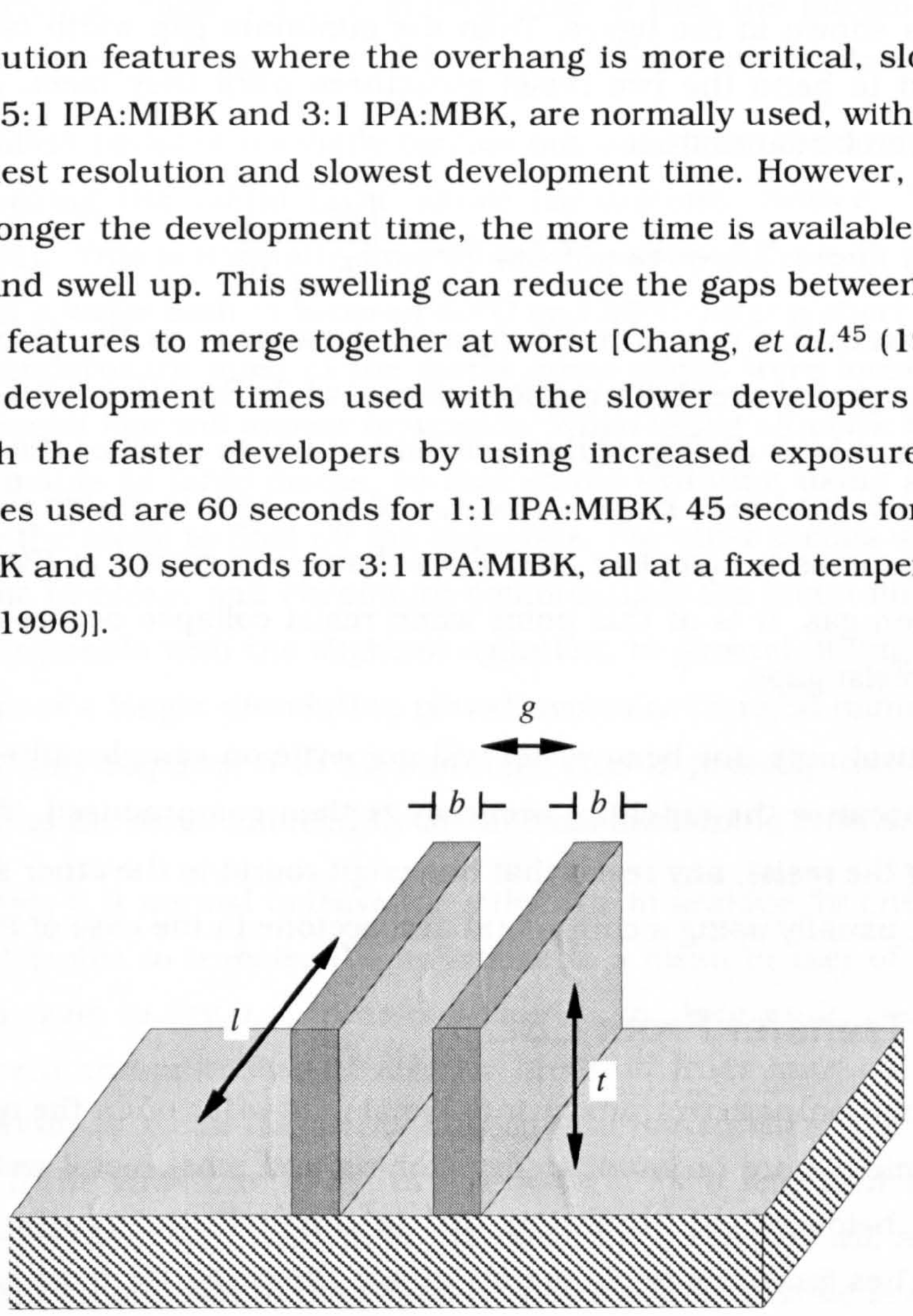


Figure 3.5.5: Two high aspect ratio resist lines separated by a short gap. Capillary forces can play a significant role during the development of such features.

Features at high aspect ratios can result in resist collapse due to a combination of the lack of rigidity of the resist and the action of capillary forces within the resist gaps upon removal of the substrate from the developer solution [McCord, Wagner and Donohue[70] (1993)]. This affects thicker resist layers more as can be argued qualitatively from

Figure 3.5.5. The lateral stiffness, k_{resist}, of the resist structure is given by (as calculated using Equation K.51):

$$k_{resist} = \frac{E_{resist} l b^3}{4t^3}, \tag{3.5.14}$$

where E_{resist} is Young's modulus for the resist and the other parameters are the resist dimensions as shown in Figure 3.5.5. The capillary forces between these two structures are proportional to the surface area and inversely proportional to the separation between the structures [Bolton[71] (1986)] so that the force, F_{resist}, between them may be given by:

$$F_{resist} = \frac{\gamma_{developer} tl}{g}, \tag{3.5.15}$$

where $\gamma_{developer}$ is the surface tension of the developer solution and the other parameters are dimensions as shown in the figure. Then the minimum gap width is determined by the force required to bend the two resist structures until they meet. As the force is mutual, this gap can be found from:

$$F_{resist} = k_{resist} g_{min} \Rightarrow g_{min} = 2t^2 \sqrt{\frac{\gamma_{developer}}{E_{resist} b^3}}. \tag{3.5.16}$$

It should be stressed that this is not a rigorous treatment as the force acts between opposite points integrated over both resist surfaces, while Equation 3.5.14 assumes that the total force acts only at the top of the resist structure. In practice the variation of the gap width with resist thickness is more linear, and gives rise to design rule (i). Typically, the development process is complete after the substrate is rinsed in IPA and blown dry with air or nitrogen gas. It is at this point when resist collapse occurs as the liquid is driven out of the resist gaps.

As a further practical note, the beamwriter will not write on samples tilted by more than about 4μm/mm because the stitching accuracy is then compromised. This means that prior to prebaking the resist, any resist that has crept round to the other side of the wafer must be removed, usually using a cotton bud and acetone in the case of PMMA.

3.5.4 Pattern Transfer From EBL

For very high resolution pattern transfer into metals, lift-off is often the only choice. This is because most metals are resistant to dry etching and most metal wet etches cannot produce features below 50nm accurately due to the etch undercutting the resist. In particular wet etches for many noble metals attack the resist far more quickly than the metal. PMMA is resistant to water, IPA, methanol, hydrochloric acid for short periods, dilute ammonium hydroxide, dilute sulphuric acid, dilute hydrofluoric acid for short periods, phosphoric acid for short periods and dilute nitric acid for short periods, amongst others. It is attacked by acetone (which is used to dissolve the resist rapidly during lift-off), chlorobenzene, concentrated ammonium hydroxide, hydrofluoric acid after a short time depending on concentration [Thoms[55] (1993-1996)].

For lift-off to succeed, the metal layer must adhere well to the substrate beneath, so that the presence of any residue can prove disastrous. Rather than attempting to use a stronger developer, which would compromise the resist undercut profile, an alternative in the form of an oxygen 'ash' can be used. Here an oxygen plasma (rich in oxygen radicals), non-directed by not placing the sample on an electrode, slowly 'burns' away a thin layer of the hydrocarbon rich resist. By controlling the ashing time, the edges of exposed features can be partially cleaned out without significantly affecting the resist overhang. However, for the highest resolution patterns, accurate control of the electron-beam exposure dose, by using an exposure test, where the feature is repeated across the substrate but with many different exposure doses, is still preferred. The thickness of metal layer that can be lifted off is limited by the thickness of the bottom layer of resist, with limits as given in Table 3.5.1. A general rule is that the maximum thickness of metallisation is between 70% and 80% of the bottom resist layer thickness.

Lift-off with a PMMA resist is normally carried out using acetone to dissolve the resist rapidly, thus freeing the metal layer above [Beaumont, Bower, Tamamura and Wilkinson[72] (1981)]. This is normally done by placing the substrate in a beaker of warm acetone heated in a water bath to between 40°C and 50°C. After a short time the acetone will have crept underneath most of the metal areas which were located on the resist surface and the metal film will appear to wrinkle. When lifting off noble metals, the metal film normally remains as large pieces, so that gentle agitation using a pipette is often required to force the metal to float off the substrate. For other metals with small grains, such as nichrome (a nickel and chromium combination) the metal film cracks up and splits into tiny fragments with the slightest agitation. In general, lifting-off metals of the former type requires a longer dissolution period, typically 15 to 30 minutes, compared to the latter which only requires 5 to 10 minutes. This is possibly due to the increased access of acetone to the resist afforded by metal films containing cracks.

For stubborn areas, it is normal to leave the substrate in acetone for one to two hours. In this case it is advisable to transfer the substrate to a clean beaker of acetone (without allowing the substrate to dry, *i.e.* transfer should take place under a continuous jet of acetone) to prevent the resettling of already lifted-off fragments onto the substrate. Resettling can result in metal fragments adhering to the substrate. The worst possible situation arises if the substrate is left in the acetone for a length of time sufficient to allow the acetone to completely boil off. In this case, the metal film will adhere strongly to the substrate due to the action of the capillary forces as the acetone is evaporated away. It is normally impossible to recover from this situation.

For extremely stubborn areas, as often occurs when the metal thickness approaches that of the resist, a syringe may be used to apply a fierce jet of acetone against the substrate [Beaumont, *et al.*[72] (1981)]. This may be accompanied by short ten second bursts in an ultrasonic bath. More drastic methods involve the gentle physical rubbing of the substrate with a cotton bud under the acetone. However, excessive application of these

techniques can severely test the adhesion of the metal with the substrate and may result in removal of the metal from areas other than those originally intended.

To make sure that lift-off has occurred, the substrate can be transferred to a low magnification microscope, whilst remaining under acetone. Various features can then be inspected. This is limited to features which the microscope can resolve, or very fine but periodic structures spread over an area large enough for the microscope to resolve (this results in an observable optical diffraction pattern). Experience shows that lift-off is most likely to fail at step edges and at convex corners, so that the pattern should ideally be on a flat surface and contain curved edges where possible (*i.e.* circles lift-off easier than squares).

Despite this, using a bi-layer of 2.5% BDH, 2.5% Elvacite, 20nm lines are relatively easily defined at 100kV by lift-off. Densely patterned features, such as gratings, of around 50nm pitch (with a 1:1 mark:space ratio) can be obtained at 100kV, with 70nm achievable at 50kV. Higher resolutions can be obtained using thin substrates which reduce the backscatter component of the exposure [Broers, Harper and Molzen[73] (1978); Beaumont, *et al.*[72] (1981)]. This is because 50kV electrons can penetrate the substrate to a depth of 5μm to 10μm before being absorbed or backscattered, so that patterning onto a membrane less than one micron thick can significantly reduce the backscatter component of the exposure [Parikh and Kyser[74] (1979)].

It has been noted that even when spot sizes of less than 3nm are used (*e.g.* in a converted scanning transmission electron microscope with low *f*-number, high numerical aperture optics), the minimum resist feature is still between 5 and 10nm. This is either attributed to the generation of secondary electrons by the incident electron beam [Broers[75] (1981)], the minimum length of the fractured PMMA chains [Chang, *et al.*[45] (1988)], the strength of intermolecular forces between the resist features during development [Chen and Ahmed[76] (1993); Chen and Ahmed[77] (1993)], or the capillary forces generated upon removal of the resist pattern from the developing solution [McCord, *et al.*[70] (1993)]. However, using more novel techniques, a wire with a width of only 3nm has been fabricated using PMMA by lift-off [Cumming, Thoms, Beaumont and Weaver[78] (1996)]. From this it can be concluded that the resolution currently achieved in PMMA on the beamwriter is not limited by the resist, but by the spot size.

For high resolution pattern transfer to other materials, such as silicon nitride, dry etching is often the only possibility. Unfortunately, PMMA offers very low dry etch resistance to the most common etching gases under normal conditions, so that more complicated multi-layer schemes, such as patterning a metal layer using lift-off and subsequently using this layer as a dry etch mask, are often necessary [Howard[79] (1981)]. In some cases, optimisation of the etching parameters allows PMMA to survive sufficiently long for pattern transfer to take place, however, erosion of the resist edges during the etch often means that the dry etched features have sloped instead of vertical

profiles [Murad[80] (1993-1996)]. It has been found experimentally that higher molecular weight PMMA offers better dry etch resistance than lower molecular weight PMMA. For this reason, 900,000 mwt PMMA has been used where greater dry etching resistance is required at the expense of lower resolution [Cameron[81] (1993-1996)].

When exposing PMMA on insulating substrates, such as glass or quartz, or exposing thick resist layers, a charge conduction layer is often necessary [Liu, Ingino and Pease[82] (1995)]. This can be done by evaporating a 30nm to 50nm thick chromium or nichrome (60%/40% nickel/chromium alloy) layer onto the resist surface prior to exposure. These metals are easily removed by wet etching prior to development without attacking the PMMA resist.

For the manufacture of 'T-gates', a resist with a sensitivity in between that of BDH (or Aldrich) and Elvacite is necessary [Cameron[81] (1993-1996)]. The resist most commonly used in the department for this purpose is copolymer which is a mixture of methylmethacrylate and methylacrylic acid [Haller, Feder, Hatzakis and Spiller[83] (1979); Hatzakis[84] (1979)].

To avoid some of the pattern transfer problems associated with PMMA, other lower resolution but more dry etch resistant resists can be used such as SAL601 and Polyimide [Patrick, Mackie, Beaumont and Wilkinson[85] (1985)]. SAL601 also has the advantage of higher sensitivity with a dose of only 11$\mu C/cm^2$ to 14$\mu C/cm^2$ required for large areas at 50kV, using Microposit MF322 developer. Unfortunately it requires a long development time, 10 to 20 minutes, but features down to 100nm can be achieved [MacIntyre and Thoms[86] (1995-1996)].

Some photoresists, such as AZ 1400-17, AZ 1400-31, AZ 1500-18 and S1818 amongst others, can also be used as electron-beam resists. These offer better wet and dry etch resistance but their resolution is lower, typically greater than 100nm, with a sensitivity comparable to PMMA, if not slightly higher [Thoms[54] (1993-1996)].

In general, negative resists cannot be used for lift-off as they usually have a rounded edge profile – exactly the opposite of that required [Chiong, Wind and Seeger[87] (1990)]. However, negative resists usually offer much better wet and dry etching resistance than positive resists. Unfortunately, this extra resistance is accompanied by a more difficult removal process. Negative electron-beam resists such as High Resolution Negative (HRN) [Thoms, *et al.*[88] (1988)] and AZ PN114 [Early, *et al.*[89] (1992)] are particularly useful when the area without resist exceeds the area with resist in the final pattern.

Resolutions down to 20nm can be achieved with 4% HRN (60nm thick when spun at 5000 rpm for 60 seconds and prebaked for 1 to 2 hours at 120°C). HRN is a styrene based resist which offers very good dry etch resistance. It is developed by four alternate 10 second dips in MIBK and IPA. 8% HRN (120nm thick) can be used for lower resolution but longer dry etching. Removal of the resist is accomplished by using hot, concentrated

sulphuric acid or fuming nitric acid, or alternatively by etching in an oxygen plasma, which is normally the preferred option. The sensitivity of this resist is the same as PMMA but the contrast is lower so gratings are limited to about 150nm in pitch [Thoms[55] (1993-1996)]. Unfortunately the resist has been discontinued so that the supply within the department has passed its shelf life.

AZ PN114 is an alternative negative resist which is currently undergoing trials in the department. It is a chemically amplified resist so that it offers very high sensitivity, with doses of 10μC/cm^2 to 14μC/cm^2 being typical. 30nm lines have recently been achieved with this resist [MacIntyre and Thoms[90] (1996)].

Most mask plates are bought in pre-coated with chromium and electron-beam resist to a high degree of uniformity. The resist is normally EBR9 and offers a high sensitivity of about 30μC/cm^2. This allows such plates to be written quickly, with the chromium being patterned using wet etching. The use of wet etching is not restrictive as the minimum feature size usable in industrial photolithography is still about 250nm, and within the department, contact photolithography is rarely used to produce features smaller than 1000nm.

Since high resolution was an important factor in this work, most electron-beam lithography was carried out with PMMA, although other resists, in particular polyimide and photoresist, were considered from time to time as a way of overcoming some of the hurdles that arose. Some initial work was also carried out using HRN.

3.6 OTHER APPARATUS

This section describes some of the other important equipment that was used in the course of development. Being one of the major centres for optoelectronics and nanolithography in the United Kingdom, the departmental facilities are fairly extensive.

3.6.1 Photolithography: Contact Mask Aligner

Two of these machines are available, and allow exposure of an area of up to 3.5 inches in diameter with reasonable uniformity. Both are based on mercury arc lamps and make extensive use of optical microscopy to align the mask to the substrate. Alignment is usually only carried out with semi-transparent ferric-oxide mask plates, generated by copying the original chromium mask plates using photolithography and wet etching, so that the substrate is visible through the mask using longer wavelength amber illumination which does not expose the resist. One of the machines is equipped with a camera and monitor system so that the mask and substrate can be viewed on a screen instead of through the microscope eyepiece.

Many photoresists are used within the department, but the most commonly used are AZ 1400-17, AZ 1400-31 and S1818. The first of these gives a layer 0.5μm thick when

spun at 4000 rpm for 30 seconds, while the other two give resist layers 1.8μm thick under the same conditions [Shipley[91] (1980-1993)]. For the work done here, only the last two resists were used. S1818 has been launched as a safer solvent substitute for AZ 1400-31, so that whilst the work here began using AZ 1400-31, a switch to S1818 was eventually made. However, it seems that rumours of the imminent cessation of production of AZ 1400-31 were greatly exaggerated, so that the resist is still available.

Both resists require a 30 minute bake at 90°C in a fan-assisted oven prior to exposure. Exposure time on the mask aligner ranges from 10 to 13 seconds. Both resists consist of a novolac polymer and photoactive compound (PAC). Upon exposure at a wavelength of 364nm the PAC breaks down [Pacansky and Lyerla[92] (1979)]. They are developed in alkaline solutions, containing potassium hydroxide, for between 60 and 90 seconds. Here, the unbroken PAC binds with the novolac polymer and becomes insoluble. The Microposit developer, known as MF322, comes in concentrated form and is diluted with water in the ratio 1:1. Prior to using the patterned resist pattern for pattern transfer, the resist is usually postbaked at 120°C for 30 minutes in a fan-assisted oven. As with PMMA, the resist is normally removed using acetone.

For lift-off with AZ 1400-31 an undercut profile can be generated by removing the sample from the oven five minutes before the prebake is complete, soaking the sample in chlorobenzene for 15 minutes before completing the remaining five minutes of the prebake. When exposed, this procedure results in an undercut profile. This is due to the chlorobenzene altering the resist sensitivity of the topmost layer in some way, possibly by lengthening the resist chains or inducing some cross-linking [Collins and Halsted[93] (1982)]. Using the mask-aligner, features down to 2μm can be produced reliably with these two resists, provided the exposure and development times are optimised accordingly. Features down to 1μm are possible with AZ 1400-17.

A contact printer is also available for exposure without alignment, although the uniformity of exposure for large substrates is not as good as that of the mask aligner and has been found to be unsuitable for three inch wafers.

3.6.2 Plassys MEB 450

This is an electron-beam evaporator which features a large sample to source separation and is therefore ideal for lift-off metallisations. It uses an electron beam (at an energy of 10keV) to melt the metal with the beam being scanned in the central area of the crucible to avoid any contamination of the metal vapour with materials located near the crucible edges. It also features a load-lock (fast-entry) system which allows samples to reach a background pressure of 10^{-6}mbar in 15 to 20 minutes using a combination of rotary and diffusion pumps. [Some resists, such as AZ PN114 have shown themselves to be sensitive to the ultraviolet radiation generated by the electron beam and cannot be used in this evaporator – Casey[94] (1995-1996).]

The evaporator contains six crucibles arranged on a motorised turntable so that the metallisation can consist of multiple layers of different metals. This makes the evaporator ideally suited for the generation of ohmic contacts. Since much of the department's work is carried out on III-V semiconductors, the metal choices within the evaporator are mainly dictated by the metal recipes required to generate ohmic contacts on these materials. The six available metals are:

(1) Gold
(2) Palladium
(3) Nickel
(4) Germanium
(5) Titanium
(6) Nichrome (60%/40% Nickel/Chromium)

The last two are normally used as 'sticky' layers to promote adhesion between the substrate and the noble metals, gold and palladium. Gold is used as the contact metal, with nickel and germanium being the key ohmic metals and palladium is normally used as a diffusion barrier to prevent the gold from diffusing into the semiconductor. Nichrome is also evaporated as a charge conduction layer on insulating substrates.

This evaporator is highly automated and features automatic ramping of the electron-beam power, automatic shutter control, automatic closed loop control of the evaporation rate at the Angstroms per second level (using a quartz crystal monitor) as well as automatic ramping down of the electron-beam power and switching of metal crucibles. The evaporator works by loading 'recipes' which are a series of procedures to be executed. These recipes can be altered to suit individual requirements, although there is usually at least one recipe which is close to any particular requirement. The evaporator also allows the optional use of an argon plasma clean, which is useful for the removal of polar molecules, such as water, and loose metallic fragments. However, due to the sensitivity of many of the shallow doped materials used in the department, this plasma treatment is usually avoided.

This evaporator can use several substrate holders, one which can hold four 20mm square substrates or eight 10mm square substrates, one which can hold two 75mm by 25mm glass slides, one which can hold a one inch wafer, one which can hold a two inch wafer and one which can hold a three inch wafer. The last of these is the maximum size that will pass through the load-lock mechanism. Originally, an aperture between the crucibles and substrate holders used to limit the evaporation area to just under a diameter of three inches, leaving a border of about 5mm uncoated for a whole three inch wafer. However, this aperture was eventually widened to allow almost edge-to-edge coverage. The samples must be held face down in all the holders, which means that the holders designed to hold individual samples (as opposed to whole wafers), have clamps which must be placed over the edges of the substrate and screwed down to ensure that the sample does not fall off during evaporation. This securing method can be a problem for delicate samples or

samples which do not have much area available for clamping. The whole wafer holders do not use a clamping mechanism, instead these holders have a hole with a diameter about 1mm to 1.5mm narrower than the diameter of the wafer, so that the wafer is simply placed on the holder from the rear and supported by the rim of the hole.

This evaporator is one of the workhorses of the department and is generally acknowledged to produce more reproducible results than the many more conventional bell-jar evaporators that are also available. However, the limited choice of metals mean that other evaporators are also in common use, particularly for other popular metallisations such as aluminium and chromium. Most of these other evaporators use resistive heating and the metals must be loaded prior to each evaporation. This means that they usually require several hours of pumping to reach a suitable base pressure.

3.6.3 Hitachi S800 and S900 Scanning Electron Microscopes

These two scanning electron microscopes (SEM) are used extensively for diagnosis and evaluation of fabrication processes. Both microscopes are located below ground to minimise vibration problems. The S800 allows substrates as large as two inches in diameter to be loaded and offers a practical resolution of about 50nm, this being limited by building vibrations. The S900 is a very high resolution microscope, capable of achieving a resolution of about 1nm, as a result of the anti-vibration measures that have been installed. This microscope can only handle substrates of up to a 5mm square, but the picture quality that can be achieved is far superior to that of the S800. The S800 is useful for viewing large features and has a large field of view which allows relevant areas to be found fairly easily. The S900 has a smaller field of view so that considerable time must be spent locating the relevant area. The high resolution of this system comes at the cost of a lengthy set-up time, adjusting the focus and astigmatism to get optimum performance. In addition, samples can also be cleaved and viewed side-on to evaluate dry etching profiles.

3.6.4 Plasma Technology Dry Etching Machines

The BP80 and RIE80 are both reactive ion etching machines, the latter having a water-cooled stage. Both are operated using radio-frequency sources at 13.56MHz and with r.f. powers of up to 400W. They are used with a wide range of gases, the more common ones being O_2, SF_6, CHF_3, C_2F_6, H_2, $SiCl_4$, SiF_4 and HBr. Both machines typically pump down in about 30 minutes. The BP80 is one of the more heavily used machines, typically used to pattern dielectric layers, such as silicon oxide, silicon nitride and quartz, and also for removing organic layers or etching resists. The self bias generated by these machines can be as high as -600V, but -300V is more typical.

Also available is an electron-cyclotron resonance-reactive ion etcher (ECR) with a microwave source at 2.45GHz and an independent 13.56MHz radio frequency source.

This machine also makes use of a magnetic field to allow greater control over the plasma density and is normally used when the self-bias created using the other machines results in an unacceptable amount of mask erosion.

In addition to these machines, a barrel etcher is often used for 'ashing' purposes and is run almost exclusively with oxygen. This produces an oxygen plasma which isotropically etches or removes resists and other organic layers.

3.6.5 Sputter Deposition

Sputtering facilities are also available within the department. One machine can handle three inch wafers and is used for depositing dielectrics, such as silicon oxide, and refractory metals such as platinum and chromium. Another, smaller machine, is used to deposit thin layers of gold/palladium as a charge conduction layer on substrates to be viewed in the scanning electron microscope

3.6.6 PECVD

A Plasma Technology plasma-enhanced chemical vapour deposition machine is also available to deposit silicon oxide and silicon nitride layers. A full day of deposition can result in silicon oxide layers over 6μm thick with silicon nitride layers reaching almost 2μm in thickness. Stress is a major problem with thicker layers, so that these films of this thickness may crack.

3.6.7 Silicon Oxidation Furnace

A furnace is also available for the thermal oxidation of silicon at a temperature of 1000°C to 1100°C. Both wet and dry oxides can be formed, but these furnaces are based on quartz tubes and can unfortunately only handle substrates which are no larger than 25mm square. A full day of oxidation will typically produce an oxide just over 1μm thick. A 1.2μm to 1.5μm thick oxide will typically take about 2 days to produce making thicker oxide layers impractical.

3.6.8 Rapid Thermal Annealer

A rapid thermal annealer is another commonly used machine within the department. It is capable of reaching a temperature of 1000°C in ten seconds. It is also a highly automated machine which can control ramping rates and atmosphere flushing. It makes use of the recipe concept (see §3.6.2) and is normally used to anneal metal-semiconductor contacts. It is capable of handling wafers of up to four inches in diameter, although temperature uniformity may not be as good at these dimensions. For temperatures of up to 400°C, a thermocouple is used to regulate the temperature, while for temperatures above this a pyrometer is used. Annealing can take place in a nitrogen atmosphere or in an argon atmosphere.

3.7 INITIAL WORK

This section describes some of the initial work carried out to prove that the bespoke probe concept was feasible. It is clear that one of the biggest technical hurdles requiring attention is that of carrying out electron-beam lithography on a non-planar substrate. The next section describes the work of a student who carried out a final year project designed to evaluate how difficult it would be to do electron-beam lithography on a non-planar substrate not dissimilar to that of the author. Subsequent sections describe the initial work of the author carried out at the beginning of the learning curve, and the eventual emergence from this work of a development strategy and a set of technical hurdles to be crossed in order that this project be successfully completed.

3.7.1 Feasibility Study

This work was carried out six months prior to the author's acceptance of this project. In that work the student, Shalim Rahman, was required to generate a three-dimensional substrate in silicon containing pyramidal structures, and evaluate the practicality of doing lithography on this substrate with the beamwriter by attempting to define narrow metal wires crossing over these structures using lift-off [Rahman[95] (1992-1993)].

Due to the existence of a thermal oxidation facility in-house, this work was carried out on the basis that silicon oxide be used as a mask for crystallographically wet etching the silicon (photoresist cannot survive in hot alkali for more than but a few minutes). A four inch silicon wafer was cut into nine 20mm squares which were thermally oxidised to a thickness of 320nm (on both sides). To determine which etching solution was most suitable for use with a silicon oxide mask, a survey of the published literature on wet etching was carried out. The conclusion was that caesium hydroxide offered a better ratio of silicon to silicon oxide etch rate than, the more commonly used, potassium and sodium hydroxide solutions [Clark, Lund and Edell[96] (1988)]. [Inorganic etching solutions, as discussed in §3.4.2, were ruled out by the department on safety grounds.]

An etching solution of 40% by weight caesium hydroxide solution was chosen and an etching temperature of 60°C was used. According to the published data [Clark, *et al.*[96] (1988)], the etch rate of the silicon {100} planes is about 14.4μm/hour, of the silicon {111} planes is 530nm/hour and of silicon oxide is 3nm/hour. This gives an etch selectivity of Si {100}:SiO2 of 4800:1, sufficient to etch 1536μm through a silicon <100> oriented wafer.

Initial work concentrated on producing pyramidal and tapered line structures in silicon along with alignment markers for the beamwriter. Good results were obtained using AZ 1400-31 resist spun at 8000 rpm for 30 seconds, exposed for 9 seconds and postbaked at 120°C for 20 minutes. The silicon oxide mask was then etched for 3.5

minutes in hydrofluoric acid buffered with ammonia and diluted with water in the ratio of 10 parts water to 1 part hydrofluoric acid. [This is available as a pre-mixed solution and is called silicon dioxide etch 10:1.] The resist was then stripped off using acetone, leaving a patterned silicon oxide mask. Silicon wet etching was carried out in a beaker heated using a hot-plate and without any stirring. This etching procedure resulted in an extremely rough surface, to the point where it was no longer reflective. [This was partly due to the omission of a silicon deoxidation step immediately prior to etching.] This sample was unsuitable for the beamwriter as this machine uses a laser beam reflected off the substrate surface as part of its automatic focusing system.

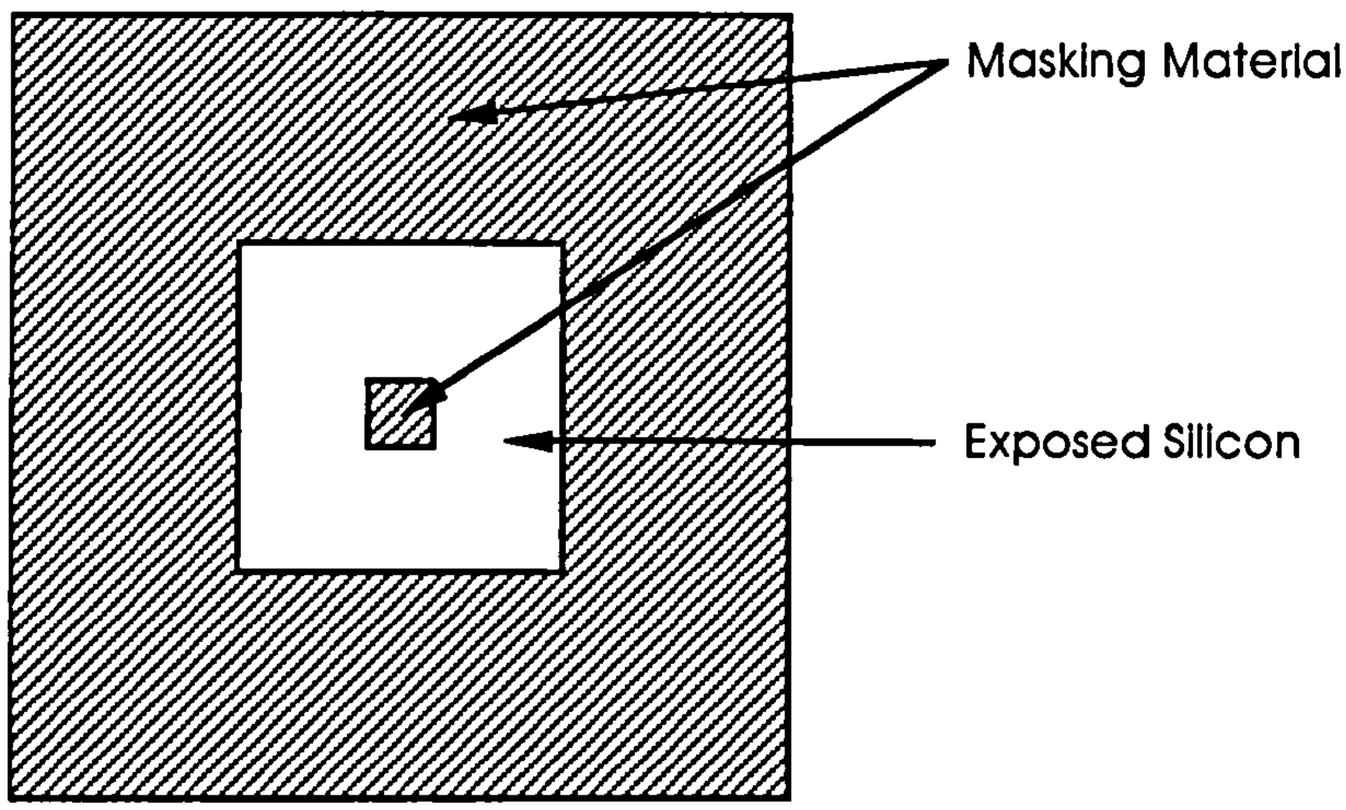


Figure 3.7.1: The pattern used by Shalim Rahman to form pyramidal structures without etching the majority of the substrate. The pyramids are located in the centre of a large clearing.

The pattern was therefore modified so that the pyramidal structures were made at the centre of localised areas of etching. This meant that the majority of the substrate would remain unetched and retain its reflective properties. Figure 3.7.1 shows how a pyramidal structure can be made in this way. The pyramids were produced from squares ranging from 5μm to 30μm in size. As time was short, the new pattern was exposed directly using electron-beam lithography, rather than by generating a new set of optical masks for photolithography. 8% HRN spun at 8000 rpm for 60 seconds and prebaked for 20 minutes at 120°C was used to pattern the silicon oxide layer. Prior to etching the silicon oxide mask, the resist was cured by a twenty minute postbake at 170°C. The mask was then etched using silicon dioxide etch 7:1 (stronger this time) and the resist was subsequently removed by ashing in an oxygen plasma. The sample was then etched as before, to a depth of 12μm. 20μm wide alignment markers were also patterned and etched simultaneously. The pattern for these simply consisted of a 20μm square opening in the masking material. This yielded substrates suitable for electron-beam lithography. The pyramids had flat tops which ranged from about 3μm to 25μm across.

Wires ranging in width from 20nm to 300nm were designed to cross over these structures. The substrate was coated with a bi-layer of PMMA: 8% BDH (85,000mwt) spun at 8000 rpm for 60 seconds, baked at 180°C for one hour and followed by 4%

Elvacite (350,000mwt) also spun at 8000 rpm for 60 seconds before being baked for another hour at 180°C. The sample was then exposed at 50kV (100kV was only just becoming available at this time) and developed in 2.5:1 IPA:MIBK for 30 seconds. Prior to metallisation the sample was ashed for 60 seconds to remove any resist residue from the exposed areas. The metallisation consisted of a thin sticky layer of titanium (probably about 12nm thick) followed by a thin gold layer (probably 15nm thick). Lift-off was carried out in warm acetone at 50°C, but stubborn areas remained due to the enhanced adhesion offered by the rougher etched areas. The sample was given ultrasonic treatment in order to clear these areas.

The results of lift-off were reasonably successful. Inspection of a sample prior to metallisation seemed to show that the resist was thicker at the edges of the flat pyramidal plateau than at the centre, as the lines seemed to fade out towards the edge. This was probably due to a combination of the high spin speed used to coat the resist and the action of capillary forces between the resist layer and the substrate. Lift-off was not very successful on these plateaux, probably due to a combination of the fading resist patterns and the application of an oxygen ash before the metallisation. However, wires appeared to cross over the tapered lines, although their continuity could not be verified due to the absence of contact pads. The alignment accuracy also appeared to be in good agreement to that obtained using more conventional metal markers. This was probably in part due to the use of the beamwriter to place the alignment markers rather than photolithography, where placement errors might have crept in.

In conclusion, the technical hurdle posed by electron-beam lithography, lift-off patterning and alignment on these substrates was not so great as to make the project unfeasible. However, the substrate differed in one crucial way from that required in the author's work: the majority of the substrate, including alignment markers, were at the same level as the substrate surface and pyramid plateaux. This meant that the substrate was smooth and reflective enough for the beamwriter autofocus mechanism to work, that resist was more likely to coat the tops of the pyramids and that the electron beam was in focus on the plateaux of the pyramids. Thus these problems would have to be overcome in the actual substrate design used for this project.

3.7.2 Fabrication Process and Development Strategy

In the beginning there was a natural learning curve that needed to be traversed, involving photolithography, micromachining, electron-beam lithography and the techniques of SPM. Therefore, while the first faltering steps along the learning curve were all unsuccessful, they served as a basis for picking a route which had the potential to achieve the objective. In order to reach a position where informed choices regarding the implementation of the fabrication scheme could be made, the first year was mainly spent exploring speculative techniques for each part of the process and noting the problems which would have to be overcome, while simultaneously eliminating certain routes from

the huge number available for exploration. This was very important if the project was to be completed within the set timescale. To blindly steer through each of the routes would have taken many years due to the wide range of options available. Several criteria were used to help choose between routes which appeared to be equally plausible, such as process stability, process cost, process compatibility, yield, volume of devices, duration of each process, total number of processes and in-house availability. Many of these were dictated by the requirements that the end result be a commercially viable product in the long run, with the flexibility and scope to adapt to the needs of the SPM community in the near future.

The first stage of this project was then to carry out very coarse development of each process after evaluating several approaches, the overall strategy being to gradually refine each process so that the final process was equivalent to a many generation evolved process, without having to refine the whole process from beginning to end for each generation. This meant that the whole project was split into sub-projects, each of which could be refined and developed individually and in parallel, with the final stages of the project then consisting of attaching all the developed processes together to produce the required result. These final stages were naturally expected to result in a few process compatibility problems appearing, so that the final stages of development would necessarily involve the process as a whole. There are several advantages to this approach:

(a) It allows for parallel development of each of the stages, and therefore shortens the development time while exposing all the problems as early as possible and allowing time and effort to be allocated to each problem in proportion to its difficulty and significance.

(b) The full knowledge and experience from each stage can be applied simultaneously to all other stages during development, so that each stage uses the latest developments, ideas and information available. This minimises the number of iterations required as each generation evolves in parallel rather than in a cascading manner where the evolution of the last stage results in an evolution of the first stage and so on.

Overall, where a project can be split into parts, and where there are many possible avenues of exploration, it is the belief of the author that this approach represents one of the shortest routes to attaining the final goal.

To start with, much of the work of Shalim Rahman had to be duplicated in the context of the thesis objective *i.e.* the results of the last section had to be explored thoroughly regarding their applicability to the fabrication scheme required here. The grand fabrication scheme as explained to me [Weaver[97] (1993-1996)] is illustrated in Figure 3.7.2. The substrate is standard <100> oriented silicon wafer with a doping type and concentration dictated by the specific microscopy application where relevant. Initial development would be carried out on one inch square pieces cleaved from a full three or

four inch wafer to maximise the number of development runs and minimise the processing and development cycle time in the early stages.

The first step is to deposit or grow a masking layer on both sides of the wafer for the bulk micromachining of the substrate. Then, using photolithography, windows would be opened in the mask for alignment markers which would ultimately penetrate the wafer and allow the through-wafer alignment of the pyramidal tip to the substrate engineered on the reverse side of the wafer – this is an example of a through-wafer alignment scheme.

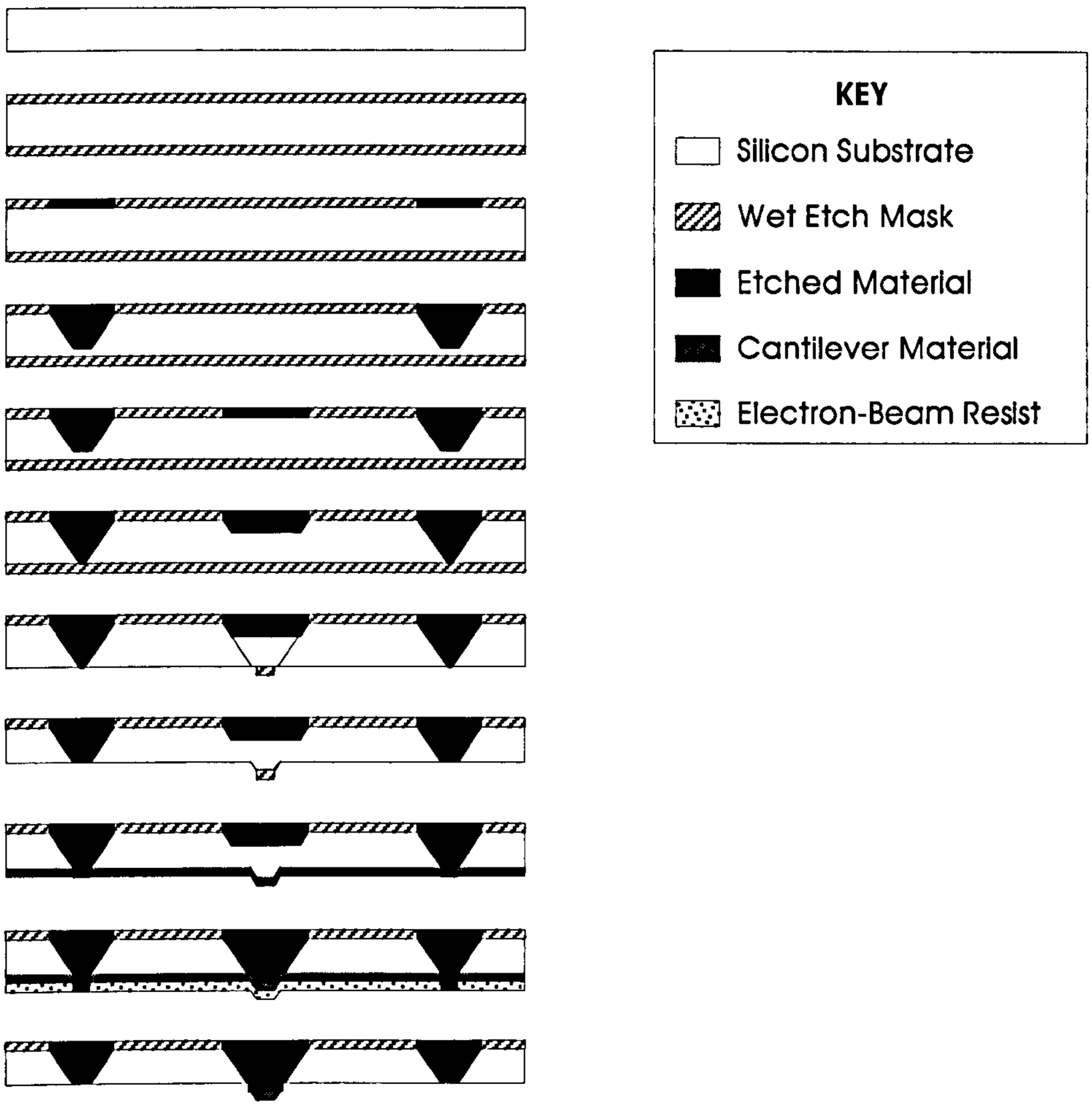


Figure 3.7.2: The device substrate fabrication method originally envisaged for the work of this thesis [Weaver[97] (1993-1996)].

The next step would be to partially etch these alignment marker windows, giving these markers a head start through the wafer over the membrane windows that would eventually result in the definition of the cantilever. These membrane windows would be opened in the next step using photolithography, by aligning to the alignment marker windows. Then the etch would be completed so that the through-wafer alignment markers penetrate the substrate while the cantilever definition windows lag behind.

As it is not envisaged that the photolithographically defined markers will be sufficiently accurate to align to any pyramidal tips, and these pyramidal tips would have to be

formed by electron-beam lithography for the ultimate in placement accuracy, an electron-beam lithography level would then be carried out, aligned coarsely to the through-wafer alignment markers (these would appear as squares of negative contrast to the beamwriter), to define both the mask for the etching of the pyramidal tips and a new set of accurately placed alignment markers. The next step would then be the formation of these pyramidal tips by wet etching, but still without letting the cantilever windows penetrate the wafer.

A layer of the material to be used for the cantilever and tip would then be deposited or grown on these pyramids. This layer would then be patterned using electron-beam lithography to form the cantilever shape (contact photolithography cannot be used here as this would destroy the tips; proximity photolithography may be possible depending on the height of the tips and the resolution required; projection photolithography would have been the best option had the facilities been readily available). Then further electron-beam lithography levels would be carried out to define the sensors on these tips. The final step would be a release etch allowing the cantilever definition windows to penetrate the wafer, thereby releasing the cantilever and tip from the remainder of the substrate.

It is clear that this process can be split into three parts, each of which is the subject of the following chapters. These are: the backface patterning and etching levels; the pyramidal tip and cantilever definition levels, and; the sensor definition levels.

3.7.3 The Photolithography Levels

The first investigation concerned the formation of membranes suitable for use as through-wafer alignment markers. This was done by extending the procedures used by Shalim in his work. A mask plate was designed for one inch square samples, cleaved from a four inch wafer approximately 500μm thick. Each device was allocated a 5 x 5 mm square area which meant that 16 devices could be produced on each sample. The pattern for each device consisted of a centrally-placed square hole, for the cantilever definition window, and four square openings placed at each corner, for the through-wafer alignment markers. The first level photolithography mask simply consisted of the openings for the through-wafer alignment markers, while the second level mask contained the cantilever definition windows in addition to these openings. The second level pattern could then be aligned to the first by lining up the edges of the through-wafer alignment markers to each other.

Since the beamwriter prefers alignment markers of 20μm to 80μm across, an average size of 50μm was aimed for to accommodate the variations in wafer thickness. Figure 3.7.3 shows the relationship between the size of opening on one side of the wafer and the resulting opening on the other side. The openings have to be aligned with the [110] directions on the surface. As silicon cleaves along the {111} planes, the major flat of a silicon wafer is aligned along the projection of this plane on the (100) surface, *i.e.* along a

[110] direction. This means that the wafers are cut parallel and perpendicular to the major flat of the wafer (the [110] axes are orthogonal on the (100) surfaces) and the openings simply have to be aligned with the edges of these wafers. To help with this process, a border was drawn around the edge of the mask pattern with the same dimensions as that of the wafer. Openings of 750μm were used for the through-wafer alignment markers, expected to result in markers about 40μm wide on the other side of the wafer. [An allowance of 10μm was made against some feature growth during the first and second photolithographic level wet etching steps.] Openings of 1000μm were used for the cantilever definition windows, expected to result in membranes about 290μm wide on the other side of the wafer (again 10μm of growth would result in 300μm membranes). This would permit the formation of cantilevers with lengths of up to 275μm.

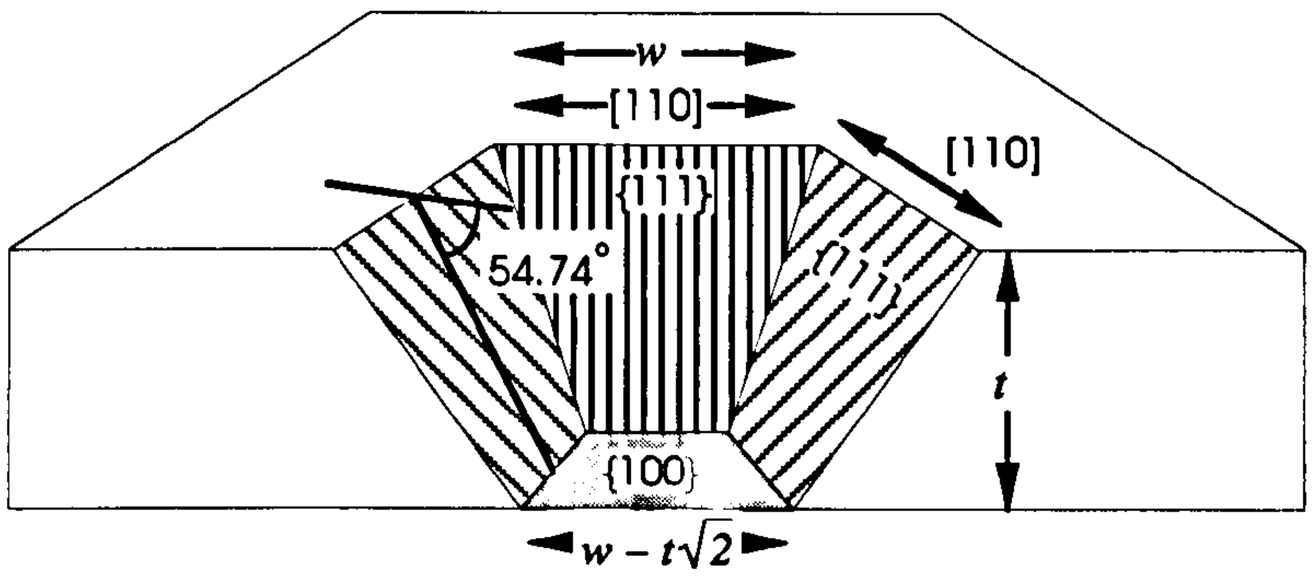


Figure 3.7.3: The geometry of silicon micromachining by crystallographic wet etching on a {100} wafer.

Since the etchant and etching conditions used by Shalim would mean that almost 35 hours of etching time would be required to go from one side of the wafer to the other, the conditions were changed in an attempt to speed up the process. This meant raising the etch temperature, which also unfortunately increases the mask erosion rate [Clark, *et al.*[96] (1988)]. As 50% by weight CsOH solution was easily available commercially, it was decided that this concentration be used at the highest temperature that etching data was to hand, *i.e.* 90°C. From this data a silicon oxide etch rate of 87nm/hour, a {111} etch rate of 840nm/hour and a {110} etch rate of 120μm/hour was expected [Clark, *et al.*[96] (1988)]. If the {100} etch rate was comparable to the {110} etch rate, then almost 1000μm of silicon could be etched with a 700nm thick silicon oxide mask. In this time the {111} planes would be etched by almost 7μm, resulting in a growth projection along the [110] axes of about 5.5μm.

Some one inch pieces from a both-side polished *n*-type silicon wafer were found and thermally oxidised to a thickness of 700nm. These pieces turned out to be 512μm thick. First level photolithography was carried out and windows opened in this masking layer as follows:

(a) AZ 1400-31 photoresist was spin coated on one side at 4000 rpm for 30 seconds (protection for opposite side of wafer) [nominally 1.8μm thick].

(b) The resist was baked at 90°C for 5 minutes.

(c) AZ 1400-31 photoresist was then spin coated on the other side of the wafer at 4000 rpm for 30 seconds [nominally 1.8μm thick].
(d) The resist was then prebaked at 90°C for 20 minutes.
(e) The resist was exposed for 13 seconds using the mask aligner.
(f) The resist was developed in AZ 1400 developer:water 1:1, rinsed in water and blown dry.
(g) The resist was then postbaked at 120°C for 30 minutes.
(h) The silicon oxide was then etched in silicon dioxide etch 4:1 for approximately seven minutes (until patterned areas became hydrophobic – this is a useful feature of silicon oxide etching on silicon), rinsed in water and blown dry.
(i) The resist was stripped off in nanostrip:water 1:1 (nanostrip is a proprietary mixture of concentrated sulphuric acid and hydrogen peroxide which heats up to over 120°C when diluted with water and is very good at removing organic materials).
(j) The substrate was then cleaned in IPA and blown dry.

All water used in the above procedure was ultra-pure water, often known as reverse-osmosis or 'RO water' due to most metal ion contaminants having been removed from it.

This substrate was then wet etched for four hours at 90°C in 50% by weight caesium hydroxide solution, and resulted in etch pits approximately 270μm deep (as measured using an optical microscope with a calibrated focus knob, by focusing successively on the non-etched and etched areas and noting the height difference). This gave a {100} etch rate of approximately 65μm/hour, almost half the {110} etch rate. All etching was carried out in a custom designed etching kit which is described in more detail in Appendix A (§A.1).

Then the second level of photolithography was carried out in similar fashion to the first, but using the second level mask, and the sample was wet etched for another four hours under the same conditions as before. However, it is clear from Figure 3.7.4a that something went wrong as the through-wafer alignment markers ended up 150μm to 200μm across, terraces were present on the walls of the etch pit, the openings were no longer square, white specks of some material had condensed onto the substrate and the silicon oxide masking layer exhibited multi-coloured fringes at the opening edges (indicating it was of non-uniform thickness).

Many reasons could be given for this result including, contamination of the etch by another user, erosion of the silicon oxide masking layer to an inadequate thickness (calculation shows that 696nm of silicon oxide out of 700nm should have been etched) and, the non-uniform patterning of the silicon oxide masking layer at the second level due to a lack of resist adhesion (so that the silicon oxide etch could access the masking layer for some way beneath the resist). One of the most probable causes for this, other than the use of a masking layer of insufficient thickness, was discovered later and will be

discussed in §4.2.3 in more detail. The white specks were by crystals of either caesium hydroxide or silicate reaction products being left behind on the surface after blow drying. This seems plausible from examining Figure 3.7.4b which is a higher magnification view of these specks. Such precipitates have been seen before [Bassous[40] (1978)].

Several solutions were possible, including a reduced wet etching time for the first level silicon etch, using thinner substrates and using a thicker silicon oxide masking layer. The last of these was not as attractive, because the time required to etch 1μm of silicon oxide would be of the order of 10 minutes using silicon dioxide etch 4:1. This length of time was generally acknowledged to be beyond what photoresist can reliably withstand [Pollock[98] (1993-1996)].

By cleaving the one inch pieces from a three inch wafer, the wafer thickness is automatically reduced by about 100μm [see §2.1]. Thus some *p*-type wafers were found, only one side polished, and oxidised to 700nm thick. It was noted that since the electron-beam lithography levels require a flat surface and the photolithography levels consist only of relatively large features, the polished side ought to be used for the cantilever and tip with the unpolished side used in the photolithography levels.

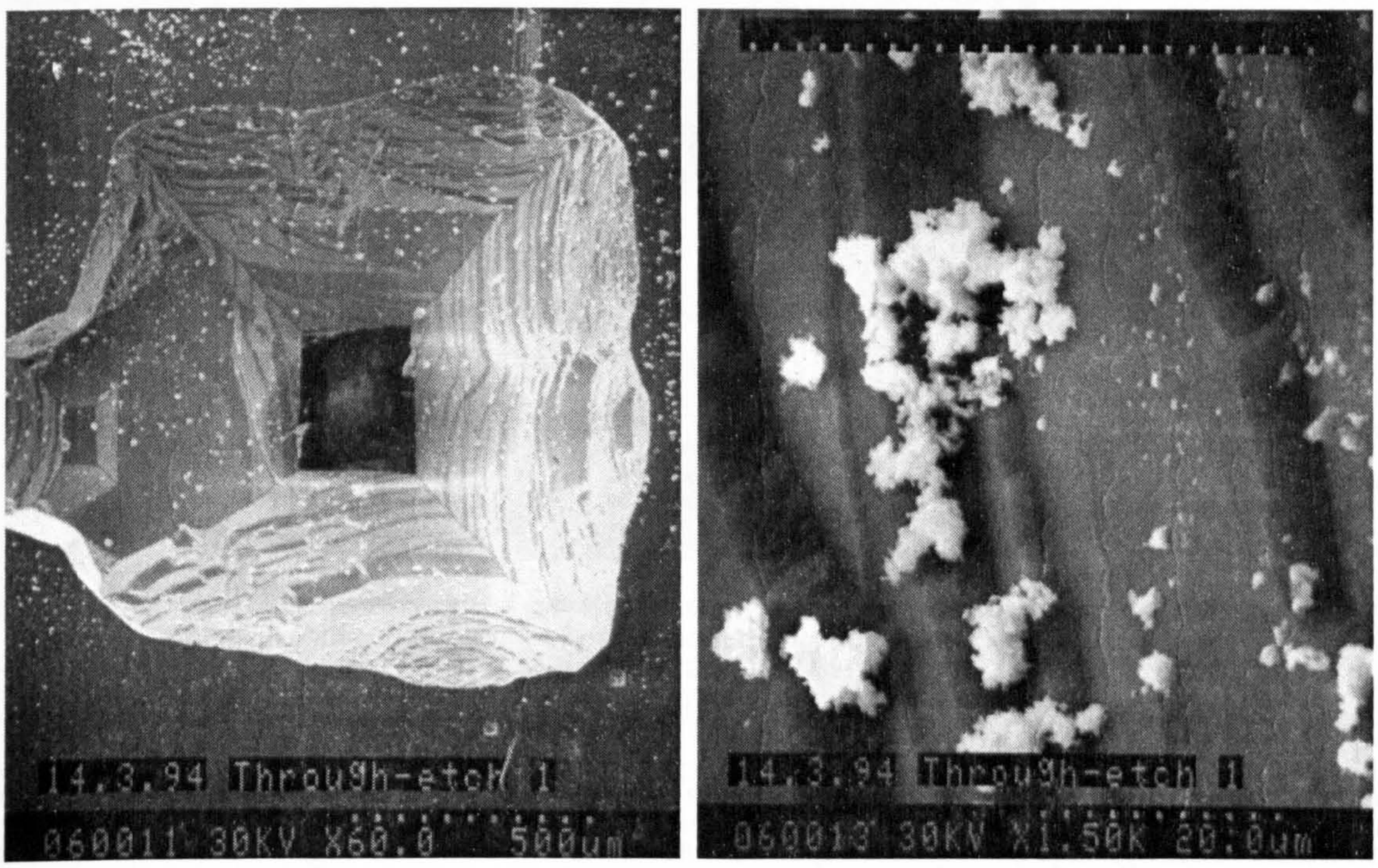


Figure 3.7.4: Micrographs illustrating the results of the author's first attempt to wet etch silicon. (a) Left Image: a through-wafer alignment marker after the second level silicon wet etch; (b) Right Image: a close-up view of the white crystals found all over the substrate.

Upon browsing through a book on the properties of metals [Lyman[99] (1961)], it had been discovered that nickel and chromium are two non-noble metals which are very resistant to attack by hot alkalis. In particular, nickel is used as one of the electrodes in NiCad batteries which are based on a potassium hydroxide electrolyte. Since the silicon oxide on the polished face would be used to form the tip, it was decided that a layer of nickel be

evaporated on top of this for added protection against the caesium hydroxide solution. A 200nm thick layer was evaporated and an adhesion test (to make sure that the layer did not fall off during etching) was carried out by placing the sample in an ultrasonic for 30 seconds. Unfortunately, the nickel layer failed this test by peeling off – not unsurprisingly as nickel is not generally used as a metal sticking layer.

However, because chromium is one of the most 'sticky' metals in common use and is very resistant to alkalis, 50nm of nichrome (Ni/Cr 60%/40%), an alloy of nickel and chromium which also functions well as a sticky layer, was used to stick down a 200nm nickel layer. This dual metal layer survived the ultrasonic test. This sample, patterned using S1818 resist this time [see §3.6.1], was then etched for three hours using a fresh bottle of caesium hydroxide solution.

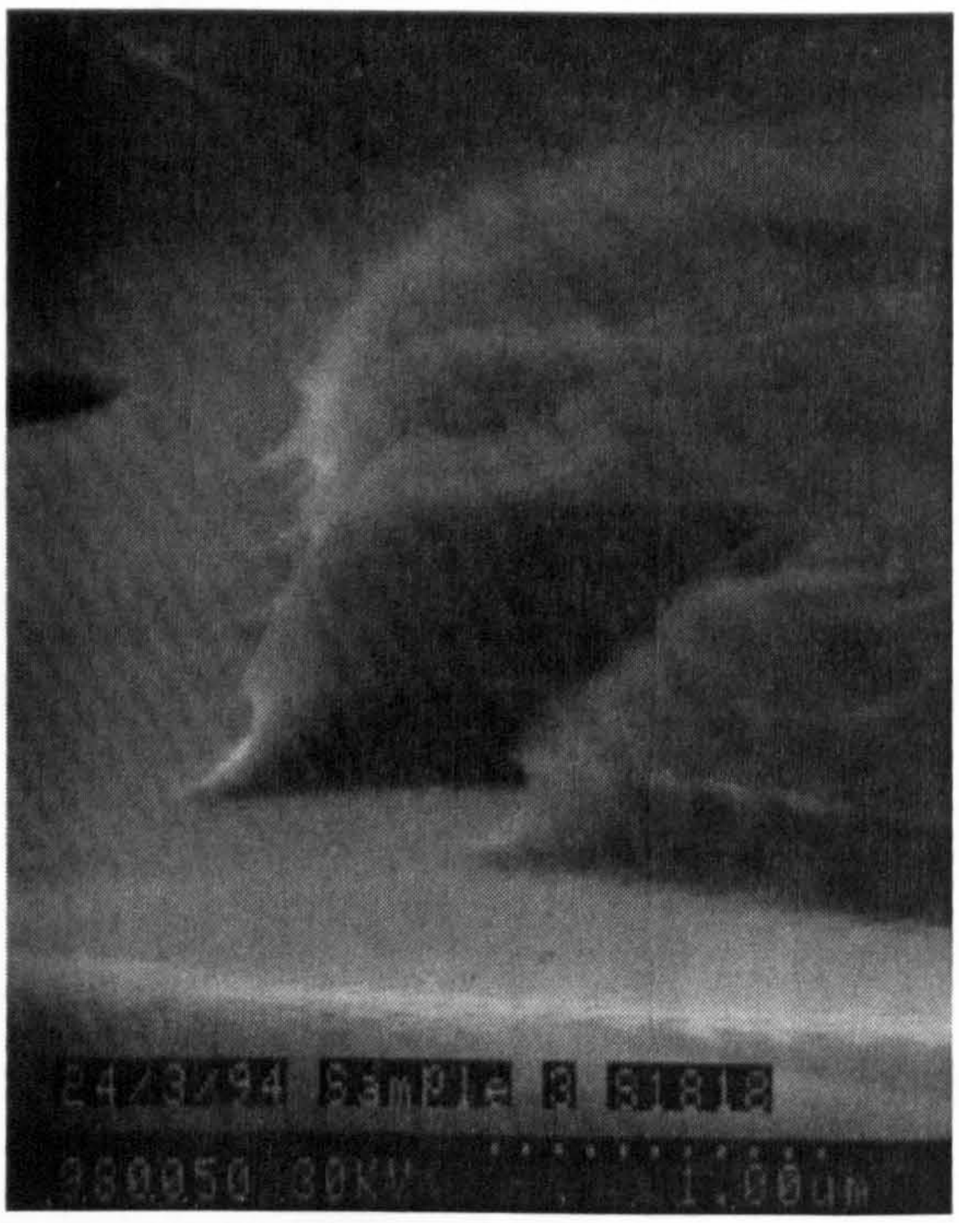


Figure 3.7.5: Micrograph illustrating the lack of etching when a protection layer was used. The rough edge of the silicon oxide can be clearly seen. It would appear that perhaps some slight etching of the silicon has taken place as a layer undercutting the silicon oxide can be clearly made out. However, there appears to be no evidence of crystallographic etching behaviour here, so this conclusion cannot be confirmed.

However, after the sample was removed, there appeared to have been very little etching of the silicon although the protection layer still seemed to be intact. Figure 3.7.5 shows the edge of the masking layer. A number of possibilities could explain this:

(a) There was some evidence that the resist had been underexposed (as the resist had been changed from the AZ 1400-31 to the safer S1818 equivalent) and so the silicon oxide masking layer had not been fully etched. This is unlikely as, by calculation, 261nm of silicon oxide should have been etched by the caesium hydroxide in this time.

(b) There was something wrong with the new etching solution.

(c) The doping type of the wafer had had a dramatic effect on the etch rate of silicon with this solution.

(d) Some kind of electrochemical cell had been set up which was somehow protecting the silicon.

The wafer was then etched for a further three hours, again with no result. Possibilities (b) and (c) were then ruled out by etching some bare silicon from the same wafer in this etching solution – the wafer etched at a rate of about 65μm/hour.

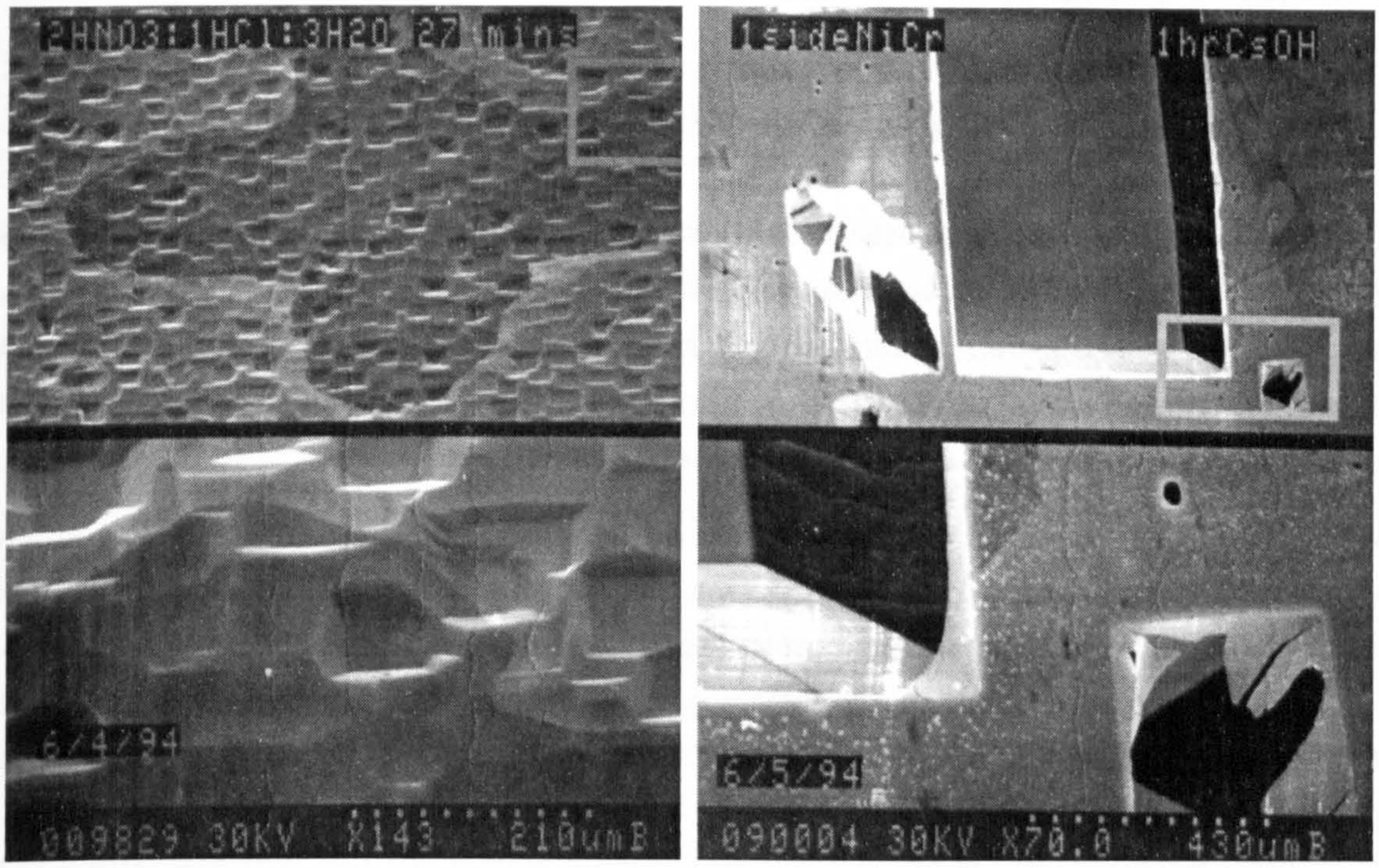


Figure 3.7.6: Micrographs illustrating the problems with using metallic layers to mask the wet etching of silicon. (a) Left Image: the effects of undercutting the photoresist masking layer in the wet etching of the nickel masking layer; (b) Right Image: the problems caused by the presence of pinholes in a nichrome layer used for masking the silicon wet etch. The bottom half of the image is a 5x magnification over the displayed scale which applies to the top half.

This sample was then abandoned and as the protection layer seemed to be so resistant, an attempt was made to use nickel in place of silicon oxide as the masking layer. It was found that nickel could be made to stick to bare silicon, even through ultrasonic treatment, if the sample was deoxidised in silicon dioxide etch immediately prior to evaporation. However, the main problem was that nickel etching required the use of aggressive chemicals. One of the etching solutions used in-house for nickel consisted of nitric acid:hydrochloric acid:water in the ratio 2:1:3 [Griffith[100] (1993-1996)]. It was found that etching nickel with this solution, using a patterned S1818 resist layer as a mask, required a variable length of time from run to run, ranging from 10 minutes to 20 minutes. The reason for this is not known. However, this etch seems to undercut the resist significantly, as illustrated by Figure 3.7.6a which shows the result of etching a 500μm long cross, 25μm wide. However, while the metal seemed to adhere well through

ultrasonic treatments, it peeled off within a few minutes of being placed in the caesium hydroxide etching solution. Some possible explanations for this are:

(a) insufficient adhesion;
(b) the difference in thermal expansion between silicon and nickel is large enough to cause cracking of the metal layer at 90°C;
(c) the native oxide between the silicon and metal layer or the metal-silicon bonds are attacked by the etchant and rapid undercutting takes place, resulting in the metal layer 'lifting off'.

To exclude the first possibility, an attempt was made to use nichrome as a substitute. The nichrome used in most parts of the department consists of 95% nickel and 5% chromium. It is found that nickel etch has to be used to pattern this. However, the nichrome used in the Plassys MEB 450 (also used to deposit the protection layer) is richer in chromium (40% chromium) so that chromium etch can be used. This etch, which consists of 150 grams of ceric ammonium nitrate to 35 millilitres of glacial acetic acid and one litre of water, is the same as that used to pattern mask plates and does not attack or undercut resist significantly [Harkins[101] (1993-1996)].

Therefore this nichrome was used and patterned as stated above. This produced a much more faithful masking pattern. However, upon wet etching, the silicon once again seemed not to have etched appreciably, even after a total of six hours immersion in the solution. To eliminate the possibility of this being an etch specific effect, an attempt was made to etch this sample in 7 molar potassium hydroxide solution at 45°C. Again the silicon appeared to have been virtually unetched.

Since a protection layer was again present on the other side of the sample, the question of whether some kind of electrochemical protection was taking place arose. This seemed highly unlikely as the electrochemical etching of silicon usually requires the application of some bias to the silicon from an external source [Soderbarg and Smith[102] (1993); Mastrangelo, *et al.*[13] (1994)], although a reference electrode is not always necessary. However, as no other more plausible explanation could be found, it was decided that this possibility should be investigated further. With the preceding sample, it was speculated that the electrochemical cell could have consisted of a *p*-type silicon electrode and a nichrome/nickel electrode in a caesium hydroxide electrolyte. These electrodes would only be separated by the thin silicon oxide layer between them. For electrochemical behaviour to take place, there must be a bias between the electrodes. However, no bias was being applied, unless the nichrome had somehow made contact with the silicon and formed a Schottky barrier. This would result in different electrode potentials. It is known that the activation energies for the etching of the {100} in silicon is in the region of 0.4V to 0.7V [Mastrangelo, *et al.*[13] (1994)] – of the same order as the barrier voltage that might be expected.

To test this theory, a check was made to see whether the metal layers on both sides of the present sample made electrical contact with each other using a simple multi-meter. The result was clearly positive. This was probably because the edges of the sample are {111} planes which lie at a shallow angle with respect to the plane of the substrate, so that some metal would lie on the edges of the wafer after evaporation. This sample was then returned to the potassium hydroxide solution for another hour of etching. At this point the sample appeared to start etching rapidly. However, the results were poor due to the masking layer having been scratched during probing with the multi-meter. It may be hypothesised that these scratches resulted in an ohmic contact forming between the metal and silicon which destroyed any potential between them and permitted etching to begin.

The experiment was then repeated using nichrome layers on both sides of the wafer and a one hour caesium hydroxide etch showed that, once again, no etching had taken place. As an experiment, the whole process was repeated without a protection layer. In this case etching did begin immediately and resulted in the etch pits shown in Figure 3.7.6b. It is possible that the wafer was scratched so that an ohmic contact was once again formed or that there is some significance to there being a metal layer on the opposite side of the wafer from that being etched.

It should be stressed that the preceding analysis is purely speculative. However, the phenomenon of *p*-type silicon passivation, when part of a *p*-*n* junction and under conditions of no external bias, has been seen before, and is caused by electrons transferring from the *n*-type layer to the *p*-type layer finding a condition of lower energy at the interface between the *p*-type silicon and the etching solution, rather than in the normal depletion region. This forms an inversion layer at the silicon-etching solution interface which protects the *p*-type silicon from attack [Ozdemir and Smith[103] (1992)]. Since it is known that both nickel and chromium intermix with silicon even at room temperature, and both act as electron donors to silicon [Franciosi, Peterman and Weaver[104] (1981); Grunthaner, Grunthaner, Madhuker and Mayer[105] (1981)], the possibility that the nichrome was acting as the *n*-type layer of a *p*-*n* junction with the *p*-type silicon cannot be ruled out. However, what remains unsatisfactory is the need for the protective metal layer on the non-etching side of the wafer, either partially or fully insulated from the silicon by thermally grown silicon oxide. Clearly, without a more thorough investigation no significant conclusions can be drawn. In any case, the pinholes present in the metal film and the time being spent on this stage of the process meant that this strategy was abandoned in favour of more conventional techniques.

Silicon nitride deposited at low pressure by chemical vapour deposition (LPCVD) acts as an excellent diffusion barrier against water, sodium and potassium [Sze[106] (1985); Mastrangelo, *et al.*[13] (1994)]. This combined with its extreme inertness makes it one of the most reliable masks for micromachining silicon. It has the advantage of allowing the use of more inexpensive etching solutions such as potassium and sodium hydroxide,

which can be made up in-house. It was decided that potassium hydroxide be used, both because it is one of the most stable silicon wet etching solutions [Bassous, *et al.*[41] (1978)] and because it can produce very smooth etching without residues being deposited on the substrate [Lambrechts and Sansen[107] (1990-1993)]. Unfortunately the departmental facilities are limited to plasma enhanced chemical vapour deposition (PECVD). So an investigation into the ability of PECVD silicon nitride to resist this etchant was carried out. It was found that 80nm of silicon nitride was etched in approximately 2.5 hours using 7 molar potassium hydroxide solution at 50°C. At this point a source of LPCVD silicon nitride coated silicon wafers was found with consequences which are discussed in more detail in Chapter 4.

3.7.4 The Electron-Beam Lithography Levels

The next part of the process was to investigate the electron-beam lithography levels, including the fabrication of the tip and associated alignment markers, the feasibility of patterning onto these tips and the alignment accuracy that could be achieved.

Normally the first electron-beam lithography level would involve alignment to the through-wafer alignment markers. Since this part of the process was not yet developed, initial investigations were carried out without through-wafer alignment, by simply writing electron-beam lithographic markers in a dummy level for use in the following electron-beam lithography level.

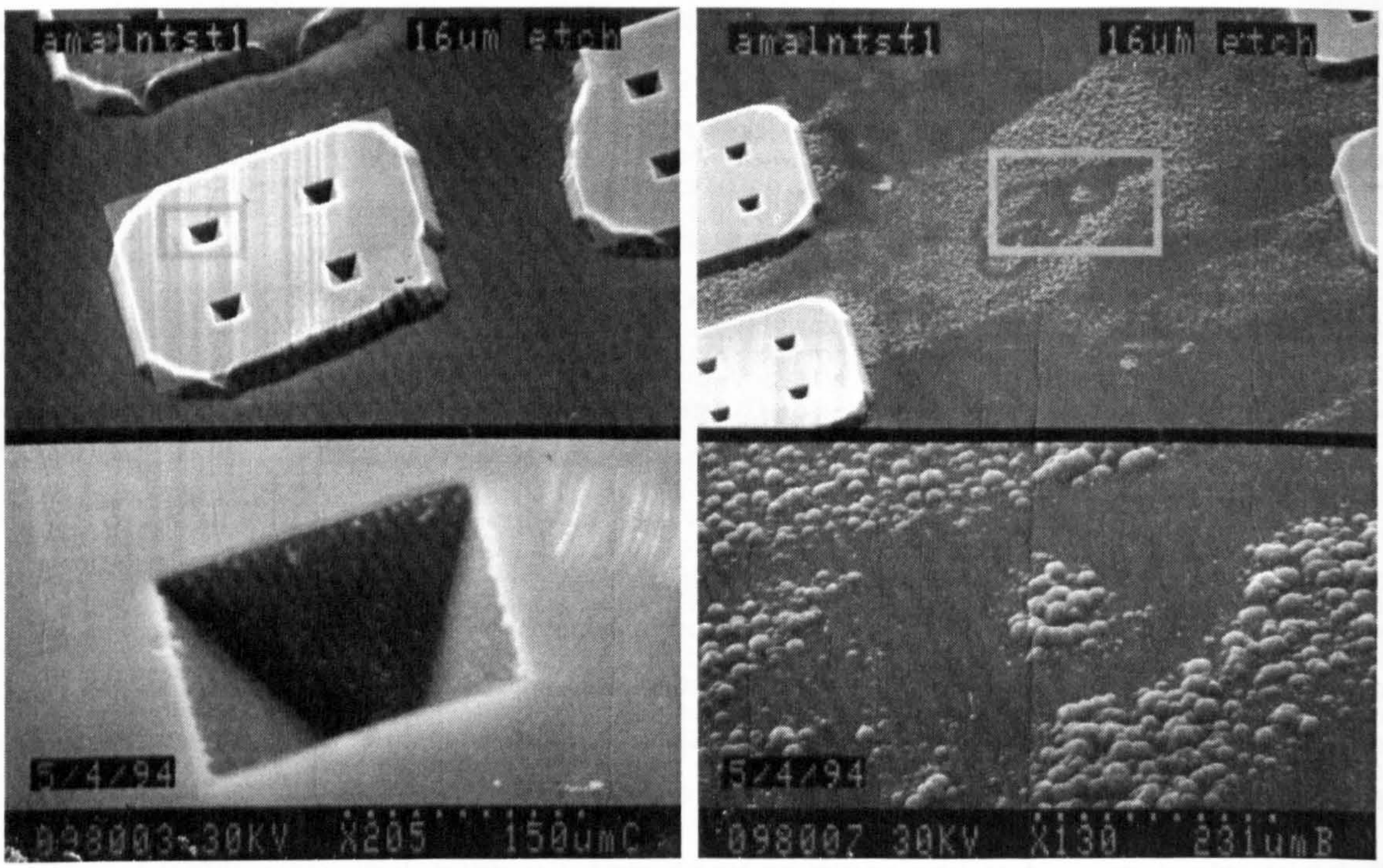


Figure 3.7.7: Micrographs illustrating the first attempt to etch pyramidal tips. (a) Left Image: an etched alignment marker structure. The effects of convex corner undercutting are apparent. The bottom half of this image is a 10x magnification over the displayed scale; (b) Right Image: areas of roughness after the etch. These areas were much less reflective and could be seen with the naked eye. The bottom half of this image is a 5x magnification over the displayed scale.

Again proceeding in light of Shalim's work, this first step consisted of using 8% HRN resist to pattern the silicon oxide layer for use as a mask in the silicon etch that defines the tips and topological markers. The marker pattern consisted of square areas 200μm across, each with a set of four 20μm alignment markers (non-written square areas), located 50μm from the edge. These structures were placed at each corner of a 1mm square area allocated to each device. A 50μm square was placed at the centre of each device to form a mask for the definition of a pyramidal tip. This meant that the majority of the substrate would be etched, potentially causing problems to the height meter of the beamwriter as mentioned previously in §3.7.1. As the pyramidal tip has to protrude from the surface in order that the tip is able to access the surface of a specimen, allowing for a 10% incline in a 200μm long cantilever requires the tip to be 20μm tall.

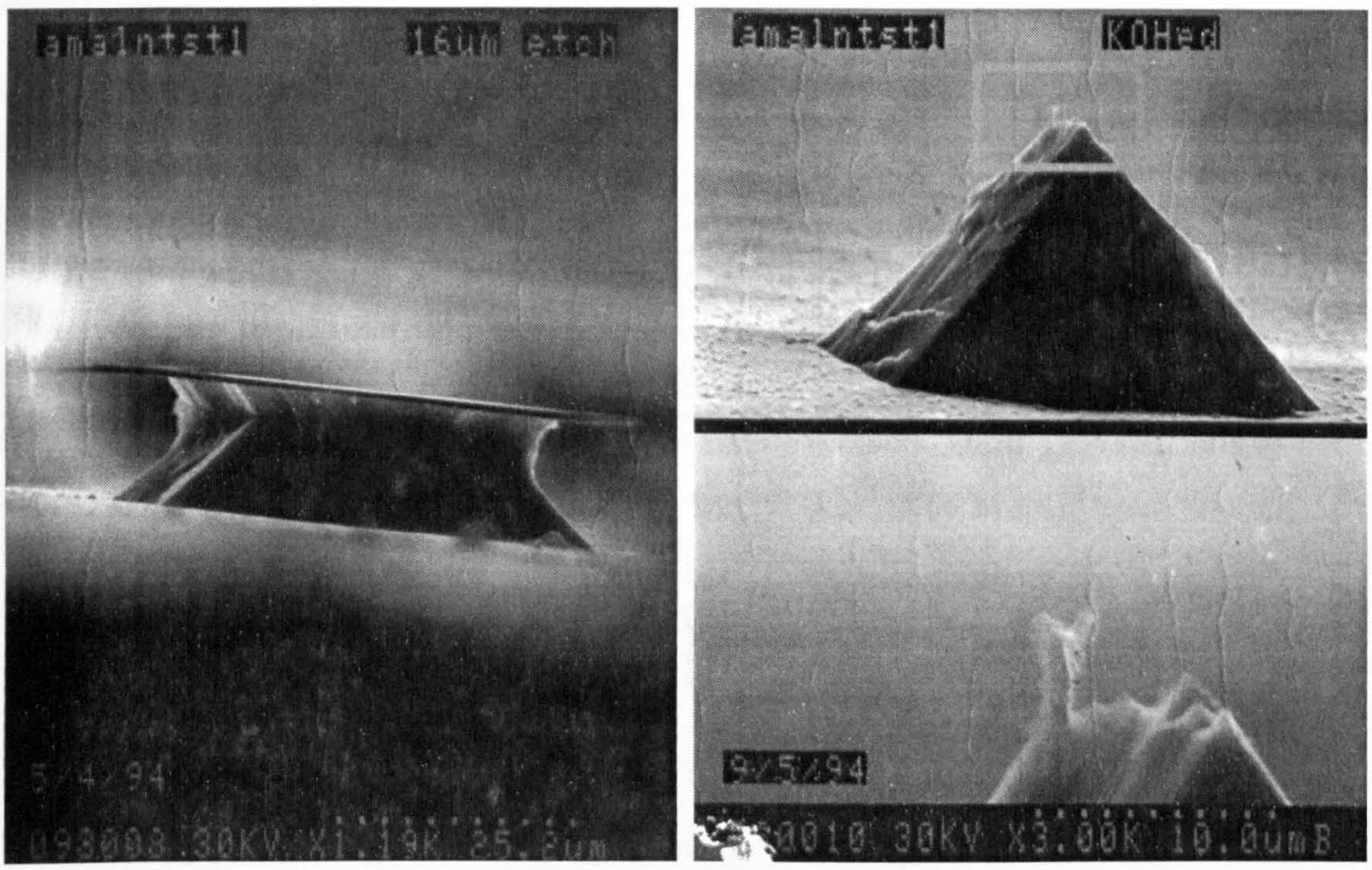


Figure 3.7.8: Micrographs illustrating an unexpected problem with etching pyramidal tips. (a) Left Image: a side-on view of an etched pyramidal tip with the silicon oxide mask still present; The 'egg-timer' profile is unexpected and undesirable. (b) Right Image: a close-up of a pyramidal tip which has 'perforated'. The bottom half of the image is a 5x magnification over the displayed scale which applies to the top half.

One inch square silicon pieces from the 400μm thick wafer as used in the previous section were oxidised to a thickness of 700nm as before. The following procedure was used to pattern these:

(a) 8% HRN was spin coated at 5000 rpm for 60 seconds [nominally 120nm thick].

(b) The sample was prebaked for 20 minutes at 120°C.

(c) The pattern was written using the beamwriter at 50kV with a dose of 300μC/cm^2 and a spot size of 112nm.

(d) The resist was developed by four ten second dips alternating between MIBK and IPA.

(e) The resist was then cured by postbaking for one hour at 180°C.

(f) The silicon oxide mask was then wet etched in silicon dioxide etch 4:1 for approximately seven minutes (using the wetting test to indicate that the mask had been etched through).

The substrate was then wet etched in 7 molar potassium hydroxide solution at 50°C for approximately one hour resulting in an etch depth of about 16μm. Figure 3.7.7a illustrates an etched alignment marker structure and the effects of convex corner undercutting [see §3.4.2] can be clearly seen against the overhanging silicon oxide masking layer. Figure 3.7.7b shows that some areas were disproportionately rough. This may have been due to the fact that the resist was not stripped off prior to etching and may have broken up during the etch, resettling on the substrate and acting as a masking layer for a short time.

During the etching of the pyramids, an interesting, but problematic effect was noticed. Figure 3.7.8a illustrates this effect which appears to result in a crystallographic 'egg-timer' like profile. As the etching time required was not yet known and the pyramids had to be made reasonably sharp (with a radius of curvature of the order of one micron), the samples were repeatedly etched and removed for examination until the pyramids appeared to be of the correct size. It can be seen from Figure 3.7.8b that what appears to happen is that the 'egg-timer' becomes narrower until the upper part of the pyramid is ripped off. This may have occurred during the blow drying step used upon removal of the sample from the etch or within the etch itself.

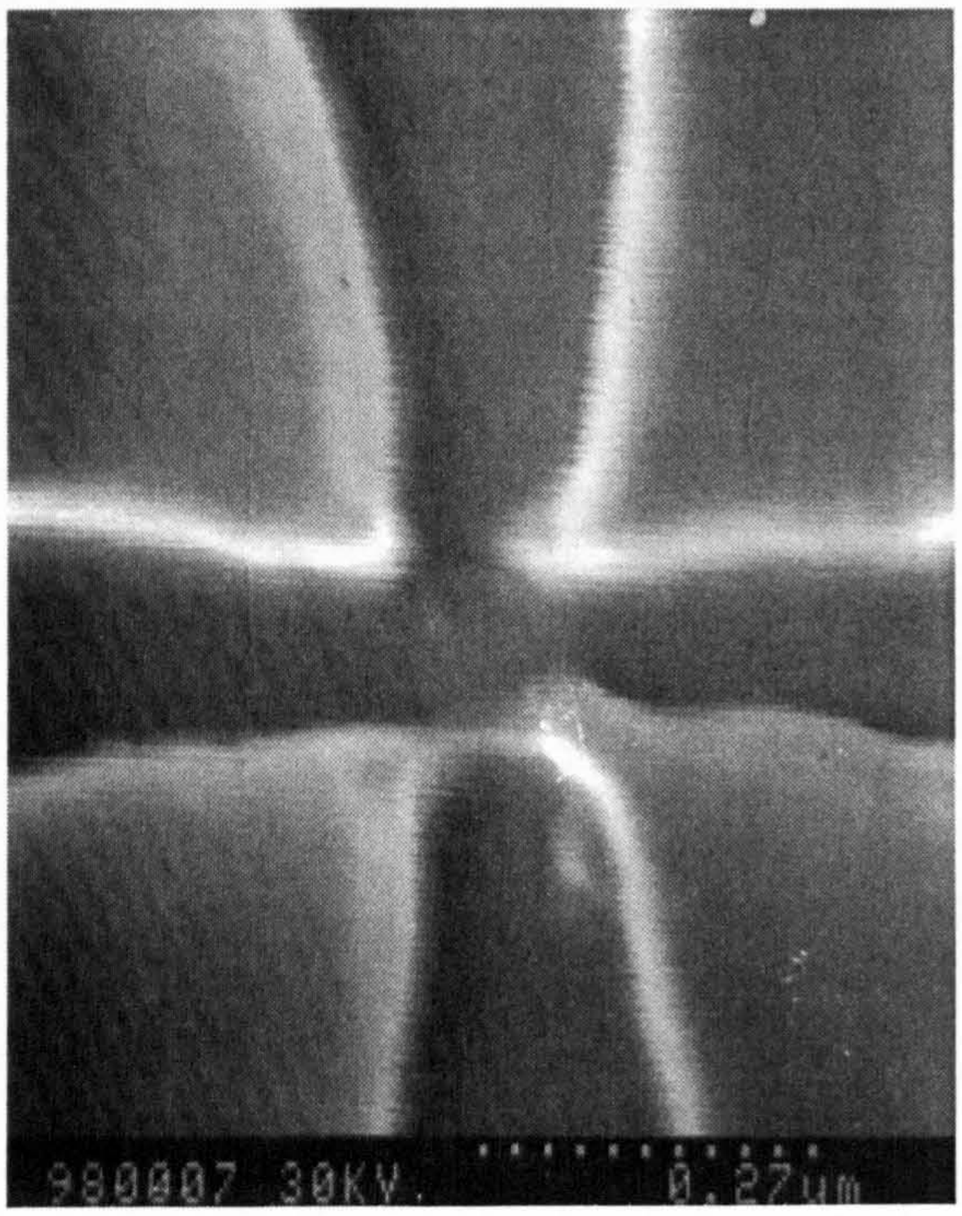


Figure 3.7.9: Micrograph illustrating the resist structure achieved on top of a pyramidal tip at the first attempt. Unfortunately the central 'crossover' is wider than desired due to the superposition of single pixel dots over this pattern.

However, even more worrying was the fact that there appeared to be a lack of consistency amongst the pyramidal tips, even though a sufficient number looked good enough for an initial attempt at the next level. This attempt consisted of writing single pixel dots and wires (the latter also being accompanied by small pads) as would be required for the SNOM and thermal/Hall-effect probes respectively [see §3.2]. In this context, a single pixel refers to the fractionation resolution of the pattern which in this case was set to the minimum of 5nm. At this scale it is the dose and spot size which define the size of the resulting pattern. As the dose required for patterning at the top of these pyramids was not known, the pattern was run as an exposure test, where the dose is incremented by a multiplicative factor from one device to the next. The resist used was a bi-layer consisting of 8% BDH and 4% Elvacite spun at 5000 rpm for 60 seconds with a one hour bake at 180°C between each layer. The sample was then baked at 180°C overnight.

The sample was originally meant to be written at 100kV with a 12nm spot. However, alignment failed due to insufficient marker contrast, so the sample was resubmitted for exposure at 50kV with a 15nm spot. This also failed to run initially because of a software related problem. The problem was eventually located and rectified, however another error resulted in the single pixel dots and wires being written on top of each other. The sample was cleaved into four pieces prior to development so that each sample could be processed separately. The resulting compound structure is illustrated by Figure 3.7.9.

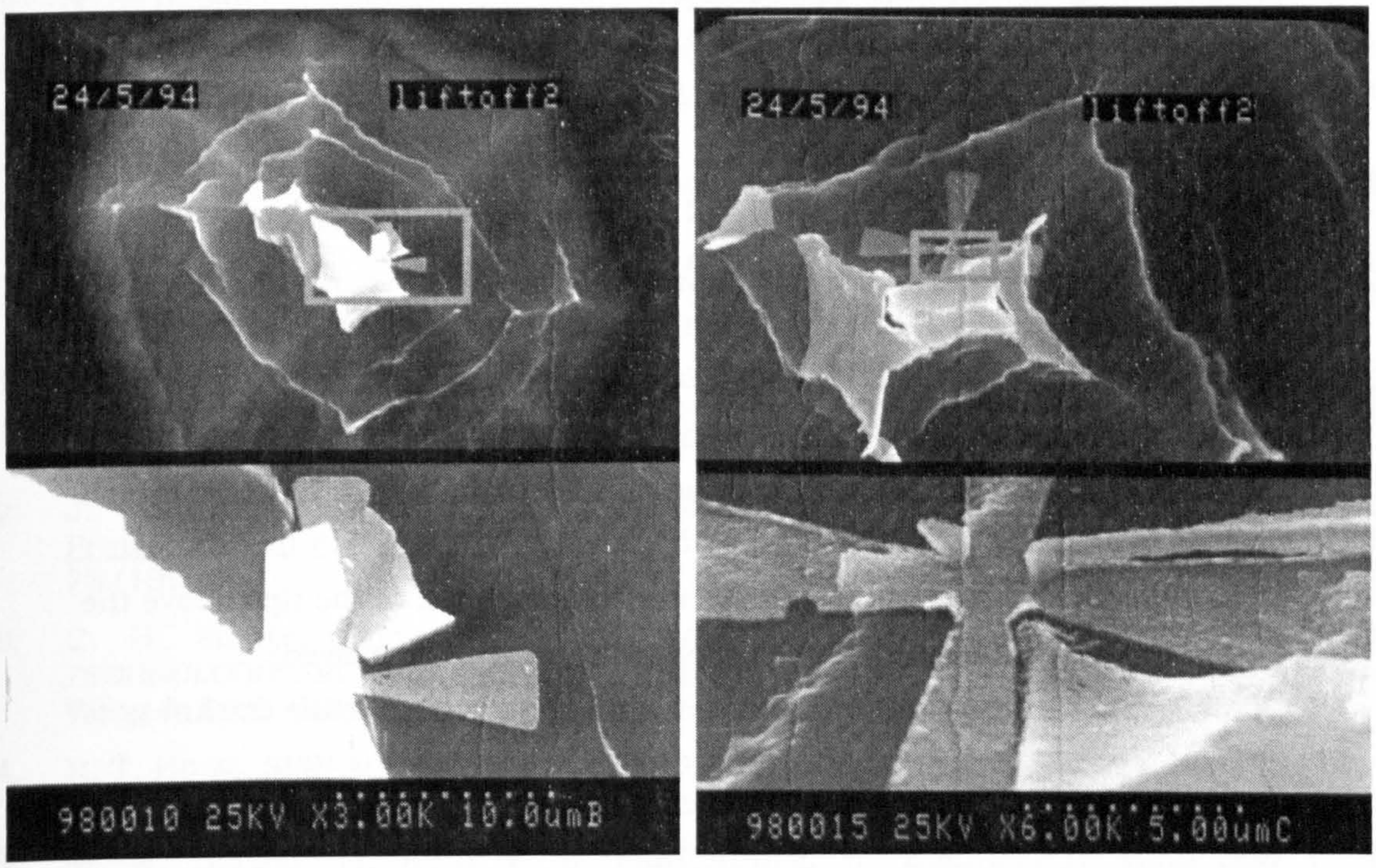


Figure 3.7.10: Micrographs illustrating the first attempted lift-off of metal structures on a pyramidal tip. (a) Left Image: one side of the cross structure has lifted off properly. The metal has not fully detached itself from the remaining sides. The bottom image is a 5x magnification over the displayed scale; (b) Right Image: two sides of the cross structure have lifted off correctly, but one stubborn corner has prevented successful pattern transfer. The bottom half of the image is a 10x magnification over the displayed scale which applies to the top half.

On one of the pieces, an attempt was made to carry out lift-off using a 12nm sticky layer of titanium and a 15nm layer of gold. This was not very successful, probably because the surface roughness on the pyramidal tip caused the metal layer to adhere too well. More success was achieved on another piece where an argon plasma clean was carried out prior to evaporation and ultrasonic treatment was used to aid lift-off. This dramatic attempt resulted in more promising probes as illustrated in Figure 3.7.10. The terraced structure of the pyramid faces were caused by breaks in the etching and can be clearly seen in Figure 3.7.10a. In general, the junction sizes ranged from 60nm to 100nm across and the resist thickness on the pyramid appeared to be between 60nm and 80nm thick. However, it was expected that with better pyramids and the spot focused at the top of the pyramid, the results would improve dramatically.

3.8 CONCLUSIONS

From the previous sections, several conclusions were drawn. These conclusions are listed in terms of hurdles to be overcome before the project goals could be achieved. It is clear that while this initial work did not expose *all* the problems that had to be tackled, it did reveal many of the significant ones:

(a) A reliable masking technique for micromachining the substrate and mask-patterning method needed to be found.

(b) A process for the formation of reliably sized through-wafer alignment markers and to which the beamwriter could align needed to be developed.

(c) A set of EBL patterned markers which could be used for accurate alignment to the pyramidal tips had to be designed and demonstrated to work at 100kV as well as the 50kV already achieved.

(d) A reliable process for fabricating pyramids with well defined apices, centres and heights had to be developed, as it is clear that the fracturing process cannot be relied upon to deliver tips properly centred with respect to the alignment markers, let alone a surface adequate for electron-beam lithography.

(e) A method for defocusing the beamwriter by the height of the tips above the substrate had to be found (once problem (d) had been resolved).

(f) A method for coating the tips with a reliable and reproducible thickness of resist had to be found.

Some other problems that could be easily envisaged at the time included: the demonstration of the wet etch releasing the cantilevers from the substrate; the formation of metallic apertures at the apices of the tips, and; the formation of wires by lift-off at the apices of the tips. Problems associated with the choice of sensor materials and process compatibility, since such materials have to function as sensors while also surviving the release etch, also had to be resolved.

3.9 REFERENCES

1. M. J. Vasile, C. Biddick and H. Huggins, *FORMATION OF PROBE MICROSCOPE TIPS IN SILICON BY FOCUSED ION-BEAMS*, Applied Physics Letters **Vol. 64,** No. 5, pp. 575-576 (1994).

2. M. J. Vasile, C. Biddick and S. A. Schwalm, *MICROFABRICATION BY ION MILLING: THE LATHE TECHNIQUE*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 4, pp. 2388-2393 (1994).

3. M. Yamaki, T. Miwa, H. Yoshimura and K. Nagayama, *EFFICIENT MICROTIP FABRICATION WITH CARBON COATING AND ELECTRON BEAM DEPOSITION FOR ATOMIC FORCE MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 2447-2450 (1992).

4. P. B. Fischer, M. S. Wei and S. Y. Chou, *ULTRAHIGH-RESOLUTION MAGNETIC FORCE MICROSCOPE TIP FABRICATED USING ELECTRON-BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2570-2573 (1993).

5. W. Gortz, B. Kempf and J. Kretz, *RESOLUTION ENHANCED SCANNING FORCE MICROSCOPY MEASUREMENTS FOR CHARACTERIZING DRY-ETCHING METHODS APPLIED TO TITANIUM MASKED INP*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 1, pp. 34-39 (1995).

6. S. Akamine, R. C. Barrett and C. F. Quate, *IMPROVED ATOMIC FORCE MICROSCOPE IMAGES USING MICROCANTILEVERS WITH SHARP TIPS*, Applied Physics Letters **Vol. 57,** No. 3, pp. 316-318 (1990).

7. T. R. Albrecht, S. Akamine, T. E. Carver and C. F. Quate, *MICROFABRICATION OF CANTILEVER STYLI FOR THE ATOMIC FORCE MICROSCOPE*, Journal Of Vacuum Science & Technology A-Vacuum Surfaces and Films **Vol. 8,** No. 4, pp. 3386-3396 (1990).

8. O. Wolter, T. Bayer and J. Greschner, *MICROMACHINED SILICON SENSORS FOR SCANNING FORCE MICROSCOPY*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 2, pp. 1353-1357 (1991).

9. K. D. Wise and K. Najafi, *MICROFABRICATION TECHNIQUES FOR INTEGRATED SENSORS AND MICROSYSTEMS*, Science **Vol. 254,** pp. 1335-1342 (1991).

10. J. Bryzek, K. Petersen and W. McCulley, *MICROMACHINES ON THE MARCH*, IEEE Spectrum No. May, pp. 20-33 (1994).

11. J. W. Gardner, *Chapter 3: CONVENTIONAL SILICON PROCESSING* in Microsensors Principles and Applications, published by John Wiley & Sons, Chichester, pp. 36-57 (1994).

12. J. W. Gardner, *Chapter 4: SPECIALISED MATERIALS PROCESSING* in Microsensors Principles and Applications, published by John Wiley & Sons, Chichester, pp. 59-78 (1994).

13. C. H. Mastrangelo and W. C. Tang, *Chapter 2: SEMICONDUCTOR SENSOR TECHNOLOGIES* in Semiconductor Sensors, edited by S. M. Sze, published by John Wiley & Sons, New York, pp. 17-95 (1994).

14. R. T. Howe, *SURFACE MICROMACHINING FOR MICROSENSORS AND MICROACTUATORS*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 1809-1813 (1988).

15. M. W. Borner, M. Kohl, F. J. Pantenburg, W. Bachler, H. Hein and W. K. Schomburg, *SUB-MICRON LIGA PROCESS FOR MOVABLE MICROSTRUCTURES*, Microelectronic Engineering **Vol. 30,** pp. 505-508 (1996).

16. K. E. Bean, *ANISOTROPIC ETCHING OF SILICON*, IEEE Transactions On Electron Devices **Vol. ED-25,** No. 10, pp. 1185-1193 (1978).

17. K. E. Petersen, *SILICON AS A MECHANICAL MATERIAL*, Proceedings of the IEEE **Vol. 70,** No. 5, pp. 421-457 (1982).

18. W. Kern, *CHEMICAL ETCHING OF SILICON, GERMANIUM, GALLIUM ARSENIDE, AND GALLIUM PHOSPHATE*, RCA Reviews **Vol. 39,** p. 278 (1978).

19. H. Robbins and B. Schwarz, *CHEMICAL ETCHING OF SILICON: I. THE SYSTEM HF, HNO_3, AND H_2O*, Journal of the Electrochemical Society **Vol. 106,** No. 6, pp. 505-508 (1959).

20. H. Robbins and B. Schwarz, *CHEMICAL ETCHING OF SILICON: II. THE SYSTEM HF, HNO_3, H_2O, and $HC_2H_3O_2$*, Journal of the Electrochemical Society **Vol. 107,** No. 2, pp. 108-111 (1960).

21. H. Muraoka, T. Ohhashi and Y. Sumitomo, *CONTROLLED PREFERENTIAL ETCHING TECHNOLOGY* in Semiconductor Silicon 1973, edited by H. R. Huff and R. R. Burgess, Princeton, p. 327 (1973).

22. A. F. Bogenschutz, W. Krusemark, K. H. Locherer and W. Mussinger, *ACTIVATION ENERGIES IN THE CHEMICAL ETCHING OF SEMICONDUCTORS IN HNO_3-HF-CH_3COOH*, Journal of the Electrochemical Society **Vol. 114,** No. 9, pp. 970-973 (1967).

23. M. J. Theunissen, J. A. Appels and W. H. C. Verkuylen, *APPLICATION OF ELECTROCHEMICAL ETCHING OF SILICON TO SEMICONDUCTOR DEVICE TECHNOLOGY*, Journal of the Electrochemical Society **Vol. 117,** pp. 959-965 (1970).

24. J. Brugger, R. A. Buser and N. F. Derooij, *SILICON CANTILEVERS AND TIPS FOR SCANNING FORCE MICROSCOPY*, Sensors and Actuators A-Physical **Vol. 34,** No. 1-3, pp. 193-200 (1992).

25. P. F. Indermuehle, C. Linder, J. Brugger, V. P. Jaecklin and N. F. Derooij, *DESIGN AND FABRICATION OF AN OVERHANGING XY-MICROACTUATOR WITH INTEGRATED TIP FOR SCANNING SURFACE PROFILING*, Sensors and Actuators A-Physical **Vol. 43,** No. 1-3, pp. 346-350 (1994).

26. R. M. Finne and D. L. Klein, *A WATER-AMINE-COMPLEXING AGENT SYSTEM FOR ETCHING SILICON*, Journal of the Electrochemical Society **Vol. 114,** No. 9, pp. 965-970 (1967).

27. A. Reisman, M. Berkenblit, S. A. Chan, F. B. Kaufman and D. C. Green, *THE CONTROLLED ETCHING OF SILICON IN CATALYZED ETHYLENEDIAMINE-PYROCATECHOL-WATER SOLUTIONS*, Journal of the Electrochemical Society **Vol. 126,** pp. 1406-1415 (1979).

28. M. J. Declercq, L. Gerzberg and J. D. Meindl, *OPTIMIZATION OF THE HYDRAZINE-WATER SOLUTION FOR ANISOTROPIC ETCHING OF SILICON IN INTEGRATED CIRCUIT TECHNOLOGY*, Journal of the Electrochemical Society **Vol. 122,** No. 4, pp. 545-552 (1975).

29. M. Asano, T. Cho and H. Muraoka, *APPLICATION OF CHOLINE IN SEMICONDUCTOR TECHNOLOGY*, Electrochemical Society Extended Abstracts No. 354, pp. 911-913 (1976).

30. M. Mehregany and S. D. Senturia, *ANISOTROPIC ETCHING OF SILICON IN HYDRAZINE*, Sensors and Actuators **Vol. 13,** pp. 375-390 (1988).

31. U. Schnakenberg, W. Benecke and B. Lochel, *NH_4OH-BASED ETCHANTS FOR SILICON MICROMACHINING*, Sensors and Actuators A-Physical **Vol. 21-23,** pp. 1031-1035 (1990).

32. U. Schnakenberg, W. Benecke and P. Lange, *TMAHW ETCHANTS FOR SILICON MICROMACHINING*. presented at Transducers '91 - 6th International Conference on Solid-State Sensors and Actuators, San Francisco, California, USA (1991).

33. U. Schnakenberg, W. Benecke, B. Lochel, S. Ullerich and P. Lange, *NH_4OH-BASED ETCHANTS FOR SILICON MICROMACHINING: INFLUENCE OF ADDITIVES AND STABILITY OF PASSIVATION LAYERS*, Sensors and Actuators A-Physical **Vol. 24,** pp. 1-7 (1991).

34. O. Tabata, R. Asahi, H. Funabashi, S. K. and S. Sugiyama, *ANISOTROPIC ETCHING OF SILICON IN TMAH SOLUTIONS*, Sensors and Actuators A-Physical **Vol. 34,** pp. 51-57 (1992).

35. M. A. Gajda, H. Ahmed, J. E. A. Shaw and A. Putnis, *ANISOTROPIC ETCHING OF SILICON IN HYDRAZINE*, Sensors and Actuators A-Physical **Vol. 40,** pp. 227-236 (1994).

36. I. J. Pugacz-Muraszkiewicz, *DETECTION OF DISCONTINUITIES IN PASSIVATING LAYERS ON SILICON BY NAOH ANISOTROPIC ETCH*, IBM Journal of Research and Development **Vol. 16,** p. 523 (1972).

37. J. B. Price, *ANISOTROPIC ETCHING OF SILICON WITH POTASSIUM HYDROXIDE-WATER-ISOPROPYL ALCHOHOL* in Semiconductor Silicon 1973, edited by H. R. Huff and R. R. Burgess, Princeton, p. 339 (1973).

38. D. L. Kendall, *ON ETCHING VERY NARROW GROOVES IN SILICON*, Applied Physics Letters **Vol. 26,** No. 4, pp. 195-198 (1975).

39. D. B. Lee, *ANISOTROPIC ETCHING OF SILICON*, Journal Of Applied Physics **Vol. 40,** No. 11, pp. 4569-4574 (1969).

40. E. Bassous, *FABRICATION OF NOVEL THREE-DIMENSIONAL MICROSTRUCTURES BY THE ANISOTROPIC ETCHING OF (100) AND (110) SILICON*, IEEE Transactions On Electron Devices **Vol. ED-25,** No. 10, pp. 1178-1184 (1978).

41. E. Bassous and E. F. Baran, *THE FABRICATION OF HIGH PRECISION NOZZLES BY THE ANISOTROPIC ETCHING OF (100) SILICON*, Journal of the Electrochemical Society **Vol. 125,** No. 8, pp. 1321-1327 (1978).

42. P. J. Hesketh, C. Ju, S. Gowda, E. Zanoria and S. Danyluk, *SURFACE FREE ENERGY MODEL OF SILICON ANISOTROPIC ETCHING*, Journal of the Electrochemical Society **Vol. 140,** No. 4, pp. 1080-1084 (1993).

43. B. H. Koek, *'MIX AND MATCH' LITHOGRAPHY*, published by Philips E-beam Lithography Department, Eindhoven, The Netherlands (1980-1993).

44. T. H. P. Chang, *PROXIMITY EFFECTS IN ELECTRON-BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology **Vol. 12,** p. 1271 (1975).

45. T. H. P. Chang, D. P. Kern, E. Kratschmer, K. Y. Lee, H. E. Luhn, M. A. McCord, S. A. Rishton and Y. Vladimirsky, *NANOSTRUCTURE TECHNOLOGY*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 462-493 (1988).

46. A. N. Broers, *LONG LIFE LAB_6 CATHODES*, Journal Of Applied Physics **Vol. 38,** p. 1991 (1967).

47. F. J. Hohn, T. H. P. Chang, A. N. Broers, G. S. Frankel, E. T. Peters and D. W. Lee, *FABRICATION AND TESTING OF SINGLE-CRYSTAL LANTHANUM HEXABORIDE ROD CATHODES*, Journal Of Applied Physics **Vol. 53,** No. 3, pp. 1283-1296 (1982).

48. T. R. Sandin, *Chapter 10: THE WAVE NATURE OF PARTICLES* in Essentials of Modern Physics, published by Addison-Wesley, North Carolina, pp. 109-121 (1989).

49. A. N. Broers, *RESOLUTION, OVERLAY AND FIELD SIZE FOR LITHOGRAPHY SYSTEMS*, IEEE Transactions On Electron Devices **Vol. ED-28,** No. 11, pp. 1268-1278 (1981).

50. R. K. Watts, *Chapter 4: LITHOGRAPHY* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 141-183 (1988).

51. Philips, *THE PHILIPS BEAMWRITER*, published by Philips E-beam Lithography Division, Eindhoven, The Netherlands (1980-1993).

52. S. A. Rishton, S. P. Beaumont and C. D. W. Wilkinson, *MEASUREMENT OF THE PROFILE OF FINELY FOCUSED ELECTRON BEAMS IN A SCANNING ELECTRON MICROSCOPE*, Journal of Physics E: Scientific Instrumentation **Vol. 17,** p. 296 (1984).

53. T. Chisholm, *SPOT-SIZE MEASUREMENT IN AN ELECTRON-BEAM PATTERN GENERATOR*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2066-2069 (1988).

54. S. Thoms, *Private Communication*, (1993-1996).

55. S. Thoms, *NOTES ON EBL*, published by Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1993-1996).

56. S. Thoms, *FABRICATORS GUIDE*, published by Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1993-1996).

57. R. L. Nicol, *Private Communication*, (1993-1996).

58. W. J. Duffin, *Chapter 7: THE LAWS OF MAGNETIC INTERACTION (IN VACUO)* in Electricity and Magnetism, 4th edition, published by McGraw-Hill Book Company, Maidenhead, pp. 162-193 (1990).

59. S. Thoms, *REGISTRATION ERRORS*, published by Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1993-1996).

60. C. D. W. Wilkinson and S. P. Beaumont, *ELECTRON BEAM NANOLITHOGRAPHY*, published by Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1980-1993).

61. R. F. W. Pease, *ELECTRON BEAM LITHOGRAPHY*, Contemporary Physics **Vol. 22,** No. 3, pp. 265-290 (1981).

62. P. D. Maker and R. E. Muller, *PHASE HOLOGRAMS IN POLYMETHYL METHACRYLATE*, Journal Of Vacuum Science & Technology **Vol. 10,** No. 6, pp. 2516-2519 (1992).

63. I. Haller, M. Hatzakis and R. Srinivasan, *HIGH-RESOLUTION POSITIVE RESIST FOR ELECTRON-BEAM EXPOSURE*, IBM Journal of Research and Development **Vol. 12,** pp. 251-256 (1968).

64. M. Hatzakis, *ELECTRON RESISTS FOR MICROCIRCUIT AND MASK PRODUCTION*, Journal of the Electrochemical Society **Vol. 116,** pp. 1033-1037 (1969).

65. W. J. Daughton and F. L. Givens, *AN INVESTIGATION OF THE THICKNESS VARIATION OF SPUN-ON THIN FILMS*, Journal of the Electrochemical Society **Vol. 129,** pp. 173-179 (1982).

66. S. W. J. Kuan and C. W. Frank, *ULTRATHIN POLYMER FILMS FOR MICROLITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2274-2279 (1988).

67. S. P. Beaumont, T. Tamamura and C. D. W. Wilkinson, *A TWO LAYER RESIST SYSTEM FOR EFFICIENT LIFTOFF IN VERY HIGH RESOLUTION ELECTRON BEAM LITHOGRAPHY*, Proceedings of Microcircuit Engineering 80, Netherlands, Delft University Press (1981).

68. J. S. Greeneich, *DEVELOPER CHARACTERISTICS OF POLY(METHYL METHACRYLATE) ELECTRON RESIST*, Journal of the Electrochemical Society **Vol. 122,** pp. 970-976 (1975).

69. M. Hatzakis, *MATERIALS AND PROCESSES FOR MICROSTRUCTURE FABRICATION*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 441-453 (1988).

70. M. A. McCord, A. Wagner and T. Donohue, *RESOLUTION LIMITS AND PROCESS LATITUDE OF X-RAY MASK FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2958-2963 (1993).

71. W. Bolton, *Chapter 16: SOLIDS, LIQUIDS AND GASES* in Patterns in Physics, 2nd edition, published by McGraw-Hill Book Company, Maidenhead, pp. 325-366 (1986).

72. S. P. Beaumont, P. G. Bower, T. Tamamura and C. D. W. Wilkinson, *SUB-20-NM-WIDE METAL LINES BY ELECTRON-BEAM EXPOSURE OF THIN POLY(METHYLMETHACRYLATE) FILMS AND LIFTOFF*, Applied Physics Letters **Vol. 38,** No. 6, pp. 436-439 (1981).

73. A. N. Broers, J. M. E. Harper and W. W. Molzen, *250 Å LINEWIDTHS WITH PMMA ELECTRON RESIST*, Applied Physics Letters **Vol. 33,** p. 392 (1978).

74. M. Parikh and D. Kyser, *ENERGY DEPOSITION FUNCTIONS IN ELECTRON RESIST FILMS ON SUBSTRATES*, Journal Of Applied Physics **Vol. 50,** p. 1104 (1979).

75. A. N. Broers, *RESOLUTION LIMITS OF PMMA RESIST FOR EXPOSURE WITH 50 KV ELECTRONS*, Journal of the Electrochemical Society **Vol. 128,** No. 1, pp. 166-170 (1981).

76. W. Chen and H. Ahmed, *FABRICATION OF 5-7 NM WIDE ETCHED LINES IN SILICON USING 100 KEV ELECTRON-BEAM LITHOGRAPHY AND POLYMETHYLMETHACRYLATE RESIST*, Applied Physics Letters **Vol. 62,** No. 13, pp. 1499-1501 (1993).

77. W. Chen and H. Ahmed, *FABRICATION OF SUB-10 NM STRUCTURES BY LIFT-OFF AND BY ETCHING AFTER ELECTRON-BEAM EXPOSURE OF POLY(METHYLMETHACRYLATE) RESIST ON SOLID SUBSTRATES*, Journal Of Vacuum Science & Technology **Vol. 11,** No. 6, pp. 2519-2523 (1993).

78. D. R. S. Cumming, S. Thoms, S. P. Beaumont and J. M. R. Weaver, *FABRICATION OF 3 NM WIRES USING 100 KEV ELECTRON BEAM LITHOGRAPHY AND POLY(METHYLMETHACRYLATE) RESIST*, Applied Physics Letters **Vol. 68,** No. 3, pp. 322-324 (1996).

79. R. E. Howard, *MULTILEVEL RESIST LITHOGRAPHY BELOW 100 NM*, IEEE Transactions On Electron Devices **Vol. ED-28,** No. 11, pp. 1378-1381 (1981).

80. S. K. Murad, *Private Communication*, (1993-1996).

81. N. Cameron, *Private Communication*, (1993-1996).

82. W. Liu, J. Ingino and R. F. Pease, *RESIST CHARGING IN ELECTRON BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 5, pp. 1979-1983 (1995).

83. I. Haller, R. Feder, M. Hatzakis and E. Spiller, *COPOLYMERS OF METHYL METHACRYLATE AND METHACRYLIC ACID AND THEIR METAL SALTS AS RADIATION SENSITIVE RESISTS*, Journal of the Electrochemical Society **Vol. 126,** No. 1, pp. 154-161 (1979).

84. M. Hatzakis, *PMMA COPOLYMERS AS HIGH SENSITIVITY ELECTRON RESISTS*, Journal Of Vacuum Science & Technology **Vol. 16,** No. 6, pp. 1984-1988 (1979).

85. W. Patrick, W. S. Mackie, S. P. Beaumont and C. D. W. Wilkinson, *POLYIMIDE AS A NEGATIVE RESIST AND ITS APPLICATION IN CROSSOVERS AND METAL ON POLYMER MASK FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 4,** No. 1, pp. 390-393 (1985).

86. D. MacIntyre and S. Thoms, *Private Communication*, (1995-1996).

87. K. G. Chiong, S. Wind and D. Seeger, *EXPOSURE CHARACTERISTICS OF HIGH-RESOLUTION NEGATIVE RESISTS*, Journal Of Vacuum Science & Technology B **Vol. 8,** No. 6, pp. 1447-1453 (1990).

88. S. Thoms, I. McIntyre, S. P. Beaumont, M. Al-Mudares, R. Cheung and C. D. W. Wilkinson, *FABRICATION OF QUANTUM WIRES IN GAAS/ALGAAS HETEROLAYERS*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 1, pp. 127-130 (1988).

89. K. Early, D. M. Tennant, D. Y. Jeon, P. P. Mulgrew, A. A. MacDowell and O. R. Wood II, *CHARACTERIZATION OF AZ PN114 RESIST FOR HIGH RESOLUTION USING ELECTRON-BEAM AND SOFT-X-RAY PROJECTION LITHOGRAPHIES*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 2600-2605 (1992).

90. D. MacIntyre and S. Thoms, *HIGH RESOLUTION STUDIES ON HOECHST AZ PN114 CHEMICALLY AMPLIFIED RESIST*, Microelectronic Engineering **Vol. 30,** pp. 327-330 (1996).

91. Shipley, *Technical Manual*, published by Shipley Photo Resist (1980-1993).

92. J. Pacansky and J. R. Lyerla, *PHOTOCHEMICAL DECOMPOSITION MECHANISMS FOR AZ-TYPE PHOTORESISTS*, IBM Journal of Research and Development **Vol. 23,** No. 1, pp. 42-55 (1979).

93. G. G. Collins and C. W. Halsted, *PROCESS CONTROL OF THE CHLOROBENZENE SINGLE-STEP LIFT-OFF PROCESS WITH A DIAZO-TYPE RESIST*, IBM Journal of Research and Development **Vol. 26,** No. 5, pp. 596-604 (1982).

94. B. Casey, *Private Communication*, (1995-1996).

95. S. Rahman, *ELECTRON-BEAM LITHOGRAPHY ON NON-FLAT SURFACES*, Final Year Project, Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1992-1993).

96. L. D. Clark, J. L. Lund and D. J. Edell, *CESIUM HYDROXIDE (CSOH): A USEFUL ETCHANT FOR MICROMACHINING SILICON*. IEEE Solid-State Sensor and Actuator Workshop, Hilton Head (1988).

97. J. M. R. Weaver, *Private Communication*, (1993-1996).

98. L. Pollock, *Private Communication*, (1993-1996).

99. T. Lyman, *METALS HANDBOOK*, published by American Society for Metals, Novelty, Ohio (1961).

100. A. W. Griffith, *Private Communication*, (1993-1996).

101. R. Harkins, *Private Communication*, (1993-1996).

102. A. Soderbarg and L. Smith, *ELECTROCHEMICAL ETCH STOP OBTAINED BY ACCUMULATION OF FREE CARRIERS WITHOUT P-N JUNCTION*, Journal of the Electrochemical Society **Vol. 140,** No. 1, pp. 271-275 (1993).

103. C. H. Ozdemir and J. G. Smith, *NEW PHENOMENA OBSERVED IN ELECTROCHEMICAL MICROMACHINING OF SILICON*, Sensors and Actuators A-Physical **Vol. 34,** pp. 87-93 (1992).

104. A. Franciosi, D. J. Peterman and J. H. Weaver, *SILICON-REFRACTORY METAL INTERFACES: EVIDENCE OF ROOM-TEMPERATURE INTERMIXING FOR SI-CR*, Journal Of Vacuum Science & Technology **Vol. 19,** No. 3, pp. 657-660 (1981).

105. P. J. Grunthaner, F. J. Grunthaner, A. Madhuker and J. W. Mayer, *METAL/SILICON INTERFACE FORMATION: THE NI/SI AND PD/SI SYSTEMS*, Journal Of Vacuum Science & Technology **Vol. 19,** No. 3, pp. 649-656 (1981).

106. S. M. Sze, *Chapter 9: OXIDATION AND FILM DEPOSITION* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 341-380 (1985).

107. M. Lambrechts and W. Sansen, *Chapter 4: PLANAR TECHNOLOGIES FOR MICROELECTROCHEMICAL SENSORS* in Biosensors: Microelectrochemical Devices, published by Institute of Physics, Leuven, p. 98 (1990-1993).

CHAPTER 4

SUBSTRATE ENGINEERING

This chapter is concerned with the development of all the lithographic levels associated with the fabrication of a substrate suitable for the definition of nanosensors by high resolution electron-beam lithography. This includes all the photolithography levels and the first two electron-beam lithography levels. The process and materials used for micromachining through the wafer, the method used for fabricating the tips in a reliable fashion and the method used for patterning cantilevers suitable for both contact and non-contact mode AFM are all explained in some detail. The solutions to problems which occurred during process development, such as the effects of residual stress and capillary forces, are also covered, as are some of the design features incorporated into the process to aid wafer handling and improve ease of use.

4.1 POTASSIUM HYDROXIDE ETCHING

As has been noted in the previous chapter, the use of caesium hydroxide for etching silicon with a silicon oxide mask had not been very successful in the context of the intended fabrication scheme. It is probable that it could have been made to work given more time and effort but the advantages in cost to moving to a potassium hydroxide alternative, combined with the more widespread use of this etch meant that a change to this scheme was preferable. However, another alternative that is widely used is the ethylene diamine, pyracatechol and water (EDP) etching system [see §3.4.2], which also etches silicon oxide at a lower rate than potassium hydroxide [Bassous and Baran[1] (1978)], but this was rejected on safety grounds.

4.1.1 Cost Considerations

A source of 3 inch silicon wafers, *n*-type, one side polished and about 380μm thick which could come precoated with LPCVD silicon nitride at very little extra cost was identified. Since many of the departmental facilities can only handle a maximum wafer size of three inches, it was decided that the ultimate scaling of the fabrication process would be to this size of wafer, at least for the duration of this project. These wafers have the advantage of being thinner, so that the etching time is reduced in comparison with 4 inch wafers at the cost of fewer devices per wafer. In-house thermal oxidation is incapable of producing a good quality uniform silicon oxide film on wafers as large as this and the cost of obtaining 3 inch wafers thermally oxidised to a thickness of 2μm was greater than that of obtaining the LPCVD silicon nitride coated wafers. Quotes for the former were in the range of £50 to £100 + VAT per wafer (Compart Technology), while for the latter it was about £20 + VAT per wafer (University of Edinburgh). In addition, the cost of 250 millilitres of 50% by weight caesium hydroxide solution was £120 + VAT (item 23,206-8 from Aldrich) while potassium hydroxide pellets cost about £40 + VAT for 5kg (item 10210 7B from BDH), enough to make more than 12 litres of 7 molar solution.

4.1.2 Comparison of Silicon Wet Etching Schemes

To investigate the new scheme some <100> oriented, *n*-type, 2Ωcm to 4Ωcm, 3 inch diameter, one side polished, silicon wafers, 381±25μm thick were bought with a 50nm LPCVD silicon nitride coating on both sides. One piece was placed in a 7 molar potassium hydroxide solution at 80°C for ten hours. The wafer was then examined in the Hitachi S900 SEM and the thickness of the silicon nitride at the edge of a pinhole was measured to be not less than 30nm as shown in Figure 4.1.1. As both sides of the pinhole are assumed to have been exposed to the etch this meant that the mask erosion rate was less than 1nm/hour per exposed side. The resistance of LPCVD silicon nitride to hot concentrated potassium hydroxide solution is well known [*e.g.* Lambrechts and

Sansen[2] (1990-1993); Mastrangelo and Tang[3] (1994)]. Since the silicon etch rate under these conditions is measured to be 76μm/hour, this means that over 3800μm of silicon could be etched with this masking layer, a drastic improvement over the previous scheme (*i.e.* 4nm of silicon nitride would be sufficient to etch once through the whole wafer, neglecting pinholes). An additional feature of the LPCVD silicon nitride is that it etches at less than 2nm/minute in silicon dioxide etch 4:1, compared to over 150nm/minute for PECVD silicon nitride and PECVD silicon oxide, and about 100nm/minute for thermally grown silicon oxide (from experiment). The implications of these factors are clear:

(a) greater fabrication process stability due to negligible dependence on mask thickness and uniformity
(b) ability to use standard thickness wafers (more cost effective) and easily scaleable to larger and hence thicker wafers
(c) ability to use a more common and less expensive etching solution
(d) negligible mask damage from carrying out a deoxidation step in silicon dioxide etch prior to the silicon etch.

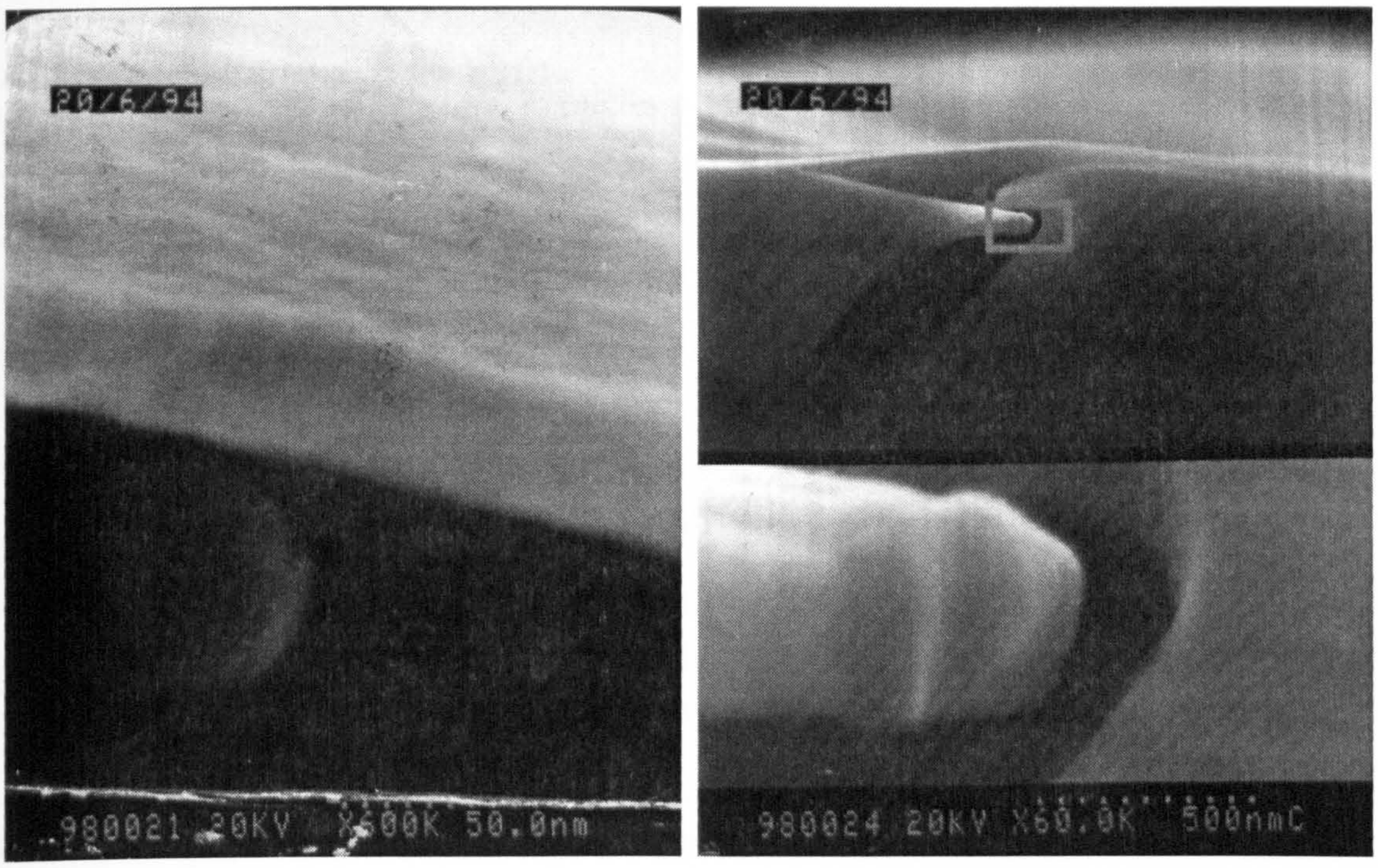


Figure 4.1.1: Micrograph illustrating the lack of etching of LPCVD silicon nitride in hot concentrated potassium hydroxide solution. See text for details.

A 7 molar concentration was selected for a number of reasons including its apparent popularity in the literature and observations that concentrations much below 30% by weight sometimes resulted in a rough finish at the bottom of {100} pits [Mastrangelo, *et al.*[3] (1994)]. Since the potassium hydroxide pellets are specified by the manufacturer to contain a maximum of 5% potassium carbonate (caused by absorbing carbon dioxide from the atmosphere), the actual solution used is 7 molar plus an additional 5% giving a 7.35 molar solution at most. This corresponds to a maximum 29.2% by weight concentration.

4.1.3 Practical Silicon Wet Etching

Based on this etch, an etching kit was designed that would have the required thermal stability to achieve an etch depth tolerance of 1μm at a reasonable rate of etching. The details of the design are given in Appendices A, B and C. Some practical methods were also established to achieve reproducible etching results, including methods for making up and storing the etching solution, preparing the specimen for etching (deoxidation), achieving a uniform etch rate (stirring) and cleaning up the specimen after etching (neutralisation). These methods are all detailed in Appendix A (§A.2).

4.2 THE PHOTOLITHOGRAPHY LEVELS

4.2.1 Processing Fundamentals

Due to the LPCVD silicon nitride layer being so thin, several precautions have been found necessary in the handling of wafers. Firstly, for handling one inch pieces, only plastic tweezers (Farnell plastic conductive tweezers, catalogue number 175-876) are used and the wafer is lifted by gripping the opposite edges of the wafer simultaneously. This minimises contact with the silicon nitride coated surface. The widely spaced teeth of these tweezers are very useful for this purpose. However, since these tweezers are not resistant to MIBK, nanostrip (a concentrated mixture of sulphuric acid and hydrogen peroxide) and concentrated sulphuric acid, a pair of polyethylene tweezers is normally used for steps involving these chemicals. These tweezers are more awkward to use due to their lack of rigidity. For handling 3 inch wafers, polyethylene wafer dippers (Fluoroware PFA Dipper, catalogue number D11-0215) are found to be ideal for all the chemicals used, as they hold the wafer securely by its edges using 'v-grooves'.

Before most processing levels, it is advisable to clean the substrate. Several cleaning processes are possible. However the cleaning processes which are used depend on the context in which they are being applied. The processes generally used for the work in this thesis are described in Appendix D.

Changing the mask from thermally grown silicon oxide to LPCVD silicon nitride meant that a change in mask patterning technique was required. This is because hydrofluoric acid etches LPCVD silicon nitride at less than 2nm/minute so that 25 minutes would be required to etch 50nm. Unfortunately, photoresist is an inadequate masking layer when more than 10 minutes of etching in hydrofluoric acid is required [Pollock[4] (1993-1996)]. After this time, the hydrofluoric acid undercuts the resist, possibly due to the loss of adhesion between resist and substrate.

The usual technique for patterning silicon nitride is dry etching. However, for initial development using one inch substrates, it was found that the development cycle time could be reduced by using an alternative scheme based on wet etching. This scheme was

modified from that used by one of the senior technicians several years previously for the fabrication of silicon nitride membranes [Gourlay[5] (1993-1996)]. The etch used here is a mixture of orthophosphoric acid and water in a particular concentration that boils at 146°C [Vangelder and Hauser[6] (1967)]. This scheme makes use of hardbaked Shipley S1818 photoresist (at 180°C) as a mask for the etching process and is described in Appendix E (§E.1). It should be noted that the manufacturer, Shipley, does not recommend that this resist be baked at temperatures exceeding 150°C [Shipley[7] (1980-1993)]. However, no special problems were encountered using this process.

This process worked well for LPCVD silicon nitride layers 50nm thick. However, it was found that a 50nm thick layer was prone to being scratched during wafer handling in subsequent processing steps, so that a change was eventually made to 100nm thick LPCVD silicon nitride coated wafers (the cost of the wafer remained unchanged). With these wafers, an etching time of 40 minutes was required using the same process but this was found to be beyond the adhesion capabilities of the photoresist. This, combined with a move to full 3 inch wafer processing, meant that dry etching became the preferred mask patterning option. The resulting scheme is described in Appendix E (§E.2).

It was found that scratches in the silicon nitride on the polished side of the wafer was still occasionally a problem, probably due to poor wafer handling during processing. As noted in §3.7.3, nichrome (60% nickel, 40% chromium) is resistant to the potassium hydroxide etch. Therefore, as an alternative to the protection of the polished side of the wafer with photoresist [steps (a) and (b) in Appendix E, §E.2] for the duration of the silicon nitride etch (the resist must be stripped off prior to wet etching), wafers were protected by evaporating a 100nm thick layer of nichrome onto the polished side of the wafer prior to any processing. This layer could then remain on the substrate until the first electron-beam lithography level. This was found to improve the situation. It has been found that the nichrome protection layer survives the silicon deoxidation step prior to silicon wet etching. It also survives the post silicon etch neutralisation step if dilute sulphuric acid is used rather than dilute hydrochloric acid.

An alternative would be to use an even thicker silicon nitride layer (200nm costs the same), however, removal of this layer after the tip definition etch then becomes more difficult, although an orthophosphoric acid etching step could be introduced.

4.2.2 The Design Concept

As mentioned in §3.7.2 and in particular, Figure 3.7.2, the first and second photolithographic levels would involve the steps shown in Figure 4.2.1. Steps 1 to 3 can be regarded as pre-processing steps involving the deposition of LPCVD silicon nitride and the vacuum deposition of a nichrome (60% nickel, 40% chromium) layer to protect the polished side of the wafer. Steps 4 and 5 constitute the first photolithographic level, while steps 6 and 7 constitute the second photolithographic level.

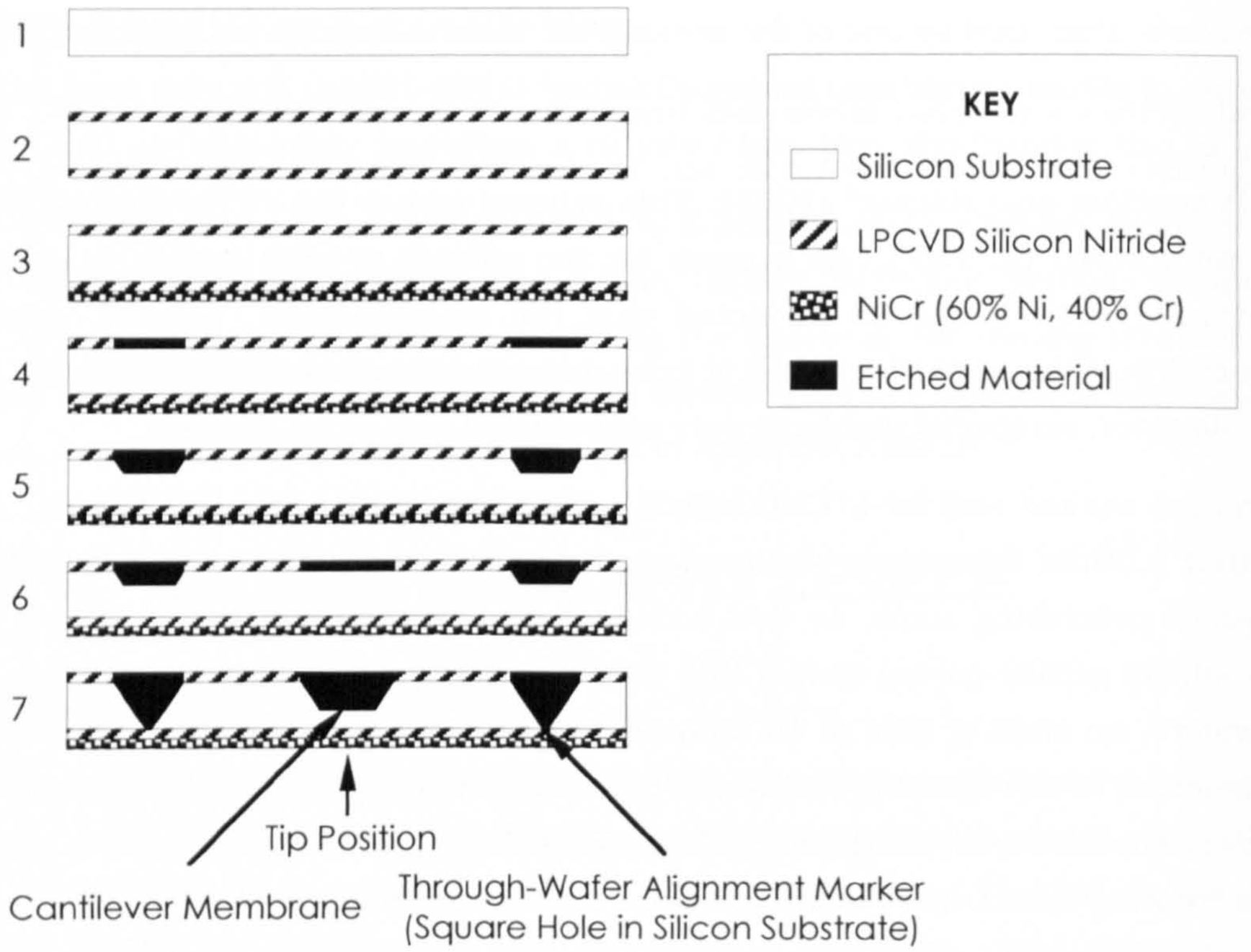


Figure 4.2.1: Schematic of the steps involved in the photolithographic levels. Compare this with Figure 3.7.2.

The purpose of the first photolithographic level is to open up windows in the backface (unpolished side) silicon nitride for the through-wafer alignment markers. The silicon is then wet etched to a depth of about 100μm to give the markers a 'head start' through the wafer. The form these markers will eventually take is a square opening in the silicon on the polished side of the wafer, bounded by the {111} planes. Thus the marker size is determined by both the thickness of the wafer and the initial opening width [see Figure 3.7.3 for the relationship between these parameters].

The second photolithographic level then opens up larger windows to form the cantilever membrane. The second silicon wet etch is carried out to complete the through-wafer alignment markers, leaving the cantilever membrane with a thickness close to that of the depth of the first silicon wet etch.

After this step, electron-beam lithography is used to pattern the frontface (polished side) silicon nitride to define new alignment markers and the masking layer for the silicon wet etching of the tip *simultaneously*, so that subsequent electron-beam lithography levels can be accurately aligned to the tip. This level is carried out by coarsely aligning to the through-wafer alignment markers and places the tip on the cantilever membrane area of the substrate. This level is described in detail in §4.3.

The cantilever membrane becomes thinner during the silicon wet etch used to define the pyramidal tip, by double the height of the pyramidal tips (etched from both sides). However, the membrane is not completely removed until the cantilevers are 'released' during the release etch described in §4.4. This is shown schematically in Figure 4.2.2.

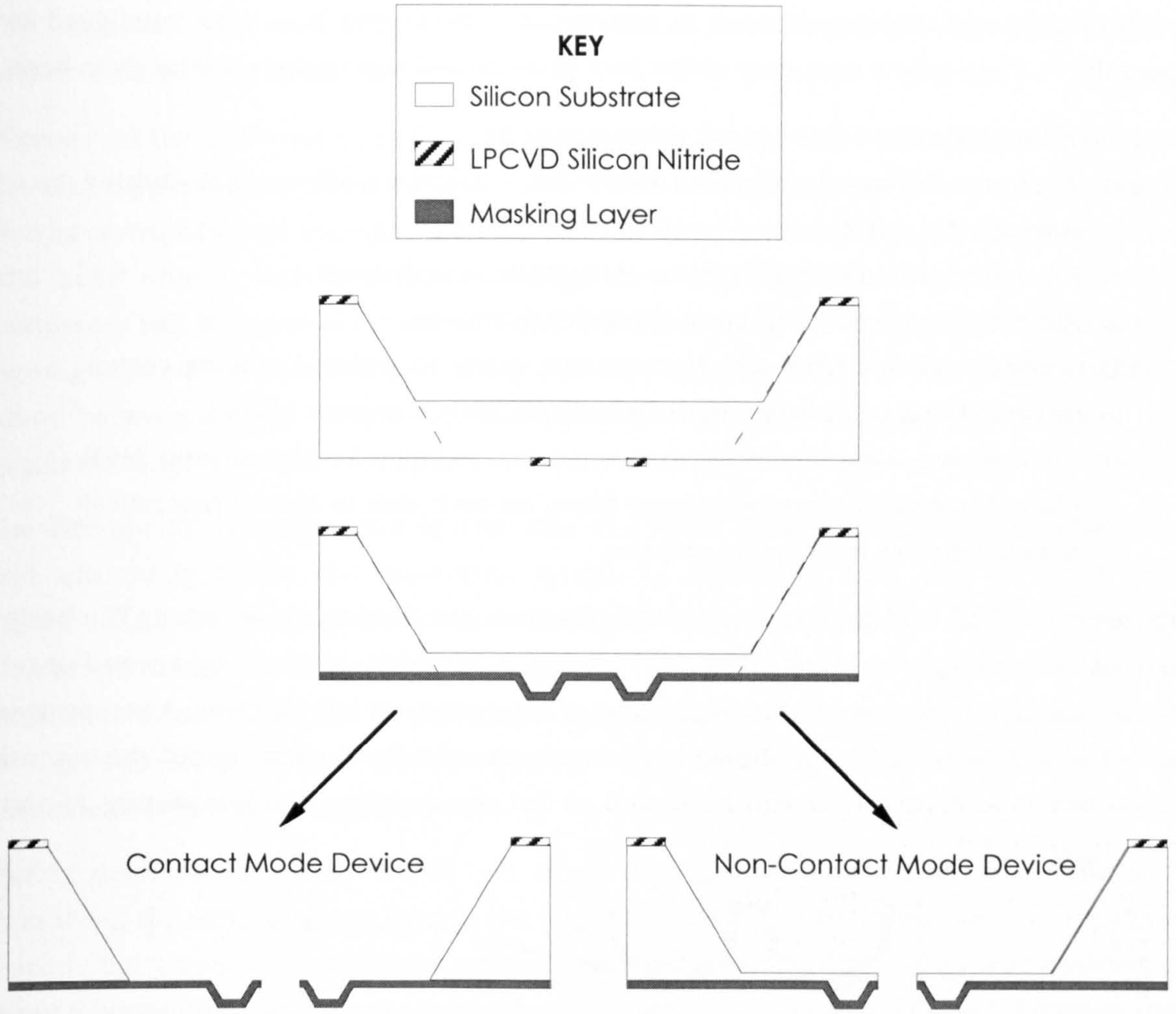


Figure 4.2.2: Schematic of the steps following the photolithographic levels. Compare this with Figure 3.7.2.

In the case of the contact mode devices, the masking material deposited after the tip definition etch becomes the cantilever material. This consists of PECVD silicon nitride, as discussed in §4.4. The release etch frees the cantilever from the substrate. As these cantilevers are extremely compliant [see §1.3], any sensors on contact mode cantilevers must be defined prior to the release etch.

In the case of non-contact mode devices the cantilevers are stiffer and consist of a part of the silicon substrate. Thus the masking material is used as a mask for etching through the cantilever membrane. This means that the thickness of this membrane must be adjusted so as to result in the required thickness of cantilever. As these cantilevers are much more rigid, sensors can be defined on these cantilevers after the release etch.

4.2.3 Mask Set Design Summary

The mask sets evolved through several iterations during the coarse of development. Initially, development was carried out on 25mm square silicon substrates, but eventually the developed process was modified to accommodate whole 3 inch wafers. Whilst the design of the mask set constituted a major part of the work carried out by the author, the

detail of this evolutionary process is somewhat tedious and has been consigned to Appendix F. Thus only a summary of the final process and main points will be given here.

The general design rules for the mask sets seem to be:

(a) Allow for 10μm of growth of every feature during silicon wet etching through the wafer.

(b) Allow for a ±5μm variation in the etch depth at each stage.

(c) Allow for a ±10μm variation in the wafer thickness.

(d) Separate all features which must not meet (to prevent convex corner undercutting) by at least 50μm, but ideally 100μm or even 150μm.

(e) All features from the first level must be widened in the second level photolithography stage by at least 50μm on each side to ensure theoretical dimensions are achieved in practice.

Figures F.20, F.21 and F.22 in Appendix F illustrate the final mask set used. The basic idea of this design is illustrated in Figure 4.2.3 which shows the second level photolithographic pattern used to define part of a substrate of AFM devices. A perspective view of four devices has been shown to illustrate the results of carrying out the second silicon wet etch. Figure 4.2.2 can be viewed as the cross-section of a line joining 'A' and 'B' in Figure 4.2.3.

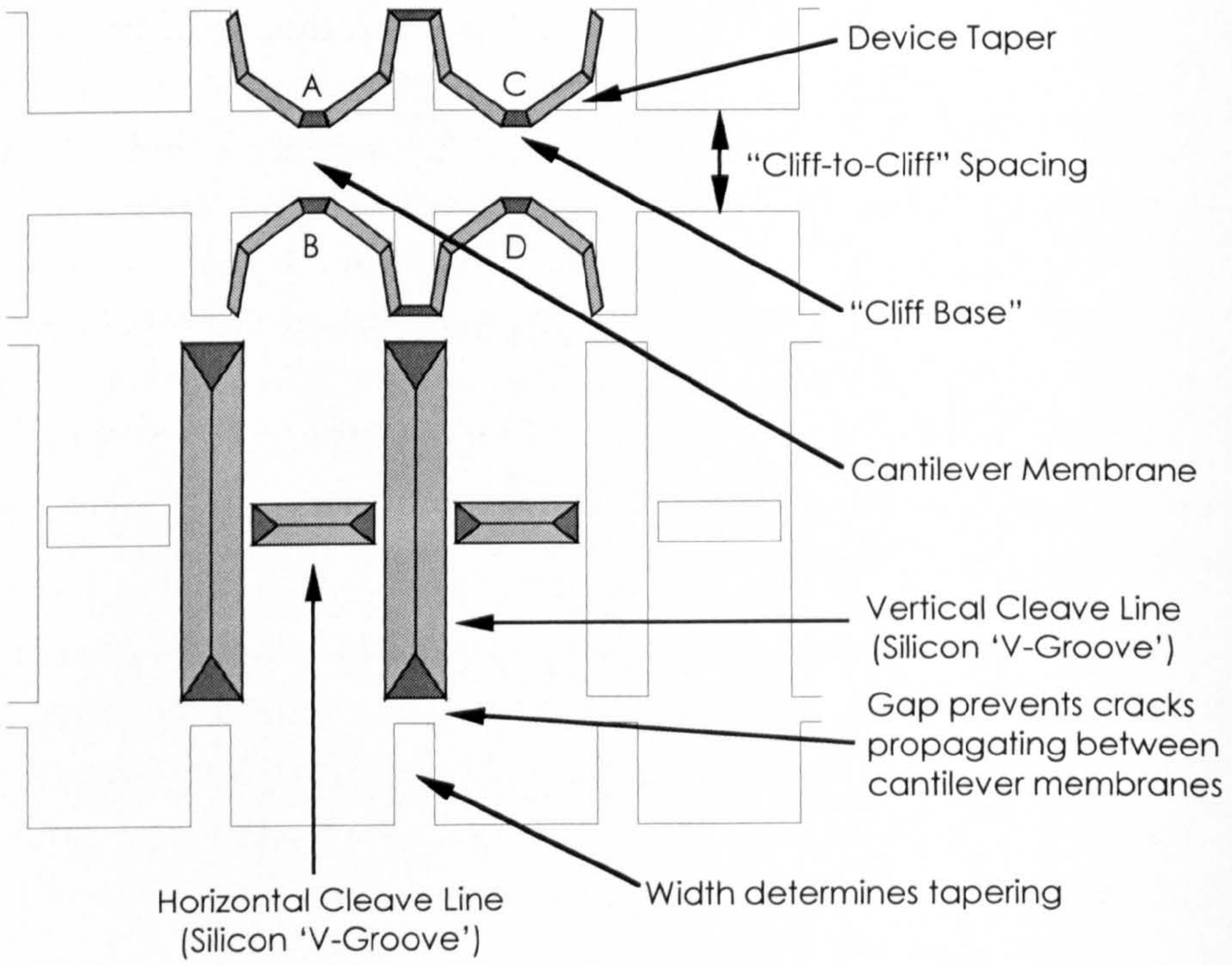


Figure 4.2.3: Schematic illustrating the eventual design used to define the AFM substrate using photolithography and silicon wet etching. A perspective view of four devices, labelled A, B, C and D is shown to illustrate the results of the second silicon wet etch.

As can be seen from the figure, {111} plane 'v-grooves' have been used at three edges of each device to aid in the cleaving of the devices. These 'v-grooves' etch terminate at a depth determined by the opening width. The widths were optimised by trial and error.

The horizontal and vertical groove patterns must remain separate to prevent convex corner undercutting, as discussed in §3.4.3.

However, at the 'cliff base' end of each device, convex corner undercutting is used to force the silicon substrate to taper towards a point. If the lateral distance between the exposed convex corners is wide enough, the etch will not quite eradicate the {111} planes at the 'cliff base' after it has penetrated through the wafer. It is from this base that the cantilevers will be made to extend into the 'cantilever membrane' area [see Figure 4.2.2, viewing this as the cross-section along a line passing through the centres of the 'cliff bases' between devices 'A' and 'B']. The taper thus serves to improve the access of the cantilever to the specimen. An optimum device width was found by iteration.

The 'cliff-to-cliff' spacing serves to determine the width of the 'cantilever membrane' area and ultimately limits the maximum length of cantilever that can be used [see Figure 4.2.2]. By making two devices face each other and share a single 'cantilever membrane' area, the number of sloping {111} planes is halved, thus increasing the number of devices that can be fabricated on a single wafer. This is at the expense of a weaker substrate as the 'cantilever membrane' area has to be twice as wide to accommodate two opposing cantilevers.

During development, it was found that stress during processing could cause cracks to form along the vertical cleave lines at the edge of a device and that these would propagate towards the 'cantilever membrane' area and then across the centre of this area destroying many devices (and sometimes the entire substrate). The presence of a gap between the vertical cleave lines and the 'cantilever membrane' area serves as both a crack arrester and as a way of allowing the width of the cleave line to be chosen independently of considerations regarding the tapering of the device and the resulting 'cliff base' width.

It was also found that in order to achieve theoretical dimensions of all wet etched features, design rule (e) was extremely important. Any features exposed in the first level which were not re-exposed in the second level (and at a larger size) would result in large and variable differences between the size of feature expected and that calculated. This was found to be due to the photoresist not adequately coating the edges of wet etched features. Thus the second level pattern had to expose these areas at a larger size to remove the photoresist bordering any feature wet etched in the first level. This meant that in effect it was the second level exposure which determined the size of wet etched features appearing on the polished side of the wafer.

Alignment of the pattern to the [110] axes was accomplished by using a grating to align to the small flat of the three inch wafer. This grating was only present in the first level pattern. The large flat was not exposed to prevent it becoming damaged as it was used as the principle means of orienting a whole three inch wafer correctly in the beamwriter. The first level pattern was exposed using a ferric oxide mask plate (made by copying a chrome mask plate defined by the beamwriter). This was semi-transparent so that alignment

could be carried out easily. The first level pattern also put down markers to allow the second level plate to be aligned correctly. The second level pattern was exposed directly from an opaque chrome mask plate, as ferric oxide copies were not as dimensionally accurate due to the nature of the exposure process used [see Appendix F for details]. Since the second level pattern is the one determining the dimensions of all wet etched features, the fact that the first level plate was a ferric oxide plate was of no consequence.

The final mask set was designed to allow the whole wafer to be easily cleaved into quarter wafers by the appropriate use of 'v-grooves' as cleave lines. However, to maximise the number of devices, the through-wafer alignment markers were placed only at four 'corners' of the wafer so that the wafer could not be cleaved until at least the first electron-beam lithography level had been carried out. As this level would define its own markers, the wafer could be cleaved up after this step (if desired) and each quarter wafer processed individually. Each quarter wafer was designed to yield 6 rows of 10 devices, *i.e.* 60 per quarter wafer or 240 devices per wafer.

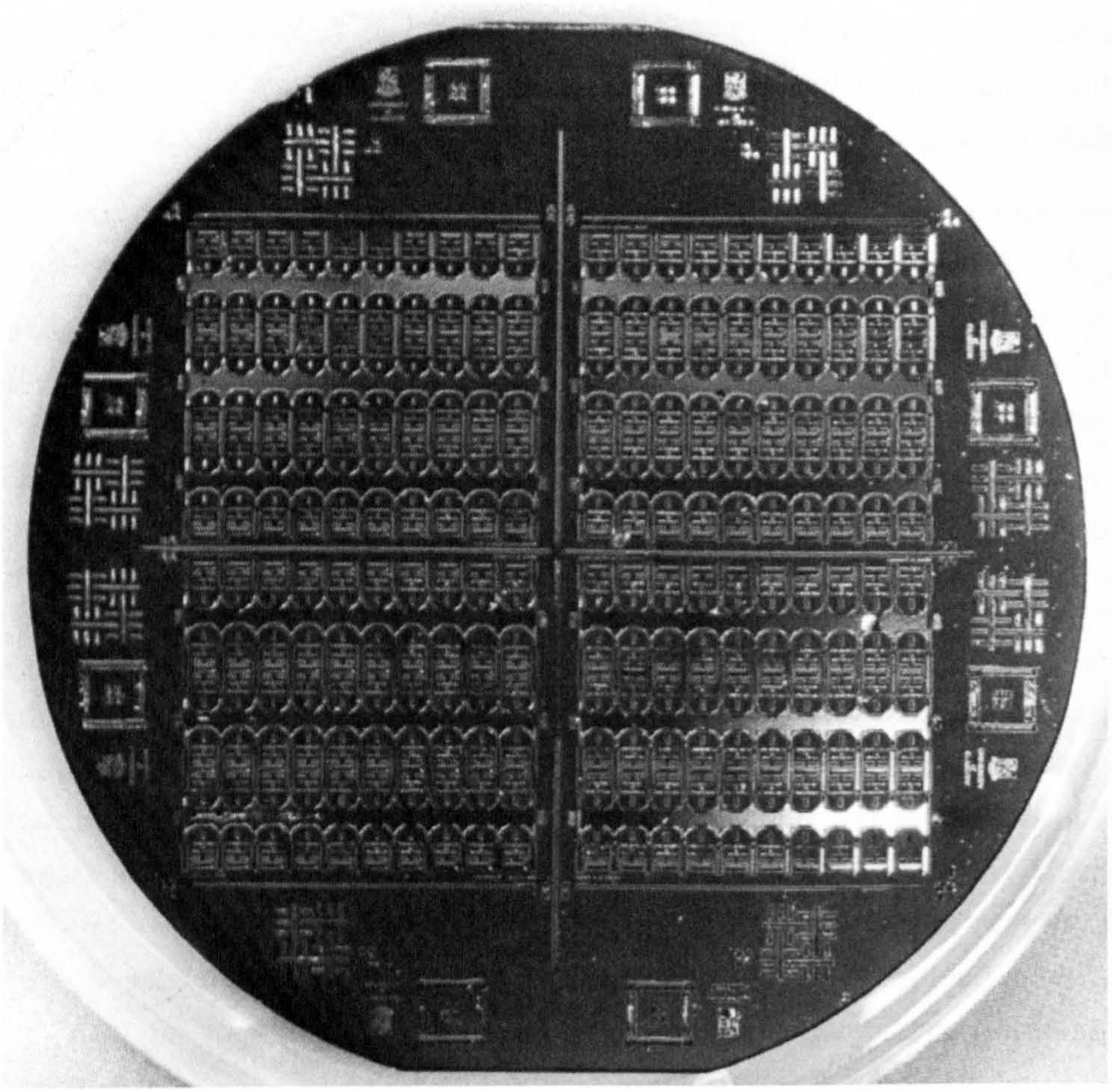


Figure 4.2.4: Photograph of a whole 3 inch wafer after the second wet etch using the final mask set ('amm0012' and 'amm0013') design.

The pattern was designed to accommodate the positions of clamps and height references used by the beamwriter and also included other features, such as the use of a triangular arrangement of 'v-grooves' on the back of each device to allow the 'pseudo' kinematic

mounting of the device onto a microscope, and the presence of the university logo for branding purposes (defined in pixels consisting of small square etch pits) – these are discussed in more detail in Appendix F.

Finally, Figure 4.2.4 is a photograph of the back side (unpolished side) of a whole three inch wafer after the second level wet etch. The developed processing steps for the photolithography levels are listed in Appendix G (§G.1).

4.3 FIRST LEVEL EBL

4.3.1 Tip Fabrication

Details regarding the method used to pattern the silicon nitride masking layer using electron-beam lithography can be found in Appendix E (§E.3). Technical aspects regarding the alignment of the electron-beam lithography level to the through-wafer alignment markers can be found in Appendix H (§H.1).

As mentioned in §3.7.4 the 'egg-timer' profile [see Figure 3.7.8a] produced during the tip wet etching step was perceived as a major problem for devices such as the thermal probe where wires would have to be patterned onto the slopes and over the apex without the loss of electrical continuity. Another important issue was the controllability with which tips could be produced with apices which were of the order of 500nm to 1000nm as this would determine the degree of specimen access and AFM imaging resolution.

One idea, based on the convex corner undercutting effect, was that as those were fast etching planes, alignment of the square silicon nitride mask used to produce the tip with favourable axes for these planes on the (100) surface, may improve reproducibility by virtue of the fact that the degree of undercut would be more immune to a slight rotational misplacement of the mask pattern. Another idea, was to use a circular mask pattern where rotational errors would have no effect. It was also clear that 'convex corner compensation' schemes, as used to compensate for this effect by extension of the corners [Bean[8] (1978)] so that the resulting structure more closely mirrors the uncompensated mask by the end of the etch, could not be used reliably due to the high ratio of etch depth to pyramidal tip plateau size (around 20) required. The compensation in this case would have had to exceed the pyramidal tip plateau size in width alone.

Thus some tests were carried out utilising circular and square masks oriented along the projections of the {111}, {212}, {313/311} planes onto the (100) surface and varying the etching conditions and compositions to determine the best method for fabricating these tips. The {212}, {313} and {311} planes have variously been identified as the fast etching corner undercutting planes in silicon [Lee[9] (1969); Declercq, Gerzberg and Meindl[10] (1975); Bean[8] (1978); Ju and Hesketh[11] (1992); Hesketh, *et al.*[12] (1993)]. The results of these tests are described in some detail in Appendix J (§J.1).

The eventual solution appeared to be the use of an alternative etching solution consisting of a mixture of IPA and potassium hydroxide solution. In the literature the addition of IPA to potassium hydroxide is known to change the etch selectivity and relative etch rates of various planes [Bean[8] (1978); Lambrechts, *et al.*[2] (1990-1993); Mastrangelo, *et al.*[3] (1994)]. This may be one reason why the 'egg timer' profile was eliminated using this etch.

As IPA is not fully miscible with water, in the quantities used it forms a two-layer solution with the IPA rich solution above the potassium hydroxide rich solution. The etch rates and mask undercutting rates in these two layers were found to differ, but etching within the lower layer produced a smoother surface. Using this etch it was found that masks oriented along the [013] axes were undercut most quickly. It was decided that this mask orientation be used to produce all pyramidal tips because the mask undercutting rate must have a local maximum with respect to orientation along the [013] axes. This would then be expected to result in more reproducible tips by reducing the effects of any misorientation of the mask with respect to the crystallographic axes. From the inclination of the slopes of the resulting tips it was concluded that the {313} planes were the fastest undercutting planes rather than the {311} (or {212}) planes.

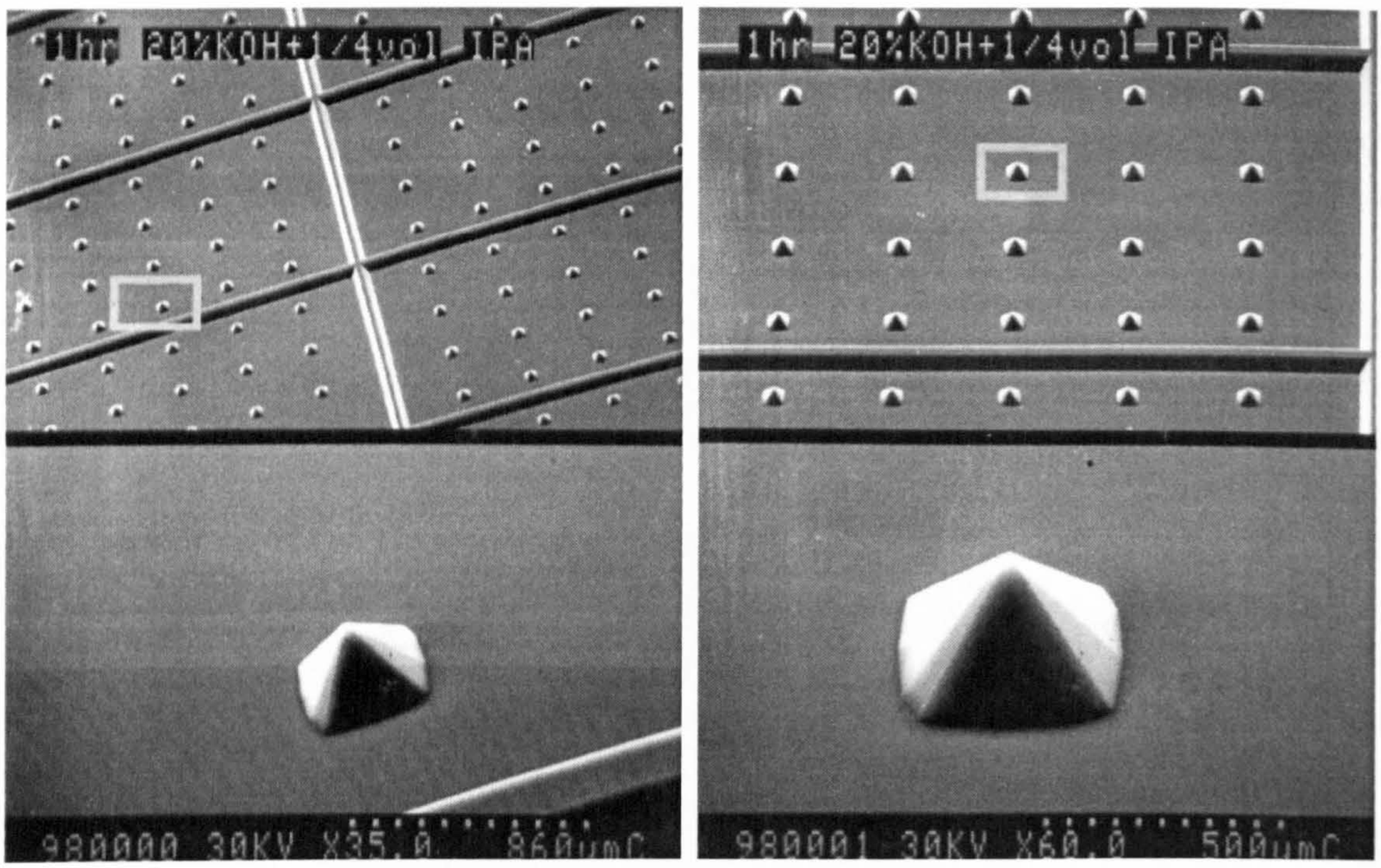


Figure 4.3.1: Micrographs showing the improvement in tip profile achieved by using the potassium hydroxide/IPA etching solution. An improvement in the surface quality is achieved by eliminating the polymer masking problem experienced during pyramidal tip etching [see Appendix J (§J.2)]. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

An additional problem with the use of this etch was the poorer quality of surface finish achieved when compared with pure potassium hydroxide etching. Some tests were carried out to establish the cause of this from which it was found that during the initial stages of this etch a polymer was forming on the surface. This served to mask the etch for

a short time in an uncontrollable manner and may help to explain the observation by Lambrechts and Sansen that IPA-potassium hydroxide etching solution results in 'hillock formation' during the first few hours of etching [Lambrechts, *et al.*[2] (1990-1993)]. Unfortunately no reason or solution had been given. However, after some further tests were carried out [Appendix J (§J.2)], it was found that by altering the deoxidation solution used (to prepare the sample for etching) from buffered hydrofluoric acid to unbuffered hydrofluoric acid the formation of this polymer could be prevented. The use of this deoxidation solution to strip off the silicon nitride after the tip etching step also seemed to eliminate the formation of a brown deposit [which may be porous silicon – see Appendix J (§J.2)] that would occasionally appear on the silicon surface and which could not be removed in any concentrated acid or by cleaning.

In order to fabricate tips with the smallest plateaux, the rate of etching was slowed down by reducing the etch temperature. A calibration sample was also found to be necessary. This simply consisted of different sizes of square masks oriented along the [013] axes and was etched immediately before etching the sample in order to calibrate the mask undercutting rate. It was found that these calibration samples had to be written on a whole three inch wafer in order to obtain a suitably accurate orientational alignment of the masking pattern to the crystal axes of the wafer – see Appendix J (§J.1).

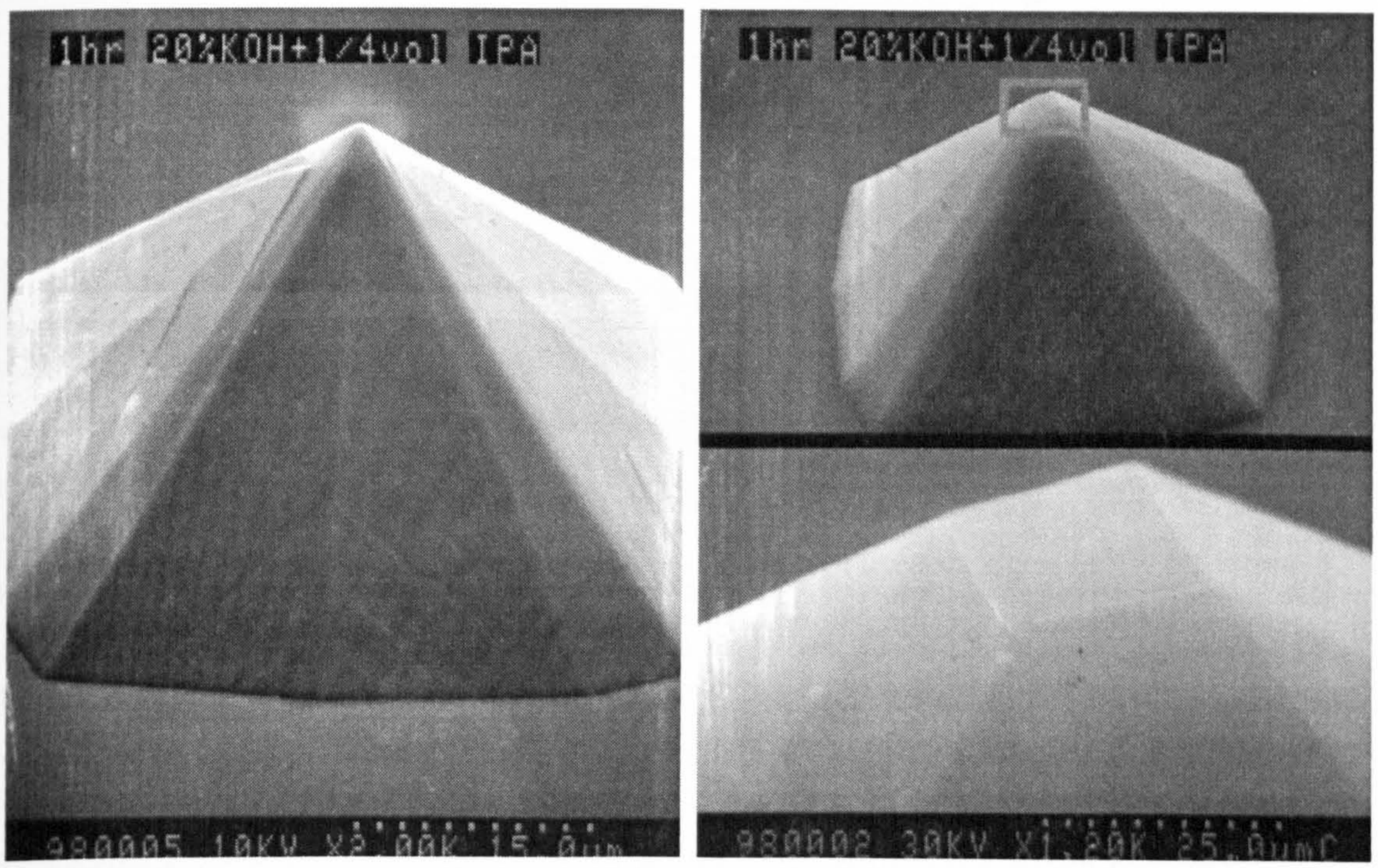


Figure 4.3.2: Micrographs showing the improvement in tip profile achieved by using the potassium hydroxide/IPA etching solution. An improvement in the surface quality is achieved by eliminating the polymer masking problem experienced during pyramidal tip etching. Notice how the surface quality of the tips themselves is also much improved over those in Figures J.4 through J.7 in Appendix J. The bottom half of the right image is a 10x magnification over the displayed scale which applies to the top half.

The final scheme then consisted of using 10μm [013] oriented masks in 20% by weight potassium hydroxide solution:IPA 3:1. Undercutting by 9μm then takes approximately 45

minutes and gives tips which are about 21μm tall. Figures 4.3.1 and 4.3.2 show some tips produced in this way. Also note the reduction in surface roughness and improved smoothness of the {111} planes of the tips achieved using the altered deoxidation stage (compare with Figures J.4 through J.7 in Appendix J). Although a calibration sample is necessary, the results are generally more reproducible and the surfaces have a near-mirror finish after etching. No satisfactory explanation was ever found for the 'egg-timer' profile obtained with pure potassium hydroxide etching, however, the prediction of the intermediate and final geometry of etched convex corner structures in silicon wet etching is widely acknowledged as a complex problem [Sequin[13] (1992); Hesketh, *et al.*[12] (1993)].

4.3.2 First Level EBL Pattern Design

4.3.2.1 Crack Arresters

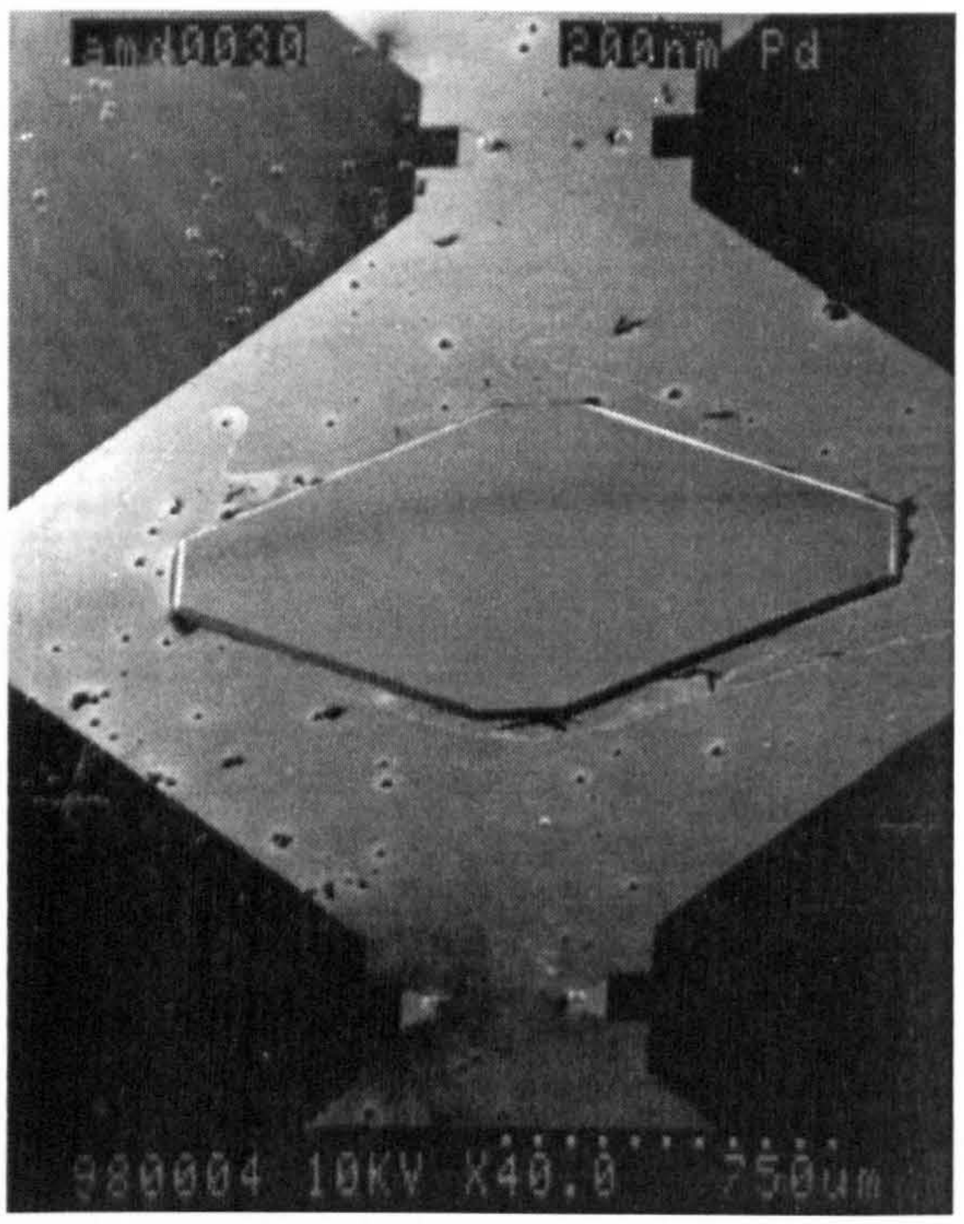


Figure 4.3.3: Micrograph showing the design of the crack arrester used in the EBL levels. The positions of the cantilevers with respect to this can be clearly seen. The crack arrester will dissolve away in the final release etch.

To help tackle the problem of the substrates cracking after the tip definition etch due to the presence of a fragile, large area, cantilever membrane, it was decided that, in addition to the measures already taken in the photolithography levels [see §4.2.3], some crack arresters be placed in the membrane area along the cleave lines. Figure 4.3.3 shows an early version of these crack arresters. They took the form of an unetched silicon area placed where the substrate was weakest.

Since protruding areas are undesirable in the final device (they could limit the access of the tip to the specimen), these arresters were designed so that they would dissolve during the release etch. Figure 4.3.4 shows how the size of the arresters were increased in the final incarnation of the design.

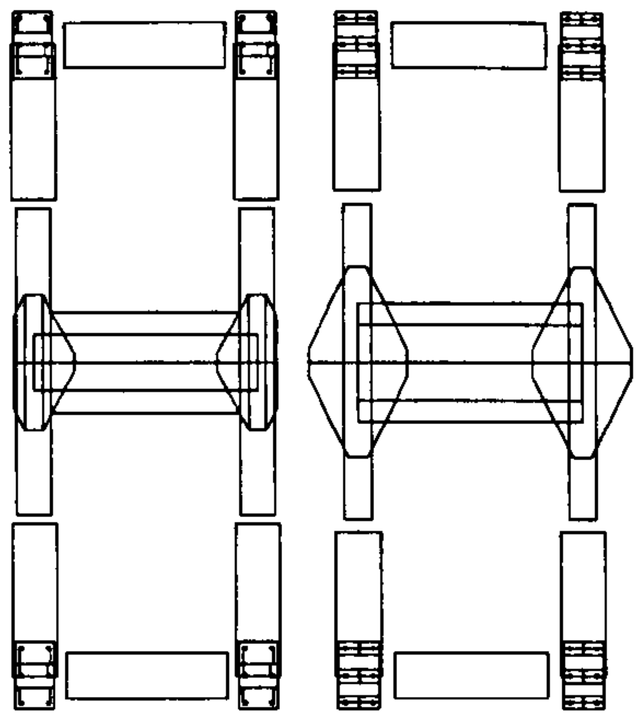


Figure 4.3.4: The pattern used to produce crack arresters during first level EBL. Their size was increased from that of Figure 4.3.3 (left pattern) in the final design (right).

4.3.2.2 Marker Placement Accuracy

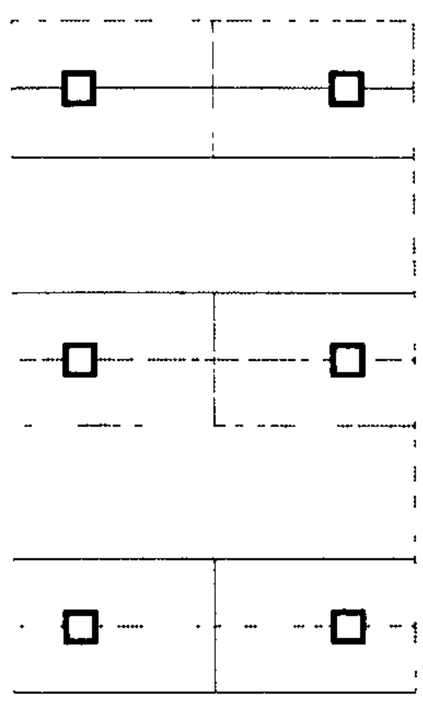


Figure 4.3.5: The final design used to produce topological markers. The marker boundaries were written in a separate matrix from that in which the plateau area was filled out.

As discussed previously in §3.5.3 the key to accurate placement of the markers with respect to the pyramids is to make sure that the first EBL level is written as quickly as possible, and certainly before the beamwriter recalibrates. In order to accomplish this, it was found necessary to split the pattern into two matrices. The first matrix would simply write the critical edges of the marker and the squares used to define the pyramidal tips. This would use a spot size of 112nm and fractionation resolution of 0.05μm to ensure that the edge profile of the resist is as sharp as possible for lift-off purposes. The second matrix would fill out the plateaux around the markers and write the crack arresters using a spot size of 400nm and fractionation resolution of 0.2μm. The final design is illustrated in Figure 4.3.5.

4.3.2.3 In-Situ Tip Etch Calibration

The final version of the photolithography mask was designed so that the first level EBL pattern had to be written as a whole 3 inch wafer, resulting in a large amount of free space becoming available at the edges of the wafer. Thus it was decided that an optical in-situ technique for determining the rate of the pyramidal tip etch be designed to make use of this area. The devised pattern consisted of six 4mm square areas with line gratings

of 7μm, 8μm, 9μm, 10μm, 11μm and 12μm in width (one size in each area) aligned along the [013] axes. These gratings were designed so that they had the same fill factor (ratio of patterned area to unpatterned area) and hence would have the same average optical intensity when reflecting visible light. However, as each area consisted of different width lines, periodically, as the etch undercuts the mask, each area would become fully undercut one by one. Since the pyramids were being fabricated from 10μm squares aligned along the [013] axes, when the area with 9μm lines became fully undercut, the pyramids would have one micron wide tops. In order to enhance the optical intensity, the metallisation used to etch the silicon nitride would be changed to 10nm nichrome (60% nickel/40% chromium), 20nm gold and 100nm titanium. During the cleaning (nanostrip) and deoxidation steps, prior to the etch, the titanium would be attacked rapidly, leaving a reflective gold surface capable of surviving the etch. As each line becomes undercut, residual stress in the LPCVD silicon nitride and metal layers would force the line to curl up from each edge towards the centre thus reducing the reflectivity of that area. By plotting the etching time for each area as a function of area line width, a straight line could then be fitted to extrapolate the amount of time required to etch the tips. Unfortunately, there was insufficient time to test how well this scheme would work in practice. However, this pattern was designed to completion and could be run as an additional matrix with a spot size of 400nm and fractionation resolution of 0.2μm.

4.3.2.4 The Final Version of the Pattern

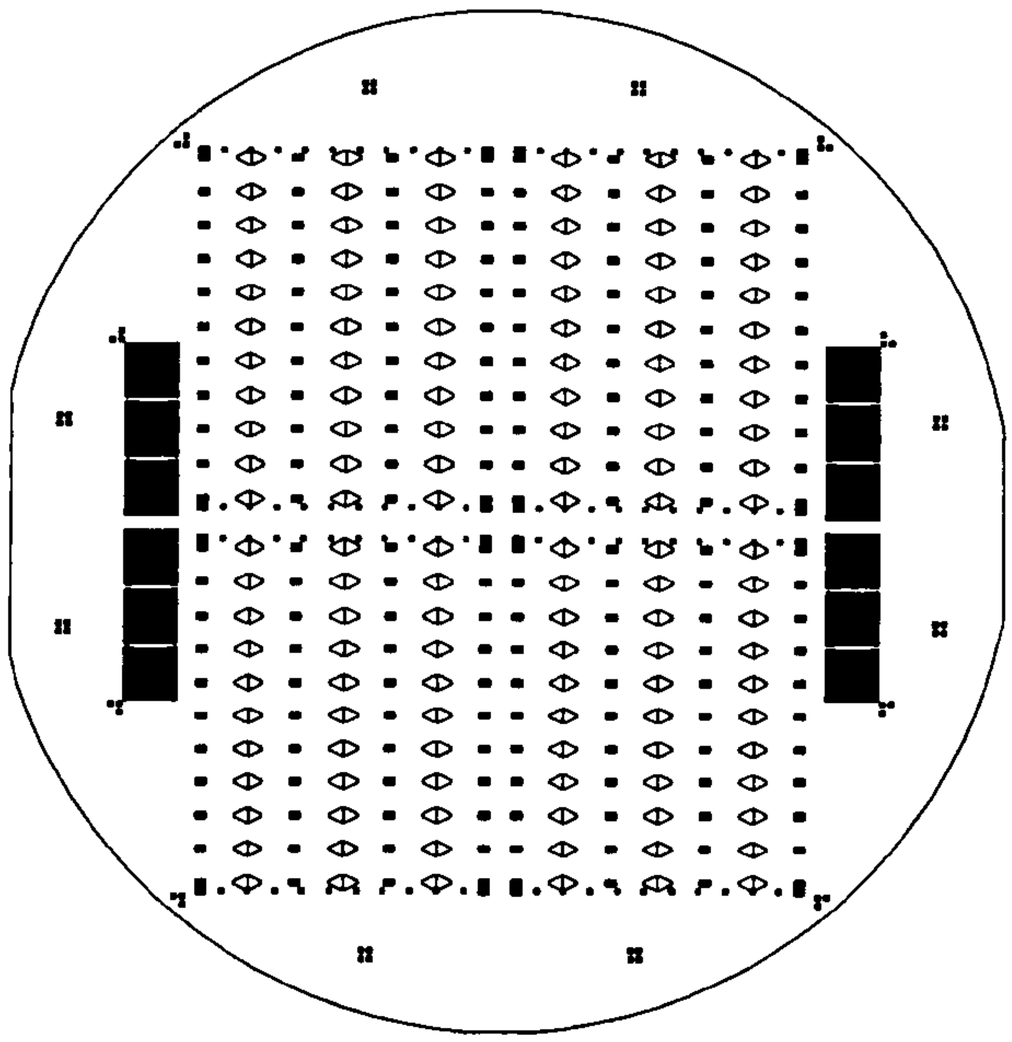


Figure 4.3.6: The final design of the first level EBL pattern for whole 3 inch wafers. The through-wafer alignment markers are also shown for comparison. The large, darker square regions on the left and right of the wafer are the optional in-situ pyramidal tip etch calibration structures as discussed in §4.3.2.3.

Figure 4.3.6 shows the final version of the pattern on the whole 3 inch wafer. Figures 4.3.7 and 4.3.8 show the pattern for each of the first two matrices on a quarter wafer basis (the third matrix is for the optional grating pattern used for in-situ tip etch calibration as discussed in §4.3.2.3). Notice that each marker plateau contains six

markers on a 300μm pitch. There is one plateau for local alignment to each pair of devices and an additional plateau at the edge of each quarter wafer for global alignment purposes.

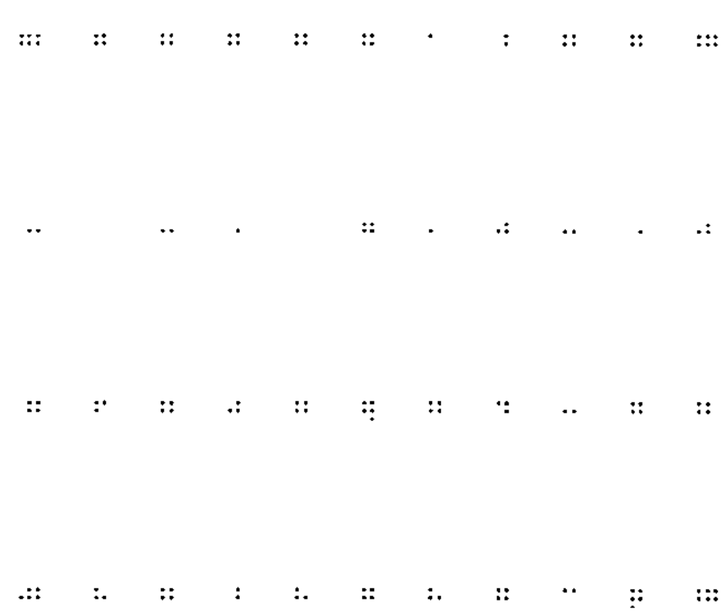


Figure 4.3.7: The final design of the first matrix of the first level EBL pattern for a quarter wafer. Only the marker boundaries are written here, quickly and with a smaller spot.

The developed processing steps for the first electron-beam lithography level are listed in Appendix G (§G.2). Detailed descriptions of the pattern file names and customised job sheet developed for this purpose along with distances from crosses to each marker are given in Appendix M.

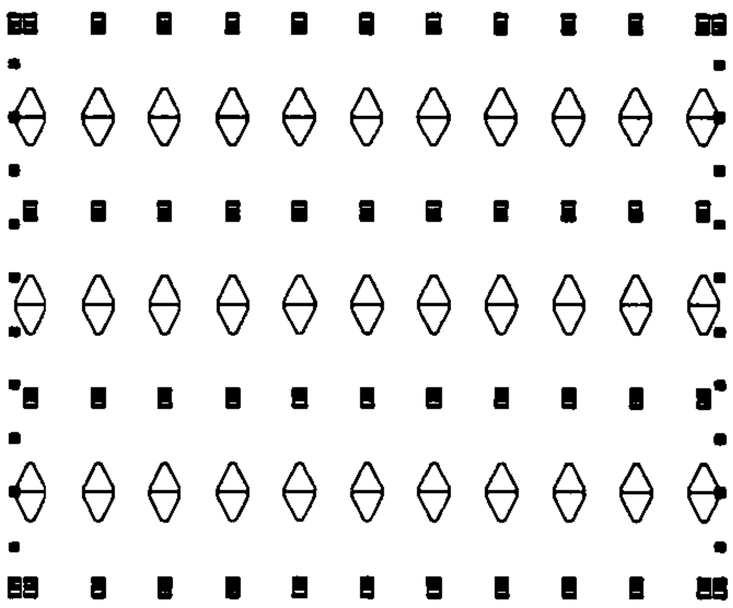


Figure 4.3.8: The final design of the second matrix of the first level EBL pattern for a quarter wafer. The marker plateaux, crosses and crack arresters are written here.

4.4 SECOND LEVEL EBL

4.4.1 Cantilever Design

As discussed in §1.3 contact mode AFM cantilevers typically have compliances in the range 0.005 to 5N/m with a low quality factor. Non-contact mode AFM cantilevers typically have compliances in the range 5 to 500N/m with a high quality factor.

Some equations for the compliance of cantilevers with simple geometries have been derived in Appendix K (§K.1). These equations will be used in the following sections which discuss the practical aspects of fabricating cantilevers. Table 4.4.1 lists some useful material parameters for use in these equations from Sarid[14] (1991).

Material	Density (kg/m^3)	Young's Modulus (GPa)
Aluminium	2700	70
Copper	8960	110
Gold	19300	80
Iridium	2250	520
Iron	7870	193
Nickel	8900	207
Platinum	24200	170
Silicon	2330	179
Silicon Dioxide	2200	60
Silicon Nitride	3100	150
Titanium	4500	116
Tungsten	19300	345

Table 4.4.1: Material parameters for some cantilever materials reproduced from Sarid[14] (1991).

It is clear from Equation K.51 that in general, the stiffer cantilevers required for non-contact mode AFM can be best achieved by increasing the cantilever thickness. The quality factor, Q, of a cantilever beam operating in atmospheric pressure varies according to [Blom, Bouwstra, Elwenspoek and Fluitman[15] (1992)]:

$$Q \propto \frac{t^2}{l^2}, \tag{4.4.1}$$

in terms of the cantilever parameters. Thus, it is also clear that the best way to reduce the compliance and quality factor of cantilevers for contact mode AFM is to increase the length and reduce the thickness. However, practical fabrication issues dictated that different materials be used for each type of cantilever.

In the case of cantilevers for contact mode AFM, the thickness required is normally less than 1μm so that making the cantilever out of silicon was impractical due to the large thickness tolerances inherent in the through-wafer process and the ease with which such a thin cantilever could be eroded by the silicon wet etch. In addition, the high mechanical quality of the single crystal material means that the quality factor of such a cantilever will be higher than that of polycrystalline or amorphous materials. The obvious materials are then silicon oxide (either thermally grown or deposited by CVD) and silicon nitride (deposited by CVD). However, a thermally grown silicon oxide layer would be more problematic because the oxide would be present on both sides of the substrate and would

therefore require an additional step to remove the oxide from the deep etched side of the substrate. In addition, a thermally grown layer is more stressed due to the elevated temperatures required for growth and has been found to be inferior to silicon nitride for the formation of AFM cantilevers [Albrecht, Akamine, Carver and Quate[16] (1990)]. As mentioned earlier [see §3.7.3 and §4.1.2] silicon oxide is attacked at a reasonable rate by the potassium hydroxide based silicon wet etch so that the adhesion of metallic features onto such a cantilever could be compromised. This then leaves silicon nitride as the only practical choice.

In the case of cantilevers for non-contact mode AFM, the thickness required is normally several microns which rules out silicon nitride because films of such thickness deposited by CVD tend to crack due to the residual stress locked into the film [Sze[17] (1985)]. In this case a high mechanical quality material is desirable so that silicon becomes a good choice of cantilever material. Using silicon nitride as a masking and insulating layer on top of the silicon cantilever ensures that adhesion and electrical isolation problems are avoided.

4.4.2 Cantilevers for Contact mode AFM

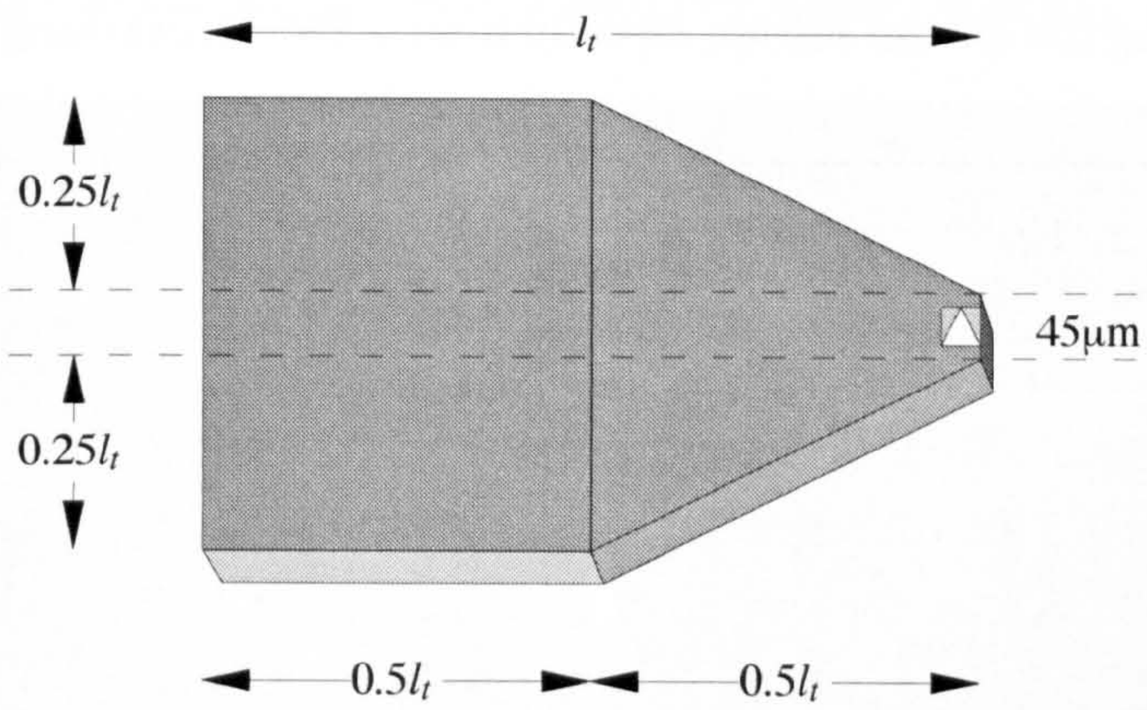


Figure 4.4.1: Diagram showing the geometry used for contact mode AFM cantilevers.

The general procedure for developing the cantilever fabrication level was as follows:

(a) Deposit cantilever material of the required thickness on the tip side of the substrate.

(b) Pattern the cantilever using electron-beam lithography and dry etching.

(c) Release the cantilever by etching the substrate in 7 molar (+5%) potassium hydroxide solution.

From Appendix G (§G.1) and Appendix A (§A.1) the device area of the substrate would be etched by 290μm to 300μm after the photolithography levels had been carried out. From Appendix J (§J.1) the tip definition etch results in tips which are between 17.5μm and 22.5μm tall. As this etch is from both sides, the cantilever membrane is reduced in thickness by a further 35μm to 45μm as a result [see Figure 4.2.2]. For a wafer between 380μm and 390μm thick this means that the cantilever membrane is between 25μm and 65μm thick. This means that on average the release etch would have to etch to a depth of

45μm, although an additional 20μm of etching would be required to clear the silicon out of the tip. However, the unmasked parts of the membrane which would etch from both sides would only require half this value of etching. It therefore made sense to make use of convex corner undercutting to give access to the pyramidal tip quicker and undercut the cantilever quicker than would otherwise be the case.

It was therefore decided that half of the cantilever would be tapered along the [013] axes with the remainder (the part attached to the bulk substrate) being oriented along the [011] axes. This is illustrated in Figure 4.4.1. Note that 30μm was allowed for the base of the tip, although the base of a 20μm tall tip typically exceeds this by approximately 10μm, facilitating the etching of silicon within the tip itself.

An initial attempt to fabricate simple cantilever beams using the precoated 100nm LPCVD silicon nitride masking layer revealed many of the problems associated with fabricating thin film cantilevers. Figure 4.4.2 shows two micrographs of these beams which were meant to be used to measure the piezoresistive behaviour of nichrome. 50nm thick nichrome resistors were patterned onto the silicon nitride. The silicon nitride cantilevers were then release etched by simply forming a silicon etch pit, with the mask defined by dry etching the silicon nitride directly from a PMMA masking layer.

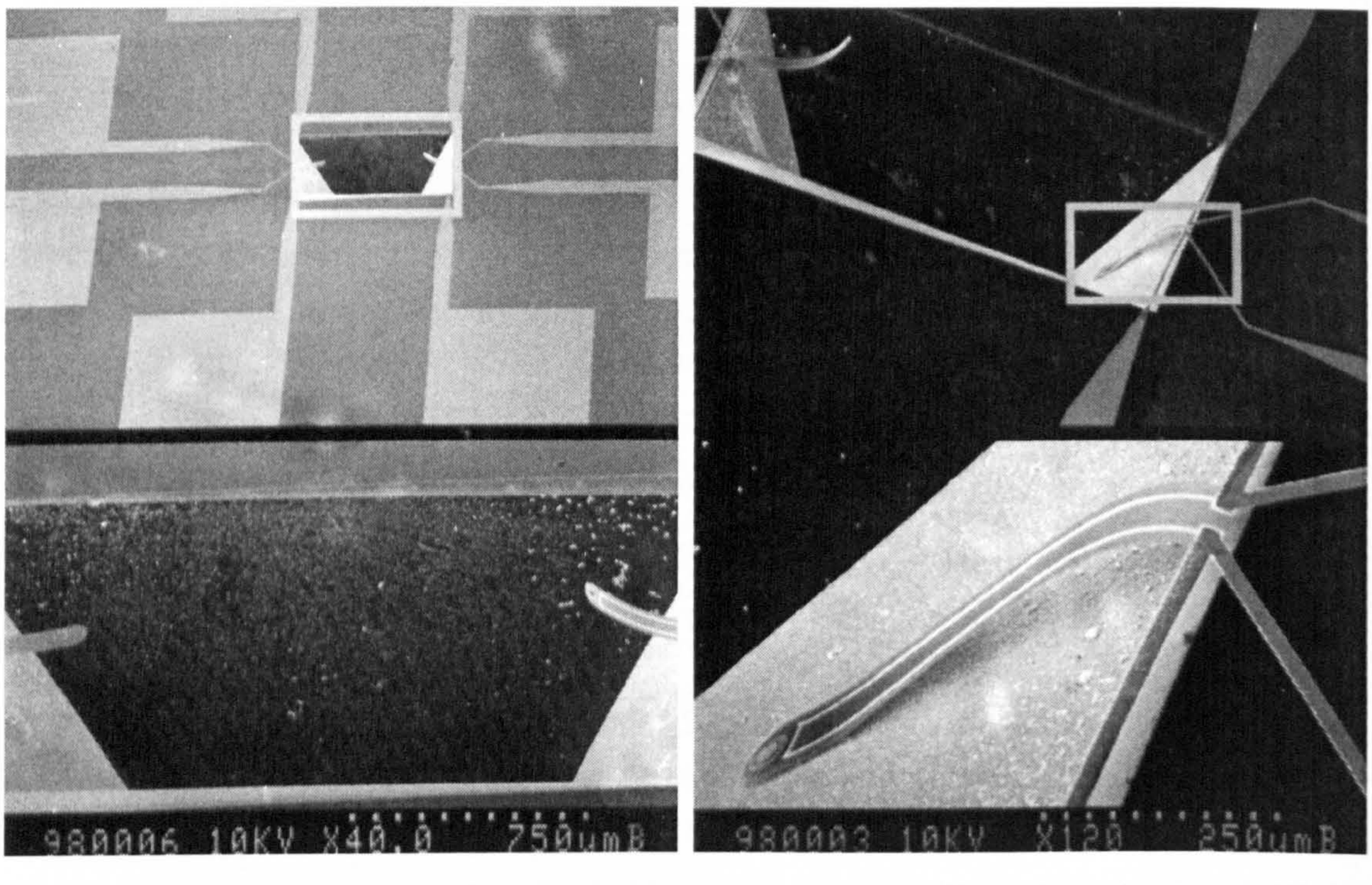


Figure 4.4.2: Micrographs showing some of the problems associated with releasing cantilevers for contact mode AFM. (a) Left Image: 100μm long, 20μm wide, 50nm thick LPCVD silicon nitride cantilever beams; (b) Right Image: 150μm long, 20μm wide, 50nm thick LPCVD silicon nitride cantilever beams. The bottom images are a 5x magnification over the displayed scales which apply to the top half.

Figure 4.4.2a shows a pair of facing 100μm long, 20μm wide cantilevers. It is clear that the metal pattern on the left beam has come unstuck during the etch suggesting that adhesion of metallic patterns through a release etch cannot be taken for granted.

Furthermore it is clear that both beams curve upwards. This is due to residual stresses in the silicon nitride layer introduced during deposition. Stress is more of a problem with LPCVD layers due to the high deposition temperature [Sze[17] (1985)]. In addition it can be seen that the cantilever with metallic features bends more than that without due to the additional residual stress introduced by the metal itself (although the difference between the thermal expansion coefficient of silicon nitride and nichrome means that an effect similar to that used in bi-metallic strips may also be playing a role here). This effect is clearly more significant for thinner cantilevers due to the cubic dependence of compliance on beam thickness [see Equation K.51].

Figure 4.4.2b shows a pair of facing 150μm long, 20μm wide cantilever beams. It can be seen that the beam on the right has become stuck to the substrate. This is caused by capillary forces acting between the cantilever and substrate as the substrate was dried after being removed from the etchant [Kobayashi, Kim and Fujita[18] (1993); Legtenberg, Tilmans, Elders and Elwenspoek[19] (1994)]. Clearly this problem has been exacerbated by the shallow angle between the cantilever and substrate in this instance. Again it is also obvious that longer cantilevers are more vulnerable to this problem due to the inverse cubic dependence of compliance on beam length [see Equation K.51].

It was therefore decided that a cantilever of intermediate length and reasonable thickness be used for development purposes. The dimensions of the test cantilever were chosen to be a length of 250nm and a thickness of 500nm, giving an expected compliance of about 0.05N/m and a resonant frequency of around 10kHz [see Appendix K (§K.2)].

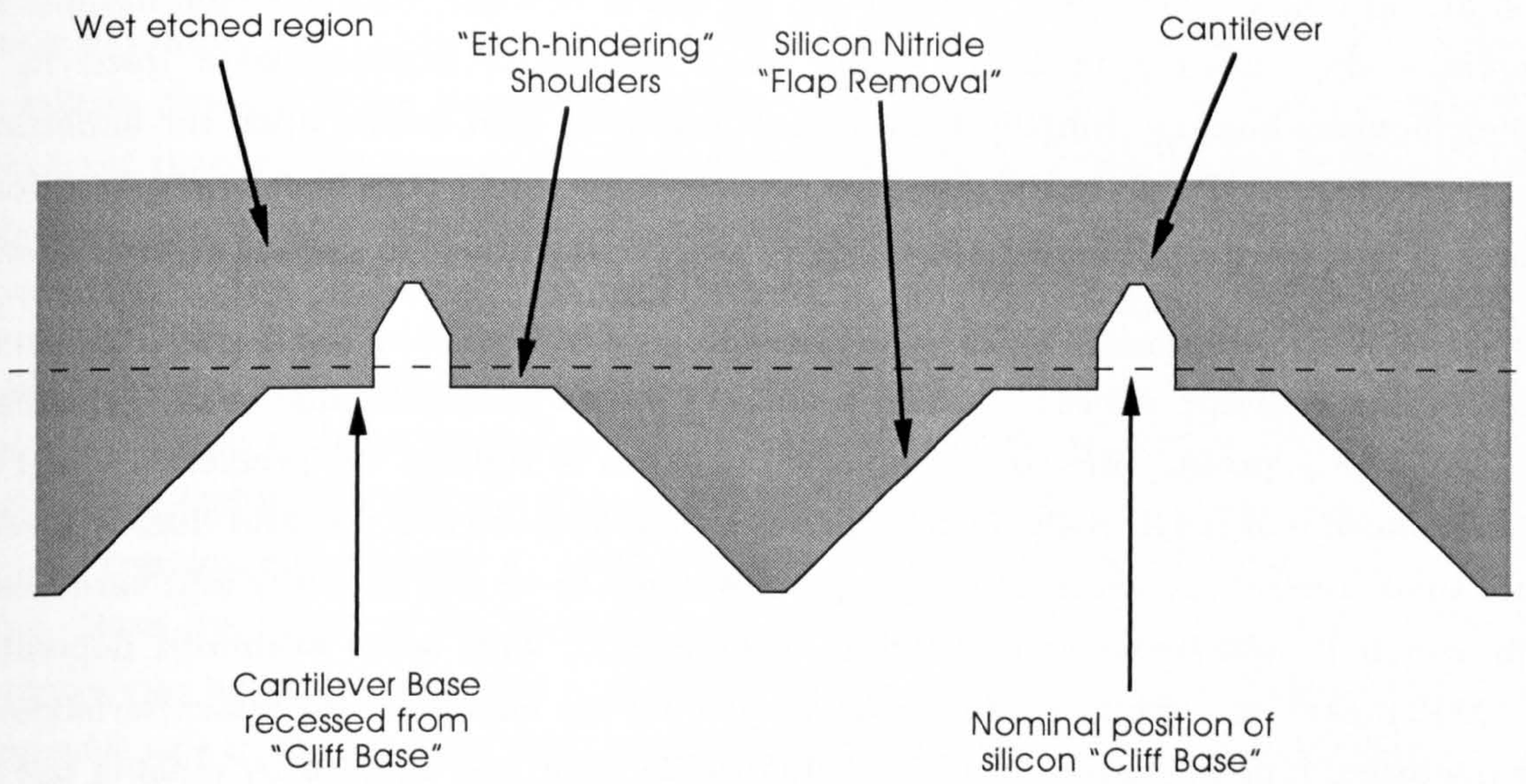


Figure 4.4.3: Diagram showing the pattern used to wet etch the contact-mode AFM masking layer.

The first problem was the patterning of films of such thickness. It is clear that whilst a thick layer of PMMA was a possibility, coating the apices of the pyramidal tips in a resist layer of sufficient thickness so as to prevent any erosion of the silicon nitride that would eventually form the tips would be difficult. Then the first possibility considered was to

use a similar technique as in the first EBL level [see §4.3]. However, unlike with the first EBL level, here the area to be protected from the etch would be far larger than that to be etched. Therefore, instead of using lift-off to create a metal masking layer, it was decided that wet etching be used since dimensional tolerances of 1μm would be perfectly acceptable for this level. To tackle the residual stress problem, Dr. Simon Hicks was able to offer PECVD silicon nitride with a controlled level of residual stress, ranging from tensile through zero to compressive.

After some development the final scheme consisted of sputtering a 100nm thick chromium layer upon the low stress silicon nitride layer already deposited on the substrate to the cantilever thickness required. The substrate was then coated with PMMA resist and patterned using the pattern illustrated in Figure 4.4.3. After developing the resist pattern, the chromium layer was wet etched to expose appropriate areas of the silicon nitride for dry etching. The silicon nitride layer was then dry etched and the chromium and resist masking layers stripped off. The development of this process and an explanation for each step is given in more detail in Appendix L (§L.1). However some of the main points will be outlined here.

Firstly, coating these substrates with PMMA was not carried out simply using the spin coating technique, as it was found that spin coating would not adequately coat the apices of the pyramidal tips. Instead it was combined with an alternative coating technique, which has been named 'float coating'. The development of this technique is discussed in detail in §5.2. Suffice to say that with this technique the resist coating is highly stressed so that the resist close to the base of the pyramids is sometimes not continuous. To overcome this, both spin coating and float coating techniques were used in a complementary manner. Initially PMMA resist would be spin coated upon the substrate and then baked overnight in the standard manner. This would then be followed by a float coated PMMA resist layer and a short bake at lower temperature to prevent resist reflow.

While initial attempts based on the use of evaporated nichrome as the dry tech masking layer were successful, subsequent attempts were not as successful due to the deposited layers cracking, peeling and curling up at a thickness of 100nm. The problem seemed to affect isolated and small scale structures more than large structures, and thicker layers more than thinner layers. Another symptom seemed to be the difficulty and variability with which it was being etched using chrome etch, with some stubborn deposits, probably nickel rich, often remaining. The cause for the change in the characteristics of the deposited film was never found and this whole issue was avoided by making use of sputtered chromium films, which proved more reliable and also had the advantage of conformally covering the surface. A 4 inch chromium sputter target was obtained for this purpose to enable the coating of whole 3 inch wafers.

Another problem which arose during development occurred during the transition that had to be made from BDH to Aldrich [see §3.5.3]. It was discovered that using the Aldrich

based float coating scheme resulted in the chromium at the top of the pyramidal tips being attacked by the chrome etch. The reason for this was never clearly identified, though one theory was that because the nature of the float coating technique is such that the PMMA is stretched over the tips [see §5.2], the higher molecular weight of the Aldrich in comparison with BDH means that the resulting layer is more porous. This increase in porosity may be sufficient to allow the active components of the etching solution penetrate the resist layer where the resist is most stressed. Working around this problem unfortunately meant the addition of an extra EBL level prior to dry etching the silicon nitride, but after wet etching the chromium layer and stripping off the PMMA. The purpose of this level was to lift-off a 10μm square of 50nm thick nichrome over the apices of the tips as a protection layer.

The pattern illustrated in Figure 4.4.3 evolved in answer to several issues which arose during the development of this process. Due to the use of convex corner undercutting to taper the silicon substrate approaching the cantilever [see Figure 4.2.3], a simple cantilever pattern would result in the formation of wide silicon nitride flaps. These flaps can break in an untidy way, and endanger the cantilevers of adjacent devices. To avoid this, the silicon nitride in these regions is removed by the patterning process prior to release [see 'Flap Removal' in Figure 4.4.3]. This then necessitates the presence of the 'etch-hindering' shoulders shown in Figure 4.4.3 to prevent the base of the cantilever being undercut by the presence of the "flap removal" areas adjacent to it. [Otherwise the base of the cantilever would be attached to silicon nitride rather than being firmly anchored on the silicon substrate. Then if the cantilever is placed under stress, as in the course of normal use, cracks may form where the base of the cantilever meets the rest of the silicon nitride.] These shoulders are also undercut by the etch, but by the time the etch front reaches the base of the cantilever, the etch will have been completed. The degree of compensation can be easily adjusted by varying the width of the shoulder. Figures illustrating the use of these features can be found in Appendix L (§L.1).

Finally, as the release etch only has to etch through about 45μm of silicon, it was decided that the release etch be carried out at a lower temperature (slower and hence more controllable etching). When the etch has penetrated the wafer, large areas of the wafer become transparent, offering a visual indicator that the etch has reached the halfway point. Since the height of the pyramids and the etch depth of the membrane definition windows can be measured prior to etching (taking precautions to avoid smashing the pyramidal tips in the process) and knowing the initial wafer thickness, the thickness of silicon to be etched can also be determined. When the substrate becomes transparent, an estimate can then be made as to how much longer the etch should be allowed to continue. In this way the etch can be made self-calibrating.

One final technical detail concerns the positioning of the base of the cantilever with respect to the expected edge of the 'cliff base' being etched from the other side of the substrate [see Figures 4.2.2 and 4.2.3, Figure 4.2.2 can be viewed as the cross-section

along a line joining the centre of the 'cliff bases' of devices 'A' and 'B' in Figure 4.2.3]. To accommodate feature growth and some misalignment, the cantilevers were deliberately recessed into the silicon substrate by 5μm [see Figure 4.4.3]. This allowed for a 5μm error in the position at which the 'cliff base' would appear after the release etch. This was done to ensure that the cantilever would always be anchored onto the silicon substrate and avoid the situation where the 'cliff base' appeared behind the cantilever.

The detailed processing steps for this level are listed in Appendix G (§G.3).

4.4.3 Cantilevers for Non-Contact Mode AFM

It is clear from Equations K.51 and 4.4.1 that if a high quality factor is required for the same stiffness of cantilever, then it is better to increase the cantilever thickness at the expense of a reduction in width. Thus a uniform rectangular beam seemed to be the most appropriate cantilever geometry.

Initial thoughts on fabrication focused on the idea of simply vertically dry etching through the silicon membrane using a nichrome metal mask. In order to end up with a beam of between 10μm and 20μm thick, it would be necessary to make sure that the cantilever membrane was sufficiently thin after the tip definition etch. This could be done before the tip definition level, immediately after the second photolithography level wet etch by introducing another slower wet etch between these two levels.

After some initial attempts with the dry etching strategy [as described in Appendix L (§L.2)], it was concluded that an alternative scheme might offer a shorter route to success. This alternative was based on wet etching. The same thickness of silicon membrane as used above but with a silicon nitride mask could simply be wet etched. The advantage of this would be that the resulting cantilever beam would be half the initial thickness of the silicon membrane, as etching takes place from both sides, allowing a factor of two improvement in controllability [see Figure 4.2.2].

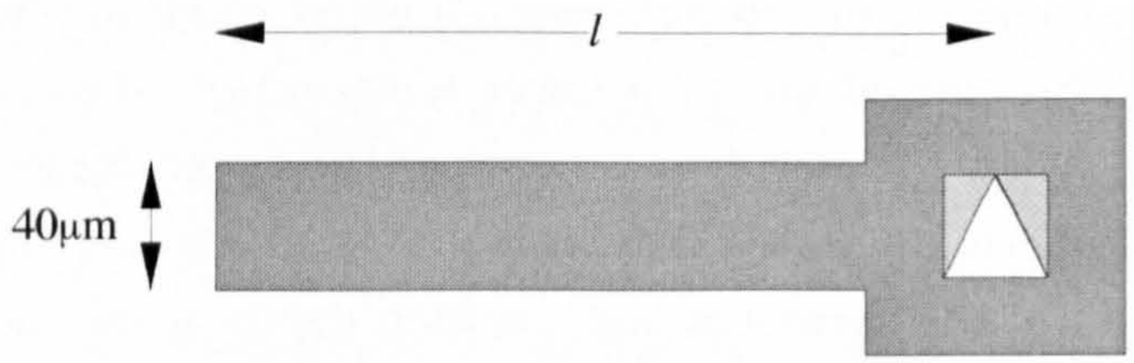


Figure 4.4.4: Diagram showing the geometry used for non-contact mode AFM cantilevers.

The pattern used for such a beam is illustrated in Figure 4.4.4. Note the extra width given around the tip region to offset the effects of convex corner undercutting during the cantilever release etch. The beam would be 40μm wide, but the etch stopping {111} planes would mean that the sides of the beam would not be vertical, but slope at an angle of about 54.74° to the horizontal [see §3.7.3]. This would have the effect of making the cantilever stiffer than a vertical-sided beam.

From some rudimentary estimates of the compliance of such a cantilever [see Appendix K (§K.3)], it was decided that a cantilever length of 250μm and thickness of 5μm, corresponding to a compliance of about 16N/m and a resonant frequency of about 94kHz, be aimed for initially with a view to increasing the length to 350μm, corresponding to a compliance of around 6N/m and a resonant frequency of about 50kHz, in the long term. Since a membrane of down to 10μm in thickness can be produced relatively easily using the extra wet etching step detailed above, a cantilever beam of 5μm in thickness did not seem unreasonable. In any case, the same trick as used for contact mode cantilevers could also be applied here, *i.e.* knowing the thickness of the cantilever membrane prior to release etching and the time taken for the etch to penetrate the substrate by etching half way through this membrane (as indicated by large areas of the substrate becoming transparent) the etch rate could be calibrated in-situ. Then the etch would simply be allowed to continue until the beam had reached the required thickness.

The final fabrication scheme was then very similar to that used for contact mode AFM cantilevers. Assuming that the cantilever membrane had been thinned to an appropriate level by the additional silicon wet etch prior to tip definition, PECVD silicon nitride is deposited on the substrate as an isolation layer (the same thickness as for contact mode AFM substrates can be used so that deposition can be carried out on both sets of wafers together). Again in the interest of process reusablity between the contact and non-contact mode AFM cantilevers, the same thickness of sputtered chromium is patterned as before and used to mask the dry etching of the silicon nitride. Then the substrate undergoes the same release etch as used for contact mode AFM cantilevers.

As with contact mode cantilevers, the pattern used to etch the chromium also had 'flap removal' regions, 'etch hindering' shoulders and were recessed into the substrate by 5μm [see Figure 4.4.3]. Thus, it was possible to devise two similar processes to result in different kinds of cantilever for contact and non-contact mode AFM applications. A more detailed description of the development process and figures illustrating the success of this method in producing 2.5μm thick cantilevers are given in Appendix L (§L.2). The detailed processing steps for this level are listed in Appendix G (§G.3).

4.5 SUMMARY AND CONCLUSIONS

In this chapter the evolution of a process for creating substrates suitable for the definition of nanosensors on AFM tips has been described. It is clear that the process consists of several conventional bulk silicon micromachining techniques, combined with some non-conventional electron-beam lithography levels. Many of the problems discussed in §3.8 have been overcome but many more remain, particularly regarding the fabrication of the nanosensors themselves. The need to carry out the release etch for contact mode AFM devices as the final step will naturally impose a constraint on the kinds of materials

that can be used for these nanosensors: they must not be attacked by the release etching process. In addition, as the contact mode AFM cantilevers are composed of silicon nitride, which is a glassy material, it is semi-transparent so that a reflector will need to be placed upon these cantilevers in order to allow optical deflection detection schemes to be used. The non-contact mode AFM devices clearly do not suffer from this problem as they are composed of silicon and are robust enough to survive additional lithography levels.

4.6 REFERENCES

1. E. Bassous and E. F. Baran, *THE FABRICATION OF HIGH PRECISION NOZZLES BY THE ANISOTROPIC ETCHING OF (100) SILICON*, Journal of the Electrochemical Society **Vol. 125,** No. 8, pp. 1321-1327 (1978).

2. M. Lambrechts and W. Sansen, *Chapter 4: PLANAR TECHNOLOGIES FOR MICROELECTROCHEMICAL SENSORS* in Biosensors: Microelectrochemical Devices, published by Institute of Physics, Leuven, p. 98 (1990-1993).

3. C. H. Mastrangelo and W. C. Tang, *Chapter 2: SEMICONDUCTOR SENSOR TECHNOLOGIES* in Semiconductor Sensors, edited by S. M. Sze, published by John Wiley & Sons, New York, pp. 17-95 (1994).

4. L. Pollock, *Private Communication*, (1993-1996).

5. D. Gourlay, *Private Communication*, (1993-1996).

6. W. Vangelder and V. E. Hauser, *THE ETCHING OF SILICON NITRIDE IN PHOSPHORIC ACID WITH SILICON DIOXIDE AS A MASK*, Journal of the Electrochemical Society **Vol. 114,** No. 8, pp. 869-872 (1967).

7. Shipley, *Technical Manual*, published by Shipley Photo Resist (1980-1993).

8. K. E. Bean, *ANISOTROPIC ETCHING OF SILICON*, IEEE Transactions On Electron Devices **Vol. ED-25,** No. 10, pp. 1185-1193 (1978).

9. D. B. Lee, *ANISOTROPIC ETCHING OF SILICON*, Journal Of Applied Physics **Vol. 40,** No. 11, pp. 4569-4574 (1969).

10. M. J. Declercq, L. Gerzberg and J. D. Meindl, *OPTIMIZATION OF THE HYDRAZINE-WATER SOLUTION FOR ANISOTROPIC ETCHING OF SILICON IN INTEGRATED CIRCUIT TECHNOLOGY*, Journal of the Electrochemical Society **Vol. 122,** No. 4, pp. 545-552 (1975).

11. C. Ju and P. J. Hesketh, *MEASURMENTS OF THE ANISOTROPIC ETCHING OF A SINGLE-CRYSTAL SILICON SPHERE IN AQUEOUS CESIUM HYDROXIDE*, Sensors and Actuators A-Physical **Vol. 33,** pp. 191-196 (1992).

12. P. J. Hesketh, C. Ju, S. Gowda, E. Zanoria and S. Danyluk, *SURFACE FREE ENERGY MODEL OF SILICON ANISOTROPIC ETCHING*, Journal of the Electrochemical Society **Vol. 140,** No. 4, pp. 1080-1084 (1993).

13. C. H. Sequin, *COMPUTER SIMULATION OF ANISOTROPIC CRYSTAL ETCHING*, Sensors and Actuators A-Physical **Vol. 34,** pp. 225-241 (1992).

14. D. Sarid, *Chapter 1: MECHANICAL PROPERTIES OF LEVERS* in Scanning Force Microscopy with Applications to Electric, Magnetic and Atomic Forces, published by Oxford University Press, New York, pp. 1-17 (1991).

15. F. R. Blom, S. Bouwstra, M. Elwenspoek and J. H. J. Fluitman, *DEPENDENCE OF THE QUALITY FACTOR OF MICROMACHINED SILICON BEAM RESONATORS ON PRESSURE AND GEOMETRY*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 1, pp. 19-26 (1992).

16. T. R. Albrecht, S. Akamine, T. E. Carver and C. F. Quate, *MICROFABRICATION OF CANTILEVER STYLI FOR THE ATOMIC FORCE MICROSCOPE*, Journal Of Vacuum Science & Technology A-Vacuum Surfaces and Films **Vol. 8,** No. 4, pp. 3386-3396 (1990).

17. S. M. Sze, *Chapter 9: OXIDATION AND FILM DEPOSITION* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 341-380 (1985).

18. D. Kobayashi, C. Kim and H. Fujita, *PHOTORESIST-ASSISTED RELEASE OF MOVABLE MICROSTRUCTURES*, Japanese Journal of Applied Physics - Part 2 **Vol. 32,** No. 11A, pp. L1642-L1644 (1993).

19. R. Legtenberg, A. C. Tilmans, J. Elders and M. Elwenspoek, *STICTION OF SURFACE MICROMACHINED STRUCTURES AFTER RINSING AND DRYING: MODEL AND INVESTIGATION OF ADHESION MECHANISMS*, Sensors and Actuators A-Physical **Vol. 43,** pp. 230-238 (1994).

CHAPTER 5

EBL ON AFM TIPS

This chapter deals with the problems encountered when trying to align to the apices of, and carry out electron-beam lithography on, pyramidal AFM tips. This includes overcoming the focusing problem caused by the heights of these tips above the substrate to ensure that full advantage is taken of the resolution capabilities of the beamwriter. The development of a technique for coating pyramidal AFM tips with resist is also described. With these problems solved, the main lithographic barriers to achieving the aims of the work of this thesis were removed.

5.1 ALIGNMENT AND DEFOCUS

As explained in §3.5.1, the depth of focus of 100kV electrons with a 200μm aperture located 40mm above the imaging plane (as in the beamwriter arrangement) is of the order of a few microns. While this is large in comparison to the smallest spot size achievable, 12nm, it is still much less than that required for defining sub-100nm features on top of a 20μm tall pyramidal tip. A more correct treatment of the focusing error, incorporating elements of gaussian optics, will now be given here [compare with §3.5.1].

For gaussian beams, Equation 2.2.1 must be modified to [Newport[1] (1993)]:

$$d_o = \frac{4\lambda}{\pi}F, F = \frac{f}{D} \sim \frac{1}{2NA}, NA = \frac{D}{2\sqrt{0.25D^2 + f^2}} \sim \frac{D}{2f}, D << f, \tag{5.1.1}$$

where d_o is the diffraction limited spot diameter, λ is the wavelength of radiation used, F is the f-number or relative aperture of the optical system, f is the effective focal length of the optical system, D is the diameter of the exit pupil and NA is the numerical aperture of the optical system. The radius of a gaussian beam as a function of distance, x, from the focal point where its radius is r_o is given by [Newport[1] (1993)]:

$$r^2(x) = r_o^2\left[1 + \left(\frac{\lambda x}{\pi r_o^2}\right)^2\right] = r_o^2\left[1 + \left(\frac{4\lambda x}{\pi d_o^2}\right)^2\right] = r_o^2\left[1 + \left(\frac{\pi x}{4\lambda F^2}\right)^2\right], \tag{5.1.2}$$

where Equation 5.1.1 has been applied. The distance from the focal point to that where the spot radius (and diameter) has increased by √2 is known as the Rayleigh range, x_r, and is given by:

$$x_r = \frac{\pi r_o^2}{\lambda} = \frac{4\lambda F^2}{\pi} = d_o F, \tag{5.1.3}$$

where Equation 5.1.1 has been applied again. Equation 5.1.2 can then be written as:

$$r^2(x) = r_o^2\left[1 + \left(\frac{x}{x_r}\right)^2\right]. \tag{5.1.4}$$

The difference between the two positions where the spot radius (and diameter) has increased by √2 is known as the depth of focus, d_f, and from Equation 5.1.4 is clearly given by:

$$d_f = 2x_r = \frac{2\pi r_o^2}{\lambda} = \frac{\pi d_o^2}{2\lambda} = \frac{8\lambda F^2}{\pi} = 2d_o F, \tag{5.1.5}$$

where Equation 5.1.3 has also been used. [Compare this with Equation 2.2.2.]

Now from §3.5.1 the wavelength of 100keV electrons is 3.70pm. However, the f-number for the beamwriter column with a 200μm aperture is 200. Putting these values into Equation 5.1.1 shows that the diffraction limited spot diameter is 9.43Å. The minimum achievable spot diameter is 12nm due to aberrations. In order to obtain accurate values for the depth of focus and other parameters a method for treating the aberrations must

also be given. A rigorous treatment of these aberrations is beyond the scope of this thesis, however, in common with other, more simplified treatments, [*e.g.* Pease[2] (1981); Watts[3] (1988)], the aberrations will be treated as if they generate a blurring effect or circle of confusion, so that the actual spot diameter, d_a, is given by:

$$d_a^{\,2} = d_o^{\,2} + d_{ab}^{\,2}, \tag{5.1.6}$$

where d_o is the unaberrated spot diameter and d_{ab} is the spot diameter component associated with (all of) the aberrations. Then it is clear that when the system is focused, an actual spot diameter of 12nm and an unaberrated spot diameter of 9.43Å means that the aberrated spot diameter is 11.96nm.

Now, assuming the aberration diameter does not vary as a function of position, the diameter of the spot, d, at any position, is given by combining Equations 5.1.4 and 5.1.6 to give:

$$d^2(x) = d_{ab}^{\,2} + d_o^{\,2}\left[1 + \left(\frac{x}{x_r}\right)^2\right]. \tag{5.1.7}$$

It is also clear that the equation for the depth of focus is incorrect if the depth of focus is defined in terms of the minimum achievable diameter as opposed to the diffraction limited diameter. In its place another parameter, the focal range will be defined here: this is the difference in position between the points where the beam diameter is at its minimum value and has grown by $\sqrt{2}$, and is analogous to the definition of the Rayleigh range in Equation 5.1.2, but applies even when the minimum diameter is not the diffraction limited beam diameter. Then by applying Equation 5.1.7 to this definition, the focal range, x_f, is given implicitly by:

$$2d_a^{\,2} = d_o^{\,2}\left[1 + \left(\frac{x_f}{x_r}\right)^2\right] + d_{ab}^{\,2} = d_a^{\,2} + d_o^{\,2}\left(\frac{x_f}{x_r}\right)^2, \tag{5.1.8}$$

from which:

$$x_f = x_r \frac{d_a}{d_o}. \tag{5.1.9}$$

The Rayleigh range for 100keV electrons in the beamwriter can be calculated from Equation 5.1.2 to be 189nm. From this and Equation 5.1.9 the focal range of 100keV electrons in the beamwriter is 2.4μm. For the 50kV situation, with a 400μm aperture, the f-number is halved so that the diffraction limited spot diameter is 6.81Å, the aberration diameter is 11.98nm, the Rayleigh range is 68.19nm so that the focal range is 1.2μm. Thus as with the approximations in §3.5.1, the focal range is longer at 100kV, but only because the aperture used is smaller.

From the preceding paragraph it can be seen that the Rayleigh range is very small compared to the height of the pyramidal tips, so that Equation 5.1.7 can be simplified further by considering the range where $x >> x_r$.

Equation 5.1.7 can be rewritten as:

$$d^2(x) = {d_{ab}}^2 + {d_o}^2 + {d_o}^2\left(\frac{x}{x_r}\right)^2 = {d_a}^2 + \left(\frac{x}{F}\right)^2, \qquad (5.1.10)$$

where Equations 5.1.3 and 5.1.6 have been used. This shows that the growth in spot size is independent of the electron wavelength, if the minimum achievable diameter is aberration limited. The main parameter in Equation 5.1.10 is the f-number which depends on the aperture used. It is also clear from this that for large values of x, where the minimum achievable diameter can be neglected, the spot size grows linearly with position. Therefore in the limit of large x:

$$\text{actual spot size} = \text{spot size} + \frac{\text{focus error} \times \text{aperture size}}{40}, \qquad (5.1.11)$$

where the focus error and aperture size are entered in units of microns and the spot size and actual spot size are in units of nanometres.

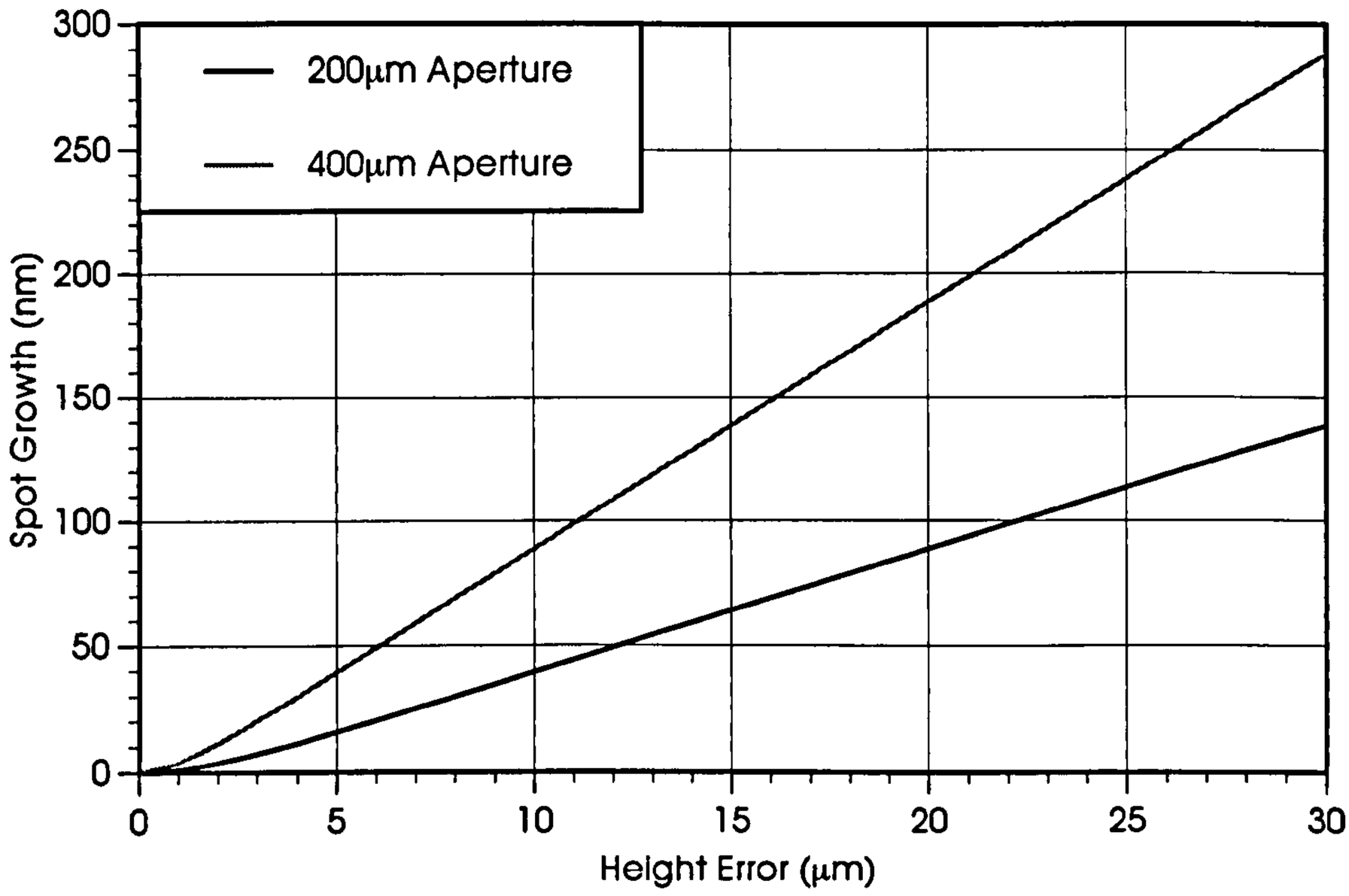


Figure 5.1.1: The growth in the actual spot size as a function of focus error for both 200μm and 400μm apertures.

Figure 5.1.1 plots the spot growth (as calculated from Equation 5.1.7) as a function of tip height for both the 200μm and 400μm aperture case. Since the spot size growth is an additive correction independent of any demagnification used to produce larger spot diameters, it is clear that the percentage growth of spot decreases with increasing spot size. It is also clear that if sensors are to be defined with physical dimensions below 100nm, and with tip heights of around 20μm, a defocusing mechanism is essential.

As mentioned in §3.5.1, the beamwriter normally automatically measures the height of the substrate, so that it is necessary to force the beamwriter to not readjust the focal point of the beam onto the substrate, but to add an offset to its measured position of the

substrate. For samples with step edges or mesas, the beamwriter provides a method for doing this:

(a) The beamwriter uses its laser as normal to measure the height of the substrate near the area to be exposed, and focuses the spot accordingly.

(b) The automatic height calibration is then turned off with the command 'set hgtinv 0'. This prevents the beamwriter from re-measuring the substrate height during writing and re-focusing the spot.

(c) The height of the pyramidal tips (as measured with the optical microscope) is put in as a parameter and is added to the measured height.

(d) The spot is now defocused using the 'adj height comp' command to accommodate this new height.

(e) After exposing the cell (or 'matrix field') the automatic height calibration is turned on with the command 'set hgtinv 1'.

This would have been a good solution, however, when this method was employed severe misalignment problems occurred with patterns being misplaced by as much as 10μm. This size of error was orders of magnitude larger than the beamwriter specification [Philips[4] (1980-1993)] and so an investigation was warranted. To confirm that the problem was related to the defocusing procedure a sample consisting of a 5 x 5 array of cells was processed through the beamwriter. The pattern used consisted of two identical matrices, one with defocus on and one with defocus off. Lift-off was carried out on these samples to indicate the placement of patterns with respect to the pyramidal tips. [This lift-off failed due to problems with coating resist over the pyramidal tips as discussed in §5.2.] Since the same set of markers were used for both patterns, marker problems could not explain the apparent misalignment of one pattern with respect to the other. By measuring the misalignment, a correlation was established between this and their position in the matrix.

Figure 5.1.2a shows position (x, y) = (5, 1). It is clear that the misalignment is only in the y-direction while Figure 5.1.2b shows position (1, 5) where the misalignment is only in the x-direction. Figure 5.1.3 shows position (3, 3) where the misalignment is in both directions. From these and additional results it was concluded that the misalignment was due to a rotation error. An approximate calculation suggested that the centre of this rotation was the bottom left marker of the first cell and the magnitude of the error seemed to correlate well with the value the beamwriter measured as the global wafer alignment (the angle between the wafer x- and y-axes and the beamwriter axes). This suggested that the beamwriter defocusing procedure in some way negated the automatic wafer rotation corrections.

Leica-Cambridge were contacted [Thoms[5] (1993-1996)] and they suggested replacing the 'adj height comp' command with the 'adj ff' (adjust fine focus) command [Philips[6] (1980-1993)] as they suspected that the former switched off certain corrections. However, this appeared to have no noticeable effect. The problem seemed to be related to the

'set hgtinv 0' command, but it would not be possible to leave this out as the beamwriter would then automatically negate the effects of any defocus during recalibration. Unfortunately it was unlikely that Leica-Cambridge would be able to offer a quick solution as the part of the code related to registration was written by a now deceased employee and was not well understood by their engineers.

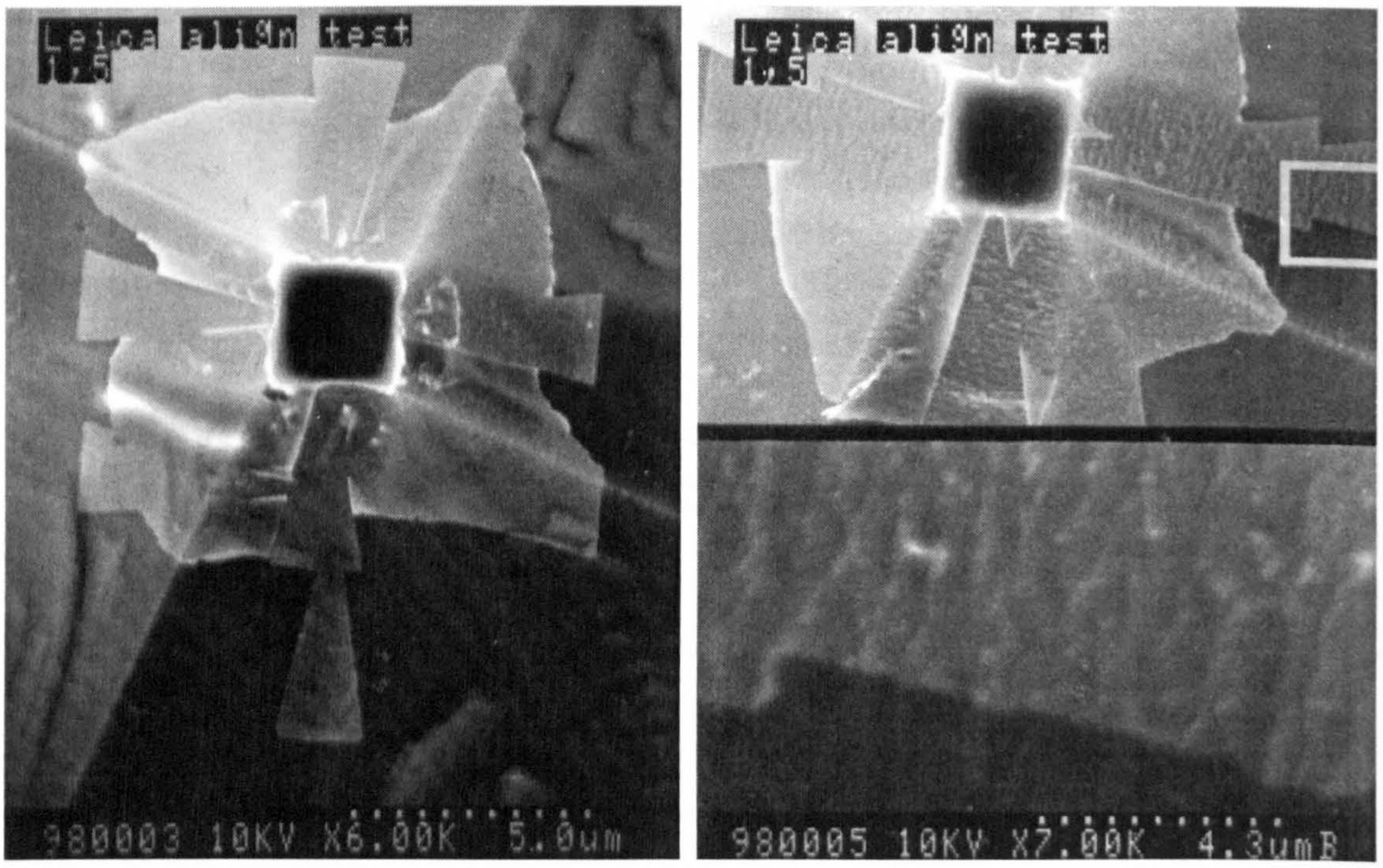


Figure 5.1.3: Micrographs showing rotational misalignment by the beamwriter with automatic height calibration switched off. (a) Left Image: this cell is at the bottom right of the sample and shows a large vertical misalignment; (b) Right Image: this cell is at the top left of the sample and shows a large horizontal misalignment. The bottom half of the right image is a 5x magnification over the displayed scale which applies to the top half.

It thus appeared that the only solution was to write software in-house to perform alignment and pattern exposure when defocusing is required. This was not perceived to be difficult as the beamwriter already measured the corrections required and it was simply a case of knowing which routines to call to apply these corrections. The beamwriter normally uses the 'expose matrix_field' command to expose the cell of a matrix in a registration job, presumably applying all the necessary corrections as given by the 'adjust wafer alignment' and 'adjust wafer distortion' commands. In the case of patterns not requiring registration there are no corrections for wafer distortions or alignment (as there are no wafer markers with which such corrections can be determined) and here the 'expose pattern' command is normally used after moving to the position corresponding to where the centre of the pattern is to be placed. This provides an obvious route for the correcting software: a custom version of the 'expose matrix_field' command which implements the corrections necessary, moves to the position corresponding to the centre of the pattern and then uses the 'expose pattern' command to write the pattern.

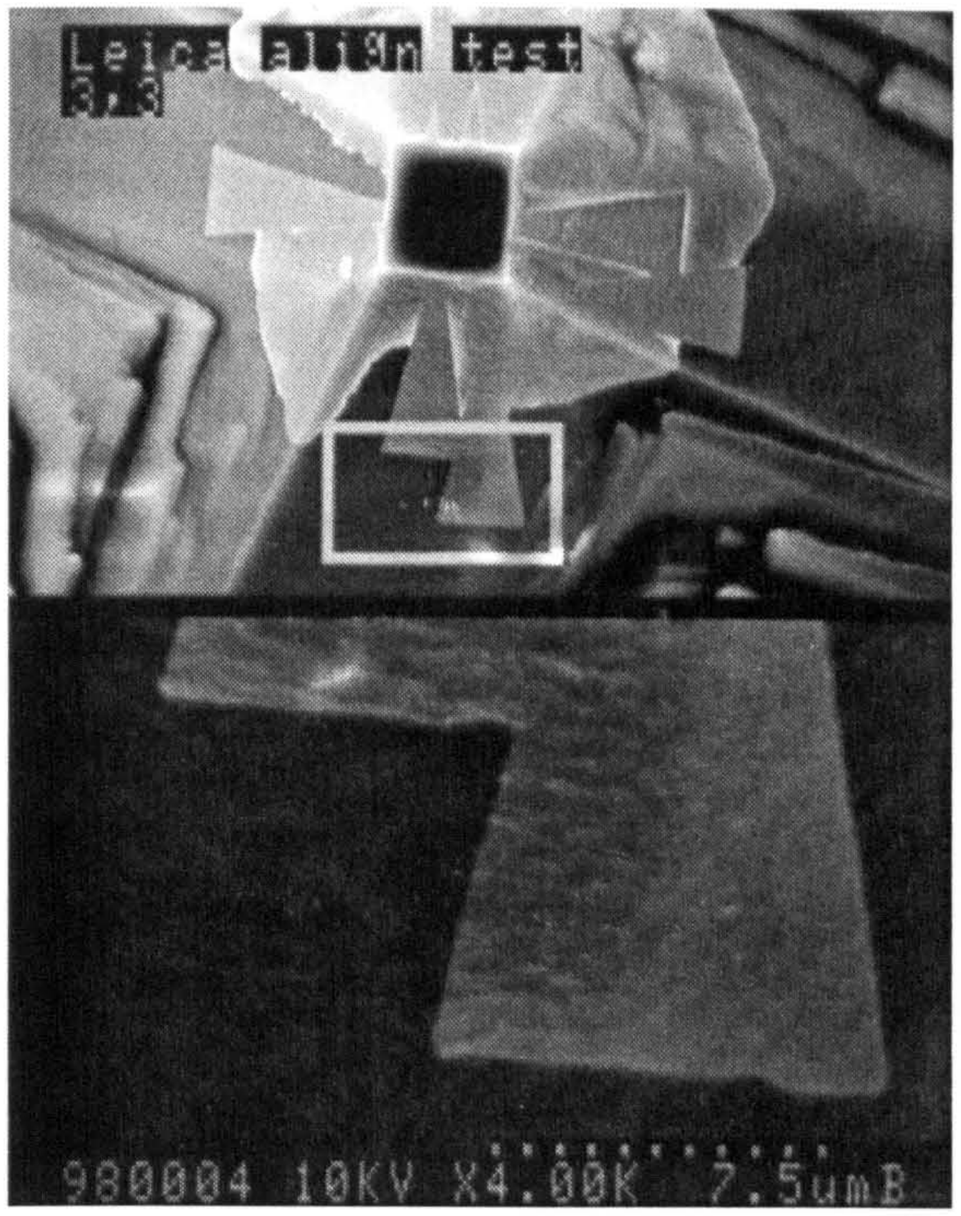


Figure 5.1.2: Micrograph showing rotational misalignment by the beamwriter with automatic height calibration switched off. This cell is in the centre of the sample and shows a large vertical and horizontal misalignment. The misalignment is less than that in Figure 5.1.2 but is in both directions. The bottom half of the image is a 5x magnification over the displayed scale which applies to the top half.

With a lot of help and advice [Thoms, *et al.*[7] (1988)] some software was written by the author by making the assumption that the 'adjust wafer alignment' and 'adjust wafer distortion' commands automatically adjusted the electron-beam scan coils for gain and rotation. Then all that remained was to calculate the best estimate of the centre position of the pattern with respect to the marker positions as measured by the beamwriter, move to that position, adjust the table compensation settings to take account of the distortions when moving from one block to the next during the 'expose pattern' command and call this command to write the pattern.

While this software was being written, a parameter that was set-up by the beamwriter installation engineers during its installation was found [Thoms, *et al.*[7] (1988)]. This parameter, which was never changed thereafter, had the symbolic name 'fl'. It was apparently related to the vertical position of the wafer table with respect to the focusing lenses and was meant to be used as a focus offset parameter. This then presented a means for using a very low level setting to offset the focusing, *i.e.* achieve defocusing, without the rest of the software interfering. An experiment was carried out to calibrate the effects of altering this parameter, with the beamwriter being made to measure the height before and after altering this parameter by a particular amount [Thoms, *et al.*[7] (1988)]. The conclusion was that adding a value corresponding to half the height of the defocus required (in microns) would work as intended. This appeared to cure the problem [see Chapter 6] and so further software development was ceased. [The factor of a half in the focus correction is due to the way the beamwriter height meter has been set-up to

allow a wider height range of samples to be run at one time. By reading out half the actual height value measured, samples over double the normal range of thicknesses can be run.]

5.2 COATING RESIST ON AFM TIPS

Initial attempts at coating surfaces with significant topography, albeit not one specifically suited for developing SPM cantilevers, had been encouraging [§3.7.1] and even though the first attempts to coat resist on a surface more akin to those suited for developing SPM cantilevers had not proved to be successful [§3.7.4], the results had not signalled the presence of an insurmountable technical hurdle. However, once the technology for fabricating pyramidal tips with flat tops had been developed [see §4.3.1], all attempts to coat these tips using the resist coating techniques of these earlier tests failed without exception. This can also be seen from the micrographs of §5.1 where lift-off seemed to fail at the apex of all pyramidal tips.

As it was known from the earlier tests that the resist became thinner as it progressed up the sides of the pyramids [§3.7.4], one strategy was based on the possibility that the resist had simply become so thin near the apex so as to be discontinuous. This strategy consisted of trying to increase the thickness of the resist. Since the resist thickness decreases with increasing spin speed and resist concentration [see Table 3.5.1 and Kuan, *et al.*[8] (1988)], these parameters became the main focus of the investigation along with the post-deposition bake temperature. This last parameter could affect the resist coating if the temperature was sufficiently high to allow the resist layer to creep away from the apex of the pyramids during the post-deposition bake [Hatzakis[9] (1988)].

A test was devised which consisted of coating the substrates with the resist, baking as required, evaporating a 20nm thick nichrome layer onto the resist and performing lift-off (without exposing the resist). Then any areas where the resist had not been of sufficient thickness were immediately revealed by the presence of a nichrome coating. These areas were easily distinguishable from the bare substrate using a scanning electron microscope. Tests were performed with 15% BDH and 4% Elvacite covering a range of spin speeds from 300 rpm to 5000 rpm [see Table 3.5.1 for thicknesses at 5000 rpm]. Multiple layers, bakes as short as ten minutes and temperatures of 80°C and 120°C were used, but in all cases the resist failed to coat the apex of the tip. The closest to complete lift-off achieved through spinning is illustrated in Figure 5.2.1 where the resist was so thick (estimated at 20μm on the flat areas of the substrate) and the spin speed so slow that it is even possible to determine the direction of spinning from this micrograph. Even had this succeeded it was clear that the majority of the substrate would be covered in a resist layer too thick to pattern properly due to the effects of charging [Liu, Ingino and Pease[10] (1995)] and resist collapse [McCord, Wagner and Donohue[11] (1993)].

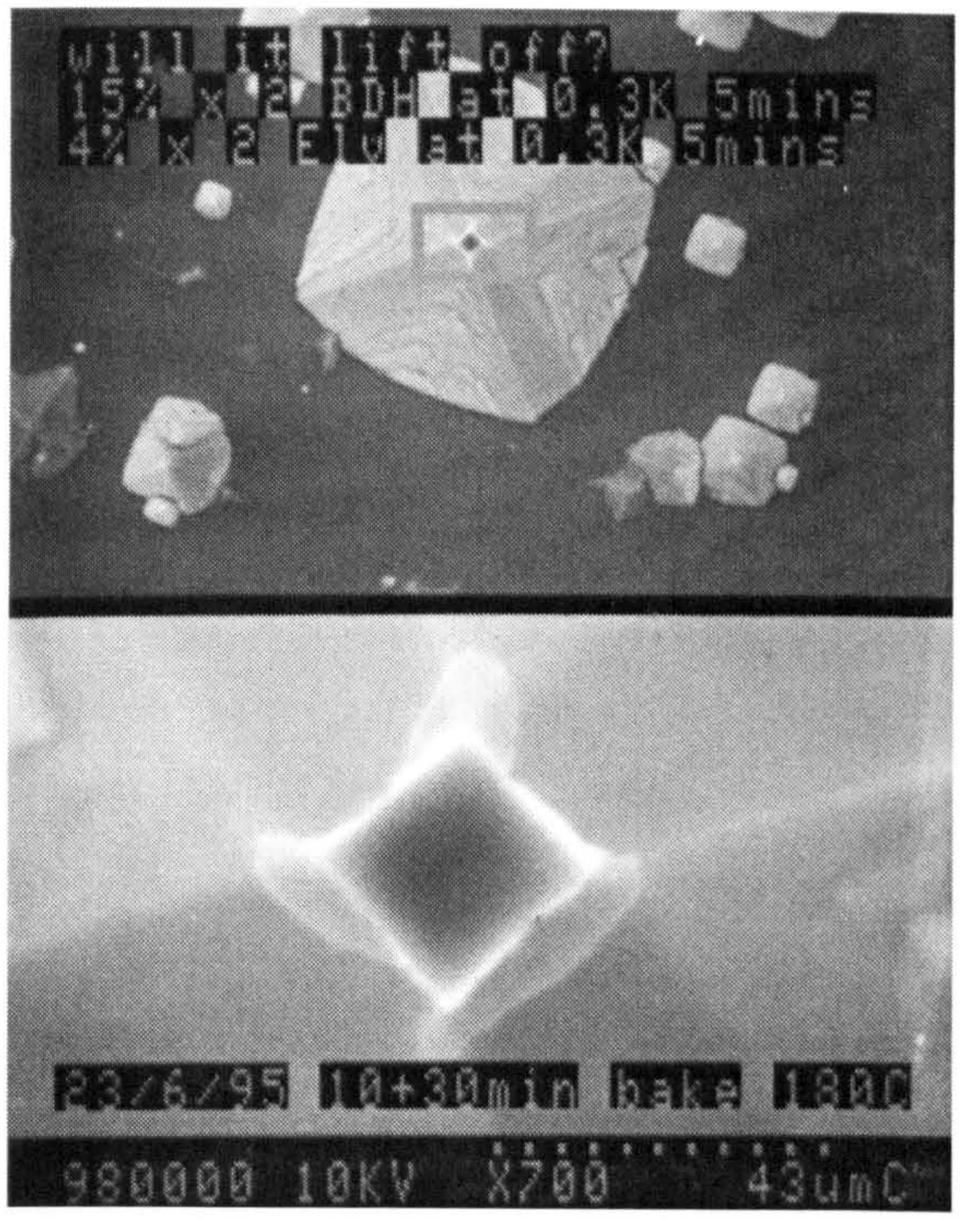


Figure 5.2.1: Micrograph showing the closest to a successful lift-off on a pyramidal tip achieved by spin-coating resist. The spin speed is so slow, at 300 rpm, that the direction of spinning can be determined from the micrograph. The resist is a bi-layer of 15% BDH followed by a bi-layer of 4% Elvacite. The bottom half of the image is a 10x magnification over the displayed scale which applies to the top half.

Another method of resist coating which hailed from over a decade previously and was devised when it was thought that spin coating would not provide resist layers of a uniformity adequate for high resolution electron-beam lithography, was suggested as a solution to the resist coating problem [Gourlay[12] (1993-1996)]. In this technique, known as 'dip coating', the substrate to be coated is mounted on a glass slide and attached vertically to the driveshaft of a motor drive. The substrate would then be submerged in the resist solution and the motor used to slowly and uniformly draw the substrate out of the resist solution, allowing the solvent to drain off and leaving a resist coating on the substrate. Speeds of 0.1cm/min to 0.5cm/min had been used with 1% BDH and 2% BDH. This had resulted in resist coatings ranging from 11nm to 45nm in thickness. The general relationship between speed and resist thickness was that the faster the wafer was pulled out of the resist, the thicker the resulting resist coating (up to a point). This technique was attempted in a crude manner using a motor speed of 1cm/min and a range of resist concentrations culminating in 15% BDH, but without success.

From these attempts it became clear that the problem was fundamentally due to the deposition of the resist in liquid form. As the resist coats an asperity, surface tension acts so as to minimise the resulting radius of curvature of the liquid surface. Minimum energy considerations then mean that close to zero thickness of resist is expected at sharp edges, as illustrated in Figure 5.2.2. This fact has even been used to advantage in the fabrication of thermal probes based around an STM tip [Oesterschulze, *et al.*[13] (1996)].

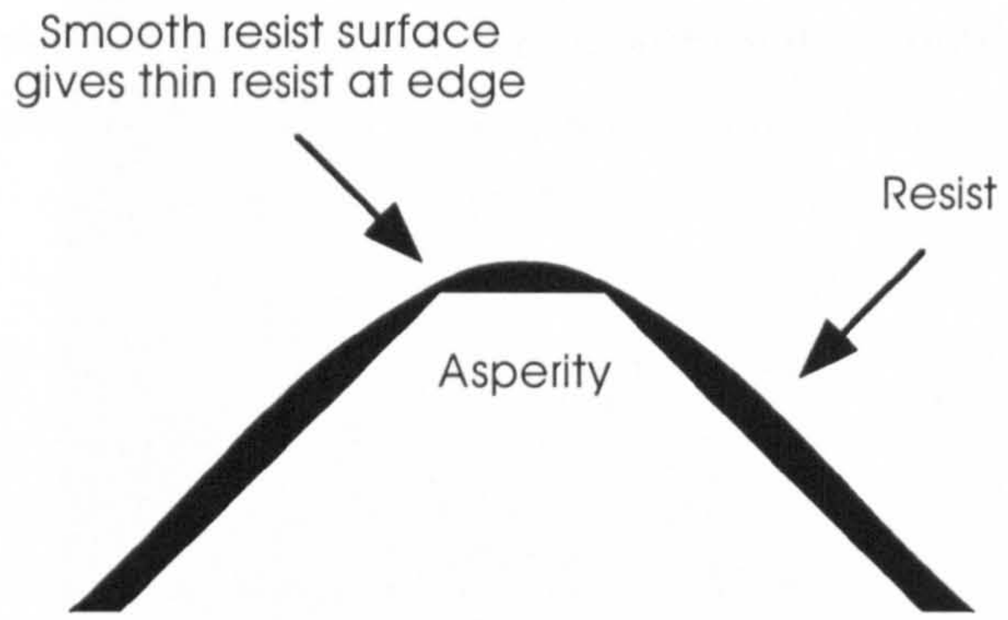


Figure 5.2.2: Diagram illustrating the problem with coating resist over sharp vertices.

A solution was suggested by Dr. J. M. R. Weaver which involved solidifying the resist prior to coating so that no flow can occur. This was done by dropping some dissolved resist into a water-filled dish. This droplet spreads out over the surface of the water, forming a thin layer and as the solvent dries a PMMA skin is formed on the surface of the water. This skin is deposited onto the substrate surface by draining the water beneath it, having previously placed the substrate at the bottom of the dish. This method, dubbed 'float coating' proved to be extremely successful as illustrated by Figure 5.2.3a. Here a single drop of 8% BDH was dropped onto a 6 inch diameter petri dish of warm water using a pipette. The water was drained off using the pipette and the PMMA allowed to drape over a sample. After a short bake at 80°C a grating was exposed over the tip and developed in IPA:MIBK 2.5:1 for 30 seconds. The resist thickness was estimated as being around 0.5μm.

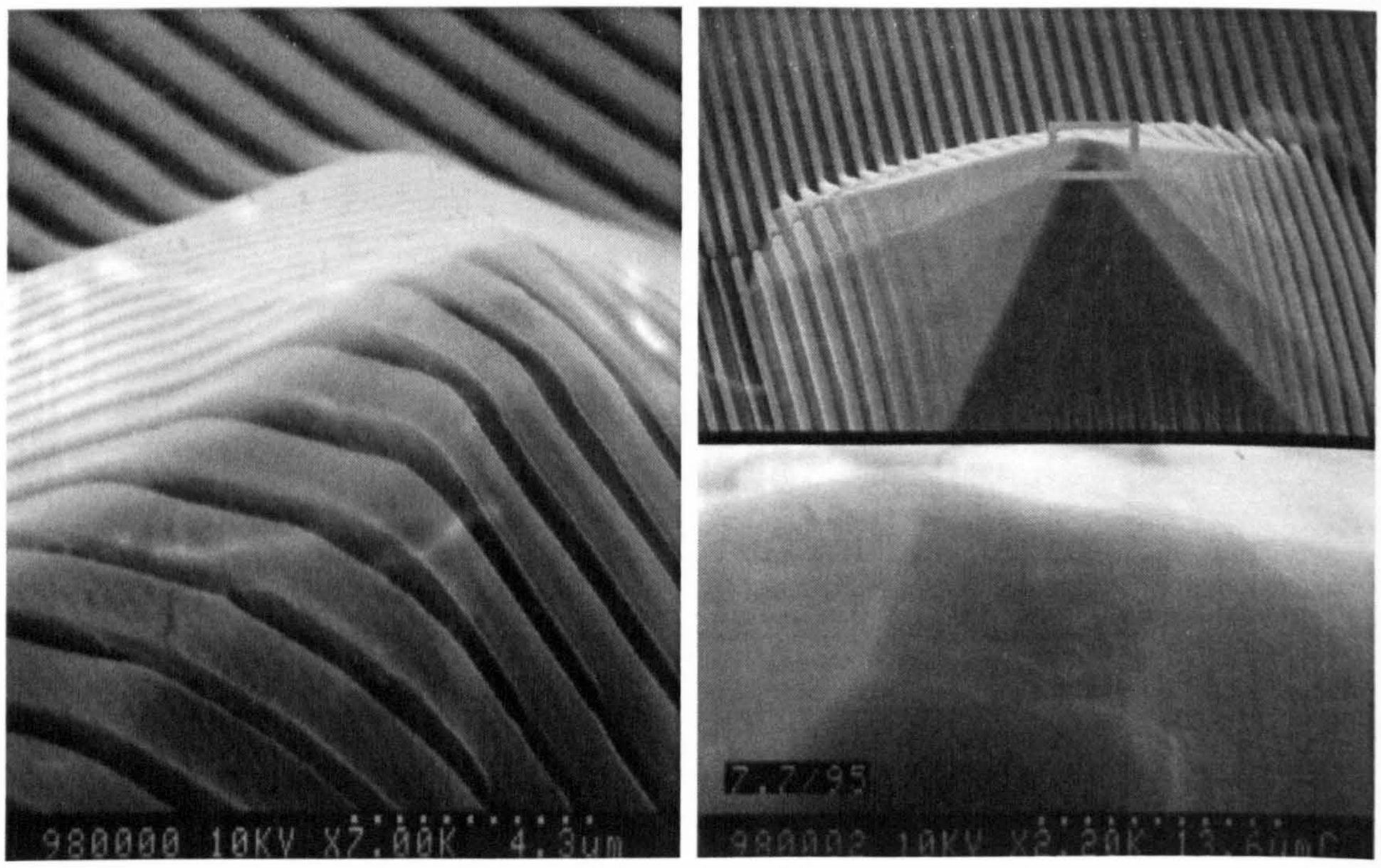


Figure 5.2.3: Micrographs illustrating attempts at coating resist on pyramidal tips by the technique of 'float coating'. (a) Left Image: the first successfully 'float-coated' tip; (b) Right Image: the failure of a variation of the 'float coating' technique. The bottom half of the right image is a 10x magnification over the displayed scale which applies to the top half.

A variation in this technique, where the skin is allowed to form as before, but instead of draining the water beneath to drape the PMMA over the substrate, the substrate is inverted and gently dropped onto the PMMA skin was attempted. The results of this is illustrated in Figure 5.2.3b. This shows that the tips are sharp enough to pierce the resist layer, leaving a region around the apex uncoated.

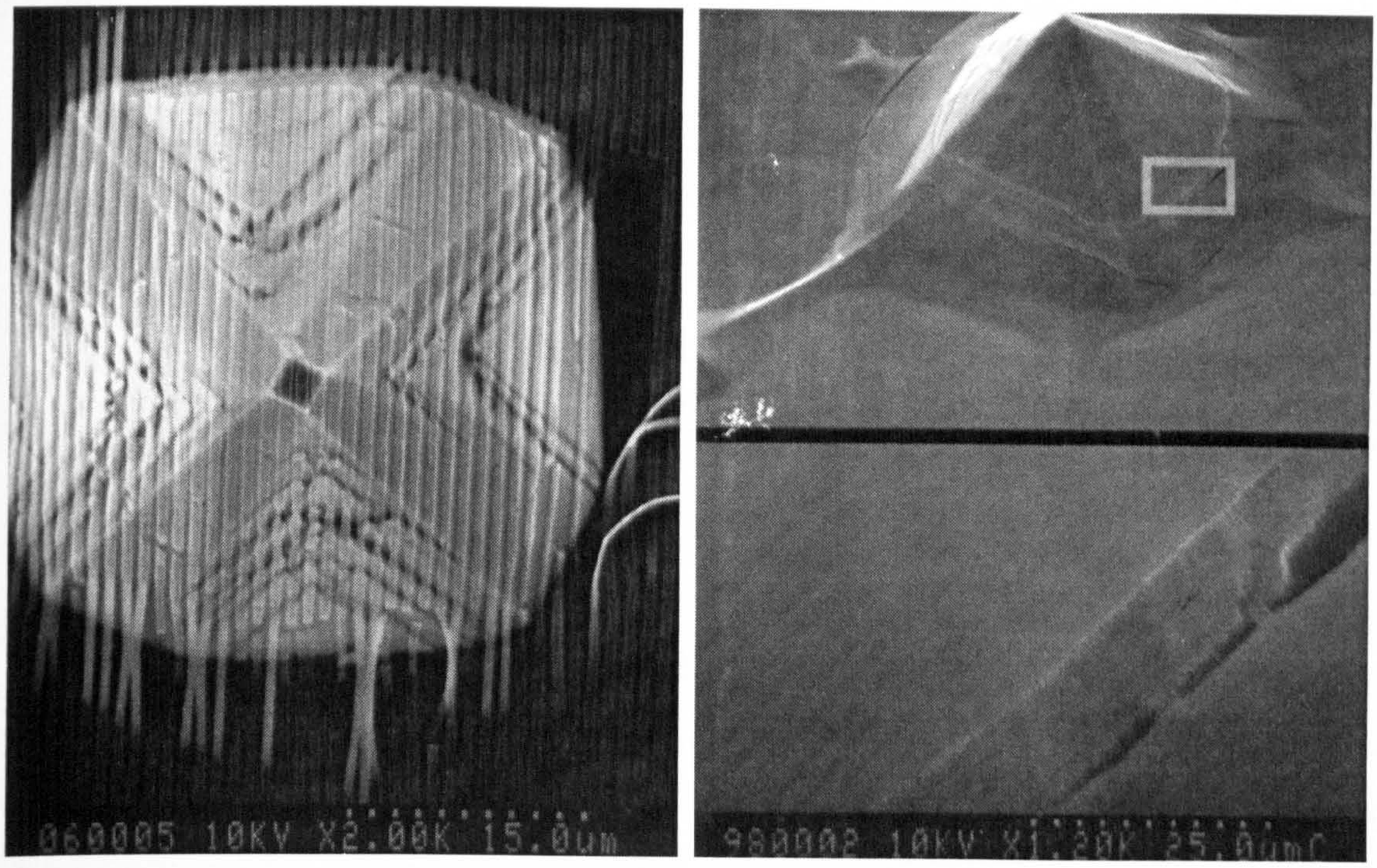


Figure 5.2.4: Micrographs showing some of the problems with 'float coated' resist layers. (a) Left Image: the resist does not properly make contact with the substrate at convex vertices; (b) Right Image: the resist is under great tension and can sometimes tear. The bottom half of the right image is a 10x magnification over the displayed scale which applies to the top half.

Several experiments were carried out to characterise this new method of resist coating. Extensive quantitative analysis of the process was not carried out as the volume of a single drop (estimated to be around $4mm^3$) was poorly controlled. Figure 5.2.4a shows a bi-layer deposited using this technique, and from this it is clear that the resist does not make good contact with the substrate at the foothills of the tip and also shows signs of stress. The stress can occasionally be large enough to result in tearing of the resist as illustrated by Figure 5.2.4b.

The effect of baking temperature was also briefly assessed by baking one sample, coated with 8% BDH, at 180°C. The result, after exposing a grating, can be seen in Figure 5.2.5a. The resist appears to have thinned near the sharp edges highlighted in Figure 5.2.2. This supports the hypothesis that surface tension was a major factor in the failure of the previously attempted coating techniques.

It was observed that surface tension plays a crucial role in the formation of the PMMA skin layer. This has its biggest effect through a quantity known as the spreading coefficient [Gray and Isaacs[14] (1975)]. This is illustrated in Figure 5.2.6. A drop of liquid,

A, resides on the surface of another liquid, B. The surface tension forces for each liquid, $\mathbf{T_A}$ and $\mathbf{T_B}$ are shown acting at one edge of drop A. Also shown is the interfacial surface tension, $\mathbf{T_{AB}}$, which acts between the surfaces of the two liquids. The drop will spread if $\mathbf{T_B > T_A + T_{AB}}$. The parameter, $T_B - T_A - T_{AB}$, is known as the spreading coefficient and is positive for liquids which will spread and negative for those that will not. The analogue to this for liquids on solid surfaces, is the contact angle, also illustrated in Figure 5.2.6 [Bolton[15] (1986)].

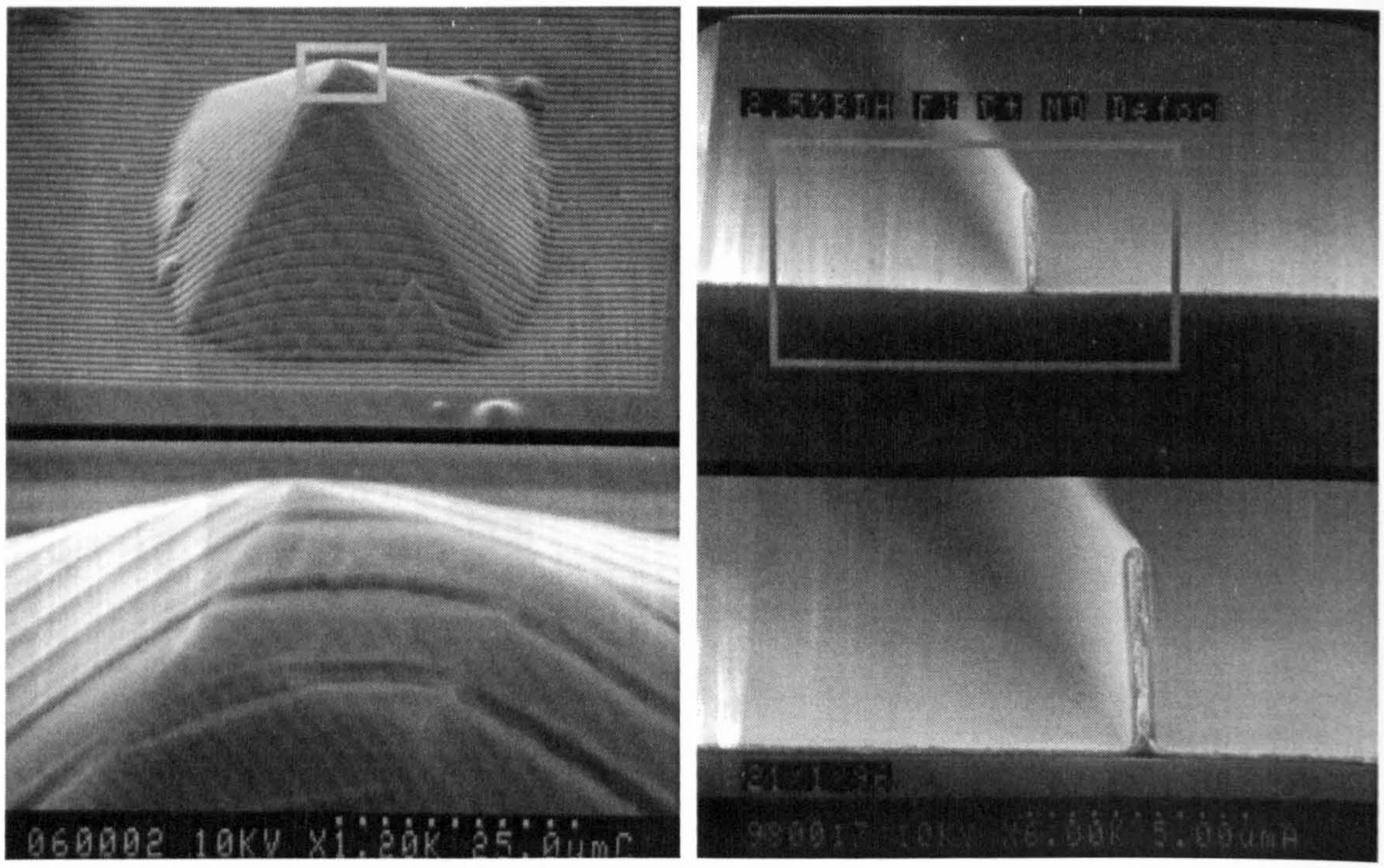


Figure 5.2.5: Micrographs illustrating some aspects of 'float coated' resist layers. (a) Left Image: if the 'float coated' resist is baked at too high a temperature, it creeps or flows away from sharp convex vertices. The bottom image is a 10x magnification over the displayed scale; (b) Right Image: the 'float coated' layer contains many resist wrinkles. The bottom image is a 2x magnification over the displayed scale.

The following are the remaining observations that were made about the 'float coating' technique:

(a) The thickness of the resist is very sensitive to the temperature of the water, with colder water resulting in thinner layers due to a reduced rate of solvent evaporation, and possibly, increased spreading coefficient.

(b) The thickness of the resist depends on the resist concentration with higher concentrations producing thicker layers due to increased viscosity and increased total volume of polymer.

(c) The thickness of the resist varies with molecular weight with higher molecular weight resist giving thicker films due to increased viscosity.

(d) The resist dries from the outside to the centre of the dish and tends to become wrinkled as the resist dries and shrinks, Figure 5.2.5b illustrates a wrinkle as viewed on a cleaved edge – these wrinkles usually fall over resulting in a thickness three times the nominal value.

(e) The resist is thicker in the centre of the dish than at the edges.

(f) The resist never touches the edges of the dish (probably caused by the non-zero contact angle between water and glass and its effect on the spreading coefficient) and therefore the thickness depends somewhat on the size of the dish.

(g) Resist dissolved in chlorobenzene cannot be used as chlorobenzene is denser than water. Resist dissolved in orthoxylene was used exclusively.

(h) The addition of detergent to the water prevents spreading of the resist due to reduced water surface tension giving a negative spreading coefficient.

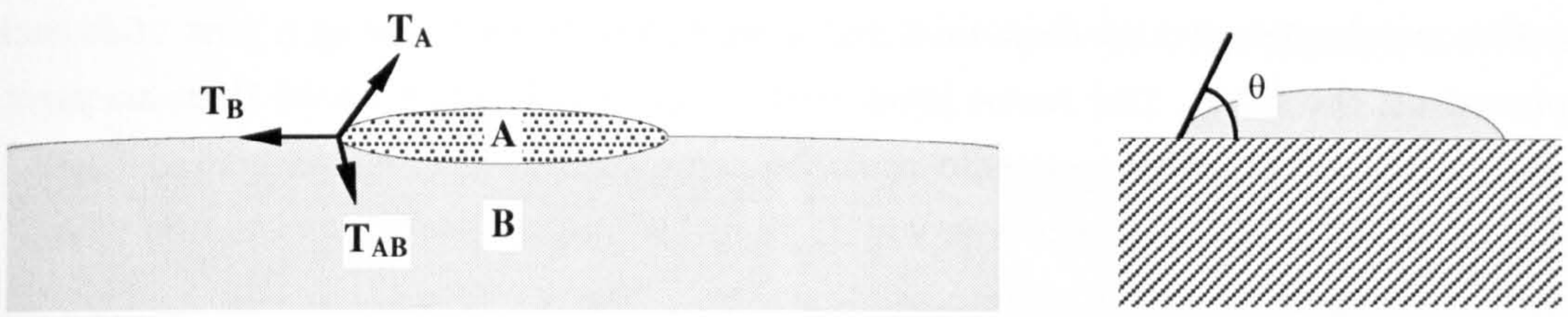


Figure 5.2.6: Diagram illustrating what is meant by the spreading coefficient and the contact angle.

These observations posed some interesting problems regarding the application of this technique for the purposes of defining sensors at the apices of these pyramids. The first concerned reproducibility. Due to the small size of the sensors required, more than a factor two change in thickness would probably be unacceptable, although the main consequence of such a thickness variation would only be that the minimum sensor size would be somewhat larger than desired. Provided that the thickness variations did not result in underexposure of the resist, this was not perceived to be a significant problem from the point of view of initial development. Resist thickness measurements for a typical run are shown in Table 5.2.1.

Resist	Float-Coating Thickness	5000 rpm Spin-Coating Thickness
8% BDH	984 nm	208 nm
4% BDH	460 nm	69 nm
2.5% BDH	156 nm	24 nm
2.5% Elvacite	160 nm	40 nm

Table 5.2.1: Some thickness measurements of 'float-coated' PMMA layers. The numbers in the right column are extracted from Table 3.5.1 for comparison.

The next concern was the tearing of the resist close to the foothills of the pyramidal tips. This was overcome by simply spin coating a resist layer prior to float coating. The spin coated layer would only coat a short way up the sides of the pyramidal tip and would leave the all-critical apex uncoated, thereby coating the region most at risk to being exposed by any tears in the float coated layer. The structures at the base of the tip could

be made much larger so that the increased thickness of the resist on the flat part of the substrate, due to both spin coated and float coated layers being present, would not be a problem.

Another concern was the method by which the fragile resist coating could be transported to the oven without damage. A method was devised whereby the substrate (one inch squares) was placed in a concave watch glass which in turn was placed on the bottom of a 6 inch diameter petri dish. Water was filled so as to submerge both the substrate and the watch glass. Then float coating was carried out as before. The water was initially drained using a large syringe until the resist layer was caught on the edges of the watch glass. The watch glass could then be lifted from the petri dish using a pair of tweezers and placed on the table. The resist layer within the watch glass could then be pierced with a pipette or hypodermic needle and the remaining water drained away until the resist layer coated the substrate. The watch glass and substrate could then be placed in the oven and baked to remove all remaining water.

Yet another concern related to the presence of wrinkles within the resist layer. However, these wrinkles were relatively sparse and were unlikely to pass through the apex of a pyramidal tip, since the resist tended to be stretched over this region. On the flat substrate, the patterns required could be made sufficiently large and doses made sufficiently high that any such wrinkles in the path of the pattern would be exposed, partly by the increased dose and partly by the proximity effect from a large exposure area. In addition, the use of a concave watch glass meant that the resist layer was stretched slightly as it draped the substrate, thereby acting so as to reduce the number and size of these wrinkles.

The final concern was the variation in resist thickness between the centre of the dish and the edges of the dish. However, it appeared that the central region was sufficiently uniform to accommodate a one inch substrate when a six inch diameter petri dish was used. This could be seen by observing the density of thin film interference fringes within the PMMA layer. However, it was also clear that accommodating whole three inch wafers with this technique might be more problematic.

One line of thinking suggested that if the resist was diluted to the point where there was insufficient PMMA within a single drop to form a continuous single monolayer, then the drying process would not begin until the resist had uniformly spread over the entire water surface. A calculation based on assuming that each PMMA polymer chain could be approximated by a 4nm diameter sphere [Chang, *et al.*[16] (1988)] suggested that resist concentrations of the order of 0.1% would be required. A quick test showed that this was indeed the case. To make up the required volume, a macroscopic quantity of resist solution could then be poured onto the surface of the water and stirred before the drying process began. This immediately removes the irreproducibility associated with the use of

'single drops' and the dependency on the water temperature by eliminating the rate of spreading as a factor in the process.

Using this method it was found that the drying process could take between 30 minutes and one hour. However the drying process was not uniform and instead of resulting in a set of wrinkles arranged with circular symmetry, patches of wrinkles (wrinkle domains) would appear, probably due to a combination of 'hotspots' being present on the water surface and the effects of air currents and eddies in transporting solvent-rich vapour away from the resist surface. However, as before this was not perceived as being a significant problem for the sparse patterns required for the work of this thesis. [Note that this method was not employed during the initial development which took place on one inch substrates.]

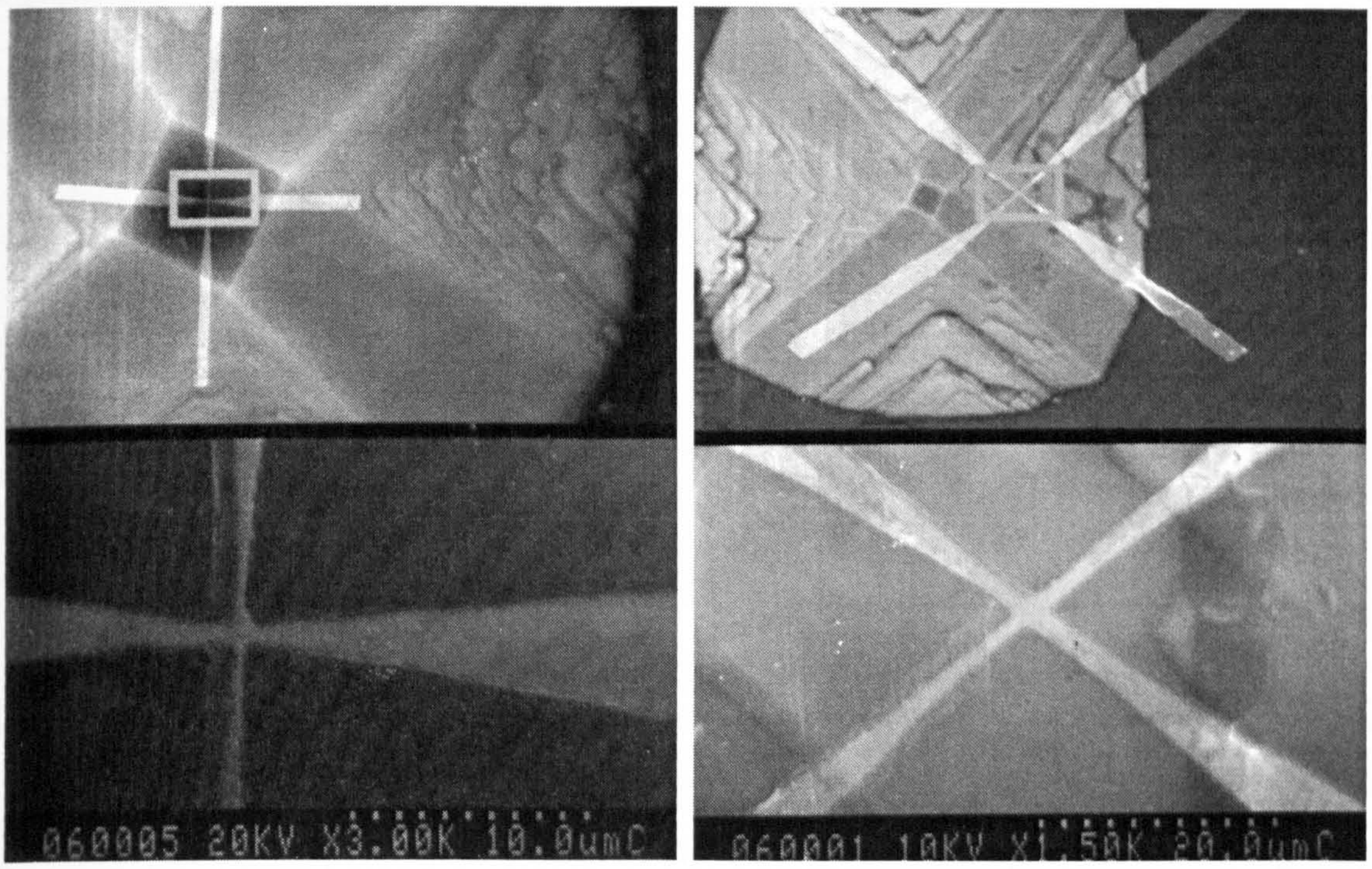


Figure 5.2.7: Micrographs showing the first successful demonstration of high resolution pattern transfer on pyramidal tips. Both images are of cross structures written by manual alignment. Lift-off was successful on both tips despite the error in aligning to the tip in the right micrograph. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

It was hypothesised by the author that if the resist could be made to dry from the centre outwards, then wrinkles would not appear. That this is reasonable can be seen by considering first what happens if the resist dries from the outside inwards, based on the assumption that as the resist dries, it shrinks. When a small region of resist at the outer edge of the dish has formed, it shrinks, attempting to pull more liquid resist outwards. However, the area of contact of the resist (with water) is significantly less than that of the remaining liquid, so that the dried resist is pulled inwards instead. [The contact area is relevant through the interfacial surface tension illustrated in Figure 5.2.6.] As this resist was located at a greater radial distance, its circumference is now too large for its current radial position so that wrinkles are formed to take up the excess resist. If instead the

resist is forced to dry outwards from the centre, the central region shrinks and tries to pull the liquid resist inwards. Again its area of contact is less than that of the liquid resist and so it fails to pull the resist inwards significantly. However, it equally cannot be pulled outwards due to the symmetric arrangement of acting forces and is therefore stretched, preventing the formation of wrinkles. This radial heating might be achievable by as simple a technique as locating a light bulb centrally over the petri dish. However, there was insufficient time to develop the process along these lines.

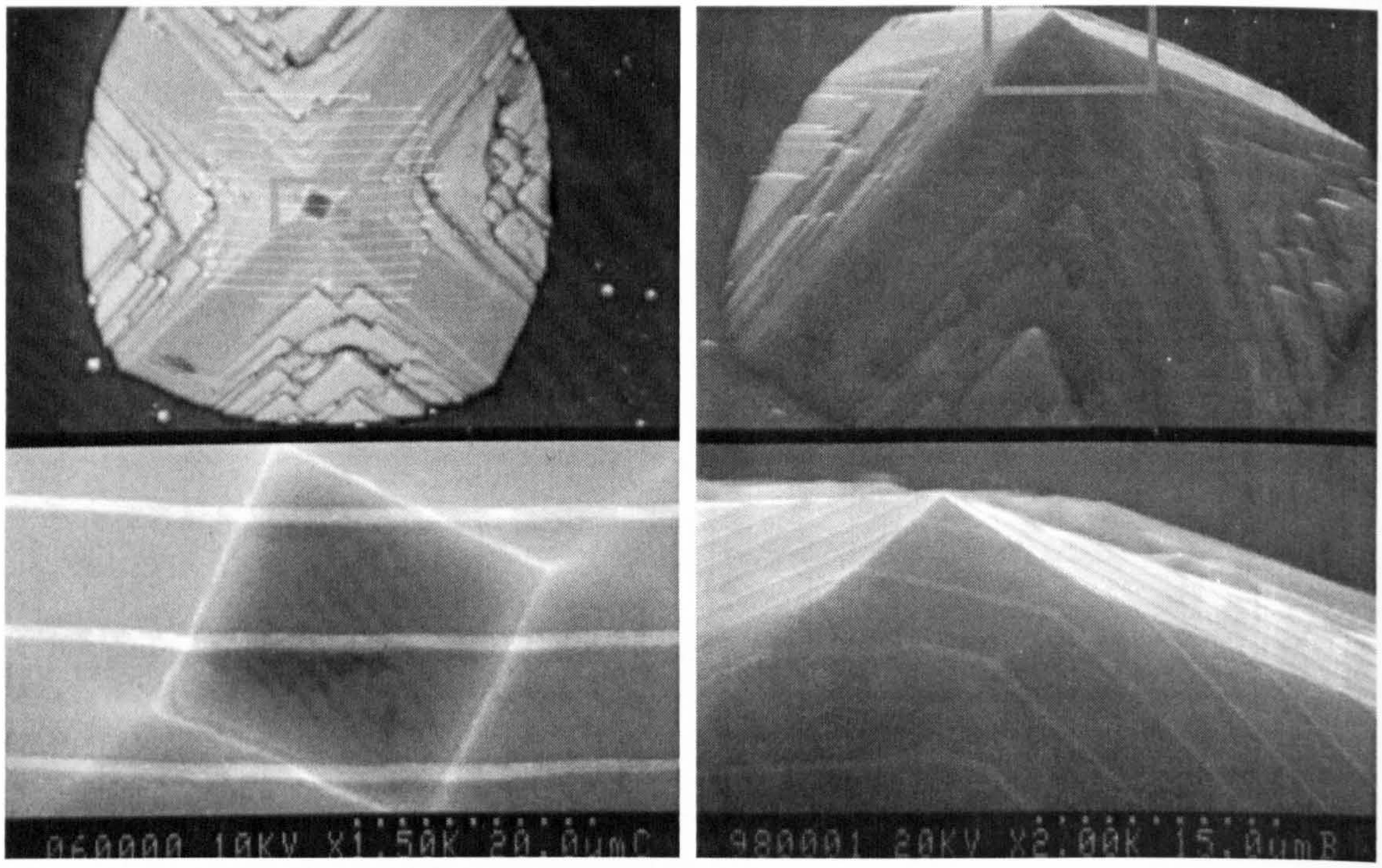


Figure 5.2.8: Micrographs showing the first successful demonstration of high resolution pattern transfer on pyramidal tips. Both images are of gratings written by manual alignment. The right image has been taken at an oblique angle to the plane of the substrate. The bottom half of the left image is a 10x magnification over the displayed scale, while the bottom half of the right image is a 5x magnification over the displayed scale.

Although a quantitative analysis of this drying process (as used in the macroscopic deposition method) was not carried out, a starting point might be the analysis of the formation of coffee stains [Deegan, *et al.*[17] (1997)]. The main difference here is that the edges of the spilt liquid (the resist) are not pinned by imperfections in the surface upon which it has been spilled (the water), because this surface is extremely close to being perfectly smooth. However, many of the processes taking place are likely to be the same or similar once this has been factored in. An analysis of the single drop method is likely to be much more complex due to requiring transient effects to be taken into account (it is a dynamic process).

That float coating would be capable of achieving features as small as required for the work of this thesis was demonstrated by Dr. J. M. R. Weaver using a converted scanning transmission electron microscope [Thoms, Beaumont and Wilkinson[18] (1989)], as the beamwriter was non-operational at this time for technical reasons. With this machine he

performed manual alignment to a number of pyramidal tips (in an SEM mode of operation) and succeeded in producing lifted off structures as illustrated in Figures 5.2.7 and 5.2.8. The resist used was a single layer of 8% BDH. Development was carried out using IPA:MIBK 2.5:1 for 30 seconds and the metallisation consisted of 5nm of titanium (sticky layer) and 20nm of gold.

5.3 SUMMARY AND CONCLUSIONS

A method was found to allow defocusing of the beamwriter during automatic alignment jobs to allow high resolution lithography to be carried out on surfaces with topography of the order of 20μm. A new technique for coating resist was developed to allow non-planar surfaces to be coated with PMMA for electron-beam lithography purposes.

5.4 REFERENCES

1. Newport, *GAUSSIAN BEAM OPTICS* in Newport Catalogue, 1989 edition, published by Newport, England, pp. N18-N21 (1993).

2. R. F. W. Pease, *ELECTRON BEAM LITHOGRAPHY*, Contemporary Physics **Vol. 22,** No. 3, pp. 265-290 (1981).

3. R. K. Watts, *Chapter 4: LITHOGRAPHY* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 141-183 (1988).

4. Philips, *THE PHILIPS BEAMWRITER*, published by Philips E-beam Lithography Division, Eindhoven, The Netherlands (1980-1993).

5. S. Thoms, *Private Communication*, (1993-1996).

6. Philips, *BEAMWRITER DOCUMENTATION*, published by Philips E-beam Lithography Division, Eindhoven, The Netherlands (1980-1993).

7. S. Thoms, I. McIntyre, S. P. Beaumont, M. Al-Mudares, R. Cheung and C. D. W. Wilkinson, *FABRICATION OF QUANTUM WIRES IN GAAS/ALGAAS HETEROLAYERS*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 1, pp. 127-130 (1988).

8. S. W. Kuan, C. W. Frank, C. C. Fu, D. R. Allee, P. Maccagno and R. F. W. Pease, *ULTRATHIN POLYMER FILMS FOR MICROLITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2274-2279 (1988).

9. M. Hatzakis, *MATERIALS AND PROCESSES FOR MICROSTRUCTURE FABRICATION*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 441-453 (1988).

10. W. Liu, J. Ingino and R. F. Pease, *RESIST CHARGING IN ELECTRON BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 5, pp. 1979-1983 (1995).

11. M. A. McCord, A. Wagner and T. Donohue, *RESOLUTION LIMITS AND PROCESS LATITUDE OF X-RAY MASK FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2958-2963 (1993).

12. D. Gourlay, *Private Communication*, (1993-1996).

13. E. Oesterschulze, M. Stopka, L. Ackermann, W. Scholz and S. Werner, *THERMAL IMAGING OF THIN FILMS BY SCANNING THERMAL MICROSCOPE*, Journal Of Vacuum Science & Technology **Vol. 14,** No. 2, pp. 832-837 (1996).

14. H. J. Gray and A. Isaacs, *A NEW DICTIONARY OF PHYSICS*, published by Longman Group Limited (1975).

15. W. Bolton, *Chapter 16: SOLIDS, LIQUIDS AND GASES* in Patterns in Physics, 2nd edition, published by McGraw-Hill Book Company, Maidenhead, pp. 325-366 (1986).

16. T. H. P. Chang, D. P. Kern, E. Kratschmer, K. Y. Lee, H. E. Luhn, M. A. McCord, S. A. Rishton and Y. Vladimirsky, *NANOSTRUCTURE TECHNOLOGY*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 462-493 (1988).

17. R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel and T. A. Witten, *CAPILLARY FLOW AS THE CAUSE OF RING STAINS FROM DRIED LIQUID DROPS*, Nature **Vol. 389,** pp. 827-829 (1997).

18. S. Thoms, S. P. Beaumont and C. D. W. Wilkinson, *A JEOL 100 CXII CONVERTED FOR USE AS AN ELECTRON-BEAM LITHOGRAPHY SYSTEM*, Journal Of Vacuum Science & Technology B **Vol. 7,** No. 6, pp. 1823-1826 (1989).

CHAPTER 6

SNOM PROBE FABRICATION

This chapter describes the attempts made to design, fabricate and model probes for scanning near-field optical microscopy. First some design issues specific to scanning near-field microscope probes are discussed. Then the problems encountered when trying to define apertures for these probes and the measures taken in resolving these problems are described in some detail. Finally, a simplistic light throughput model for the developed probe geometry is presented and used to make comparisons between these probes and more conventional optical fibre based SNOM probes.

6.1 CHOICE OF METALLIC SHIELDING

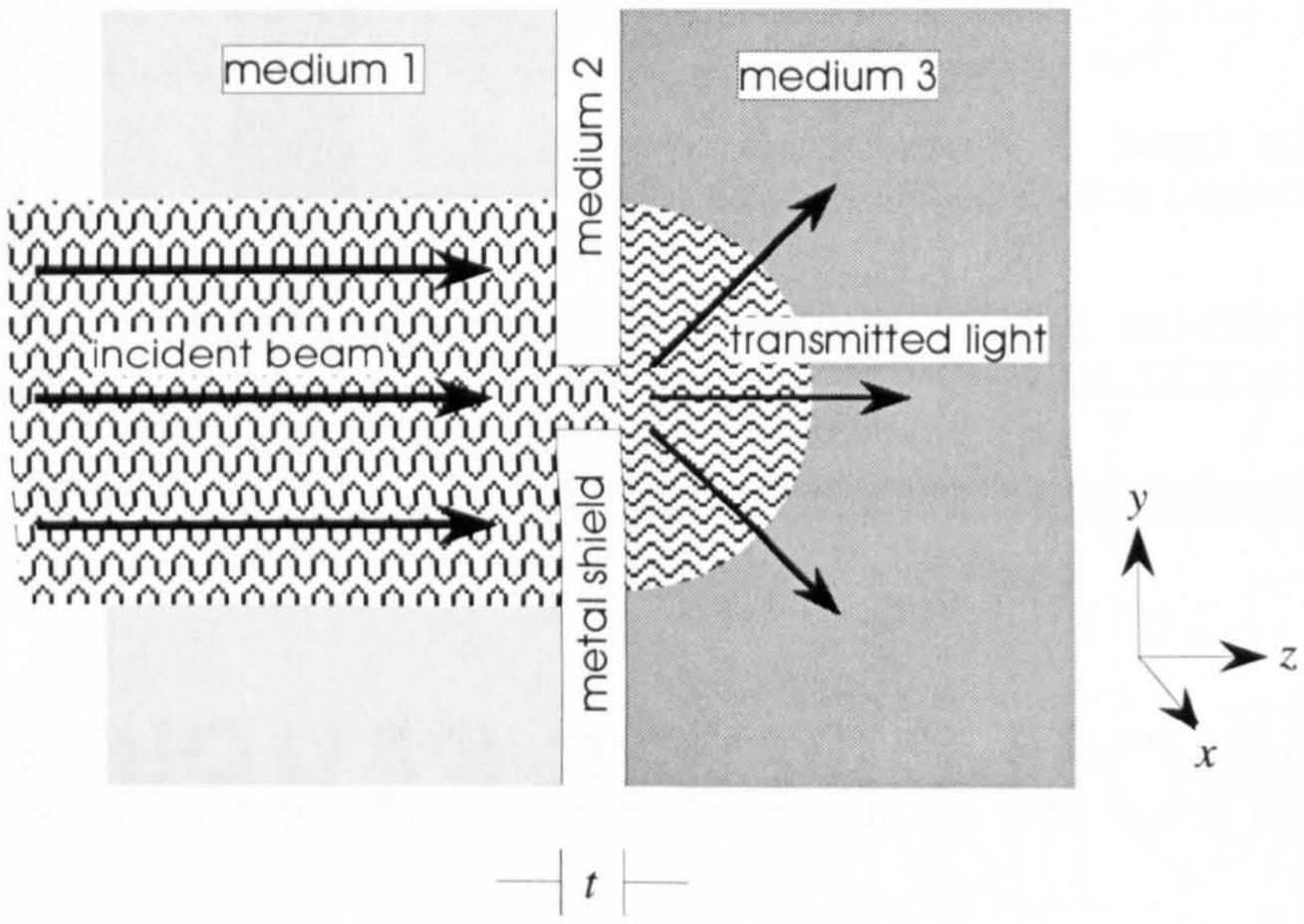


Figure 6.1.1: Diagram illustrating the function of the aperture in the shielding layer.

As mentioned in §3.2, the probe designed for use in SNOM applications consists of a contact mode AFM cantilever and pyramidal tip (as previously discussed in §4.4.2 and §4.3.1 respectively) which is coated in a metallic shielding layer in which a sub-wavelength aperture has been defined to be coincident with the apex of the tip. The choice of metal used for the shielding layer and its thickness are dictated by many considerations, both theoretical and practical. More theoretical considerations will be discussed in this section with more practical considerations being discussed in subsequent sections.

Laser	Wavelengths (nm)
Helium-Neon	632.8*
Argon Ion	488.0*, 496.5, 514.5*
Helium Cadmium	325.0, 442.0*

Table 6.1.1: The main emission lines from some lasers commonly used in SNOM. The strongest lines are marked with an asterisk, *.

As has been noted in the literature [*e.g.* Betzig, *et al.*[1] (1991)], the typical output intensity of light is six or more orders of magnitude down on the input light intensity for apertures of sub-100nm diameter. Therefore in order that an adequate signal-to-background ratio is obtained the shielding layer must have an opacity greater than six orders of magnitude. This can be achieved by adjusting the thickness of the shielding layer. In the context of the probe being designed here, the metallic shielding lies in a plane

perpendicular to the input light beam. The situation to be analysed is shown in Figure 6.1.1. In general, the choice of metal and its thickness will depend on the wavelength of operation. Table 6.1.1 lists some commonly used lasers and their emission wavelengths. The wavelengths marked with an asterisk are the strongest of these and only these will be considered here. [The reader is referred back to §1.4.1 for the definition of the variables and parameters used in the following analysis.]

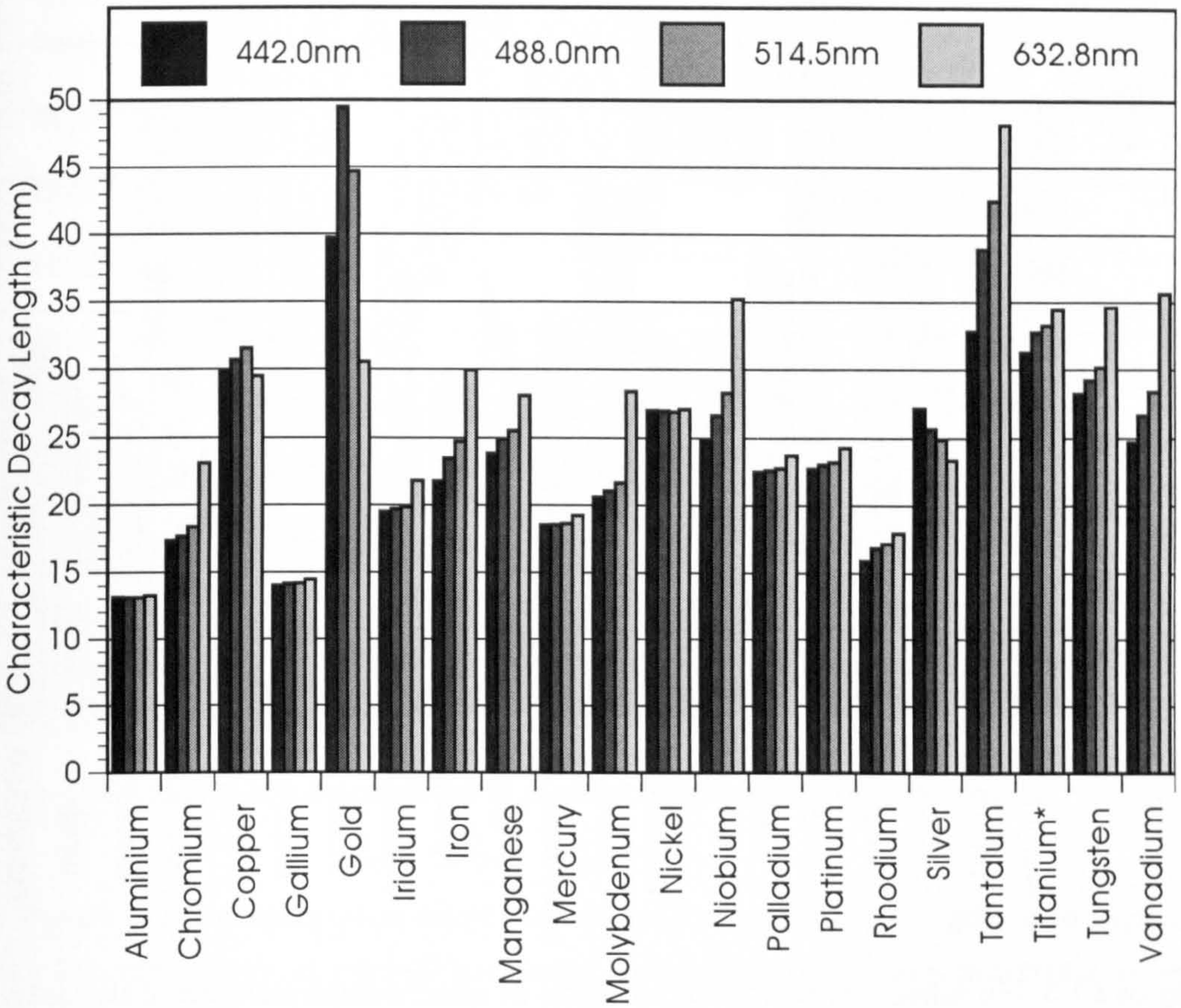


Figure 6.1.2: Characteristic (amplitude) decay lengths of various materials at selected wavelengths from Table 6.1.1. The intensity characteristic decay lengths are a factor of two shorter.

The generalised characteristic impedance, Z', is defined by extending the definition given in Equation 1.4.18 [Bleaney and Bleaney[2] (1976); Lorrain, Corson and Lorrain[3] (1988)]:

$$Z' = \frac{E}{H} = \sqrt{\frac{\mu}{\varepsilon'}} = Z_o \sqrt{\frac{\mu_r}{\varepsilon'_r}}, \tag{6.1.1}$$

where E and H are the electric and magnetic field amplitudes at the same position and time, Z_o is the characteristic impedance of vacuum as defined by Equation 1.4.18, μ is the absolute permeability of the medium and ε' is the generalised absolute permittivity of the medium defined by:

$$\varepsilon' = \varepsilon_o \varepsilon'_r, \tag{6.1.2}$$

where ε'_r is the generalised relative permittivity of the medium as defined by Equation 1.4.13. The generalised characteristic impedance is complex in nature reflecting the fact that a phase difference is introduced between the electric and magnetic field

amplitudes in conducting media [Garbuny[4] (1965); Hass and Hadley[5] (1972)]. From Equations 1.4.10, 1.4.12 and 6.1.1, this leads to a definition of the generalised index of refraction:

$$N = \sqrt{\mu_r \varepsilon_r'} \Rightarrow Z' = \frac{Z_o}{N}, \text{for } \mu_r = 1. \tag{6.1.3}$$

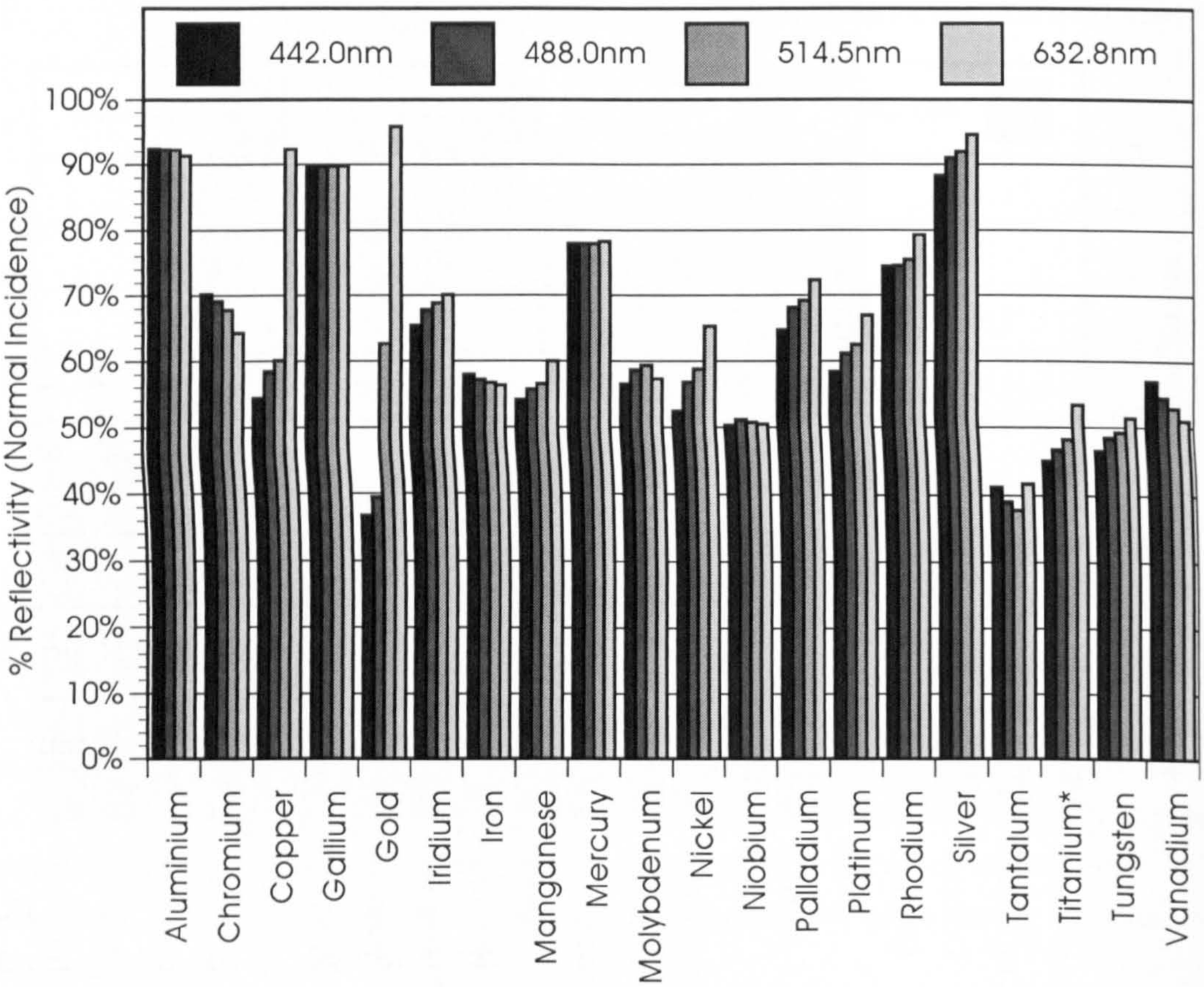


Figure 6.1.3: The reflectivity of various materials at selected wavelengths of normally incident light in vacuum.

Then assuming that the shielding material is non-magnetic, the reflectance (at normal incidence) of light going from medium 1 to medium 2, $R_{1\to 2}$, is given from the generalised form of Fresnel's equations [Garbuny[4] (1965); Hass, *et al.*[5] (1972)]:

$$R_{1\to 2} = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right)\left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right)^* = \frac{\left(n_1 - n_2\right)^2 + \left(\kappa_1 - \kappa_2\right)^2}{\left(n_1 + n_2\right)^2 + \left(\kappa_1 + \kappa_2\right)^2}, \tag{6.1.4}$$

where Z_1 and Z_2 are the generalised characteristic impedances of media 1 and 2 respectively and $N_1 = n_1 - j\kappa_1$ and $N_2 = n_2 - j\kappa_2$ are the generalised indices of refraction of media 1 and 2 respectively [from Equation 1.4.12]. Similarly the transmittance (at normal incidence) of light going from medium 1 to medium 2, $T_{1\to 2}$, is given by:

$$T_{1\to 2} = 1 - R_{1\to 2} = \frac{4\left(n_1 n_2 + \kappa_1 \kappa_2\right)}{\left(n_1 + n_2\right)^2 + \left(\kappa_1 + \kappa_2\right)^2}. \tag{6.1.5}$$

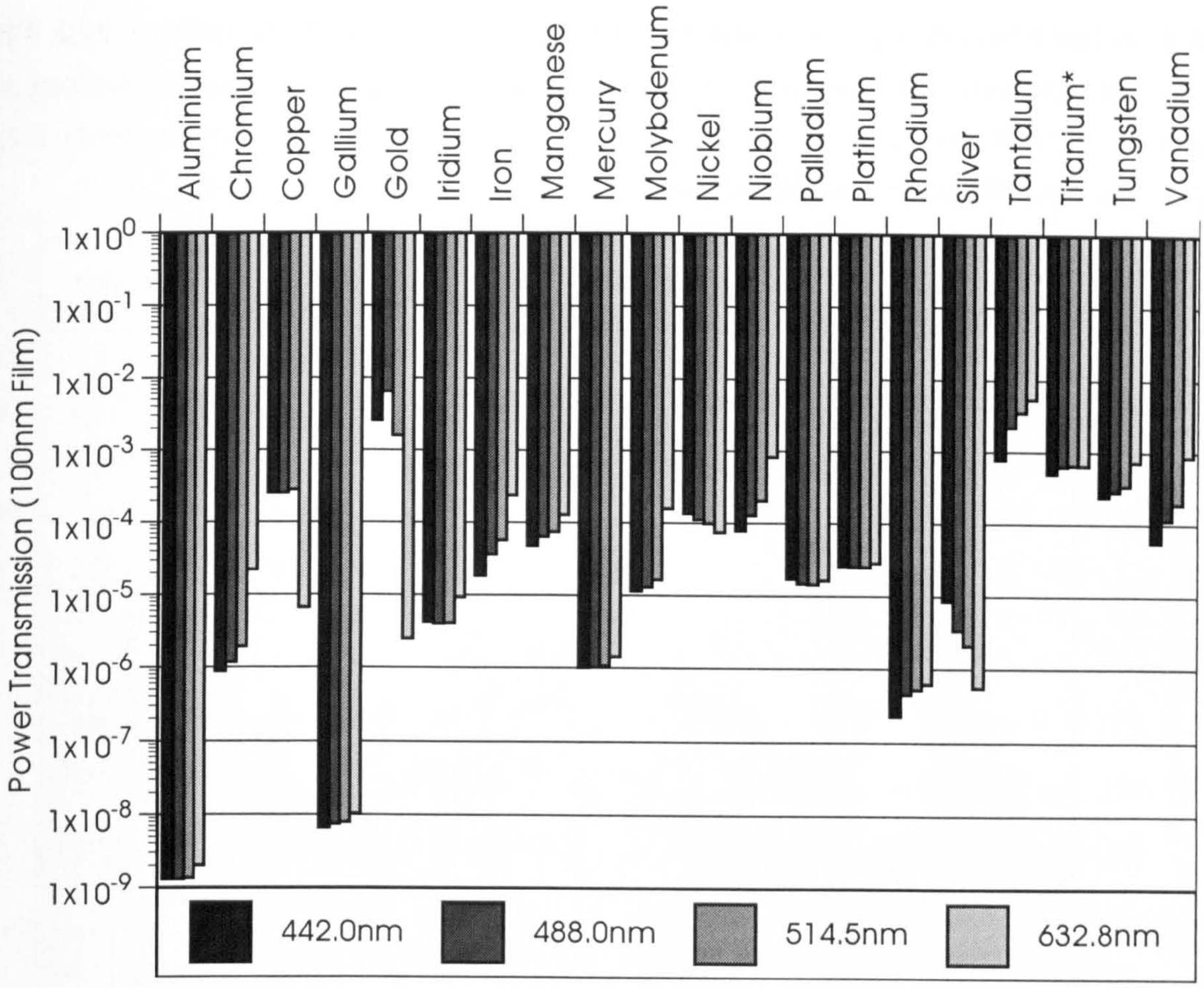


Figure 6.1.4: The intensity transmission coefficient for a 100nm thick layer of various materials at selected wavelengths.

Then from Equations 6.1.4 and 6.1.5 it is clear that the reversibility rules:

$$R_{1\to 2} = R_{2\to 1},\ T_{1\to 2} = T_{2\to 1} \tag{6.1.6}$$

apply. Figure 6.1.2 displays the amplitude characteristic decay lengths and Figure 6.1.3 displays the reflectivity at normal incidence in vacuum of various materials at selected wavelengths. These were calculated from published data [Weaver and Frederikse[6] (1996)].

The intensity of light in the direction of propagation is simply given by [Lorrain, Corson and Lorrain[7] (1988)]:

$$I = \mathrm{Re}\left(\mathbf{E}\times\mathbf{H}^{*}\right), \tag{6.1.7}$$

so that from Equations 1.4.14, 1.4.15 and 1.4.16, the intensity of light propagating in the z direction (as defined in Figure 6.1.1) in a medium with a complex refractive index is given by:

$$I = I_o \exp(-2\alpha z) = I_o \exp\left[-2\,{}^{z}\!/_{\delta}\right],\ \text{where } I_o = \mathrm{Re}\left(\mathbf{E}_o\times\mathbf{H}_o^{*}\right), \tag{6.1.8}$$

in terms of the decay constant, δ [see §1.4.1 for a formal definition of both α and δ]. Referring to Figure 6.1.1, the intensity of light in conducting medium 2 of thickness, t, after leaving medium 1 is given by:

$$I_2(z) \approx T_{1\to 2} I_1 \exp\left[-2\,{}^{z}\!/_{\delta}\right],\ 0 < z < t,\ t >> \delta, \tag{6.1.9}$$

where I_1 is the intensity of light medium 1 at the interface between media 1 and 2 and $I_2(z)$ is the intensity of light in medium 2 as a function of the distance from this interface. The thickness of medium 2 is assumed to be much larger than the decay length so that multiple reflection (thin film) effects are negligible.

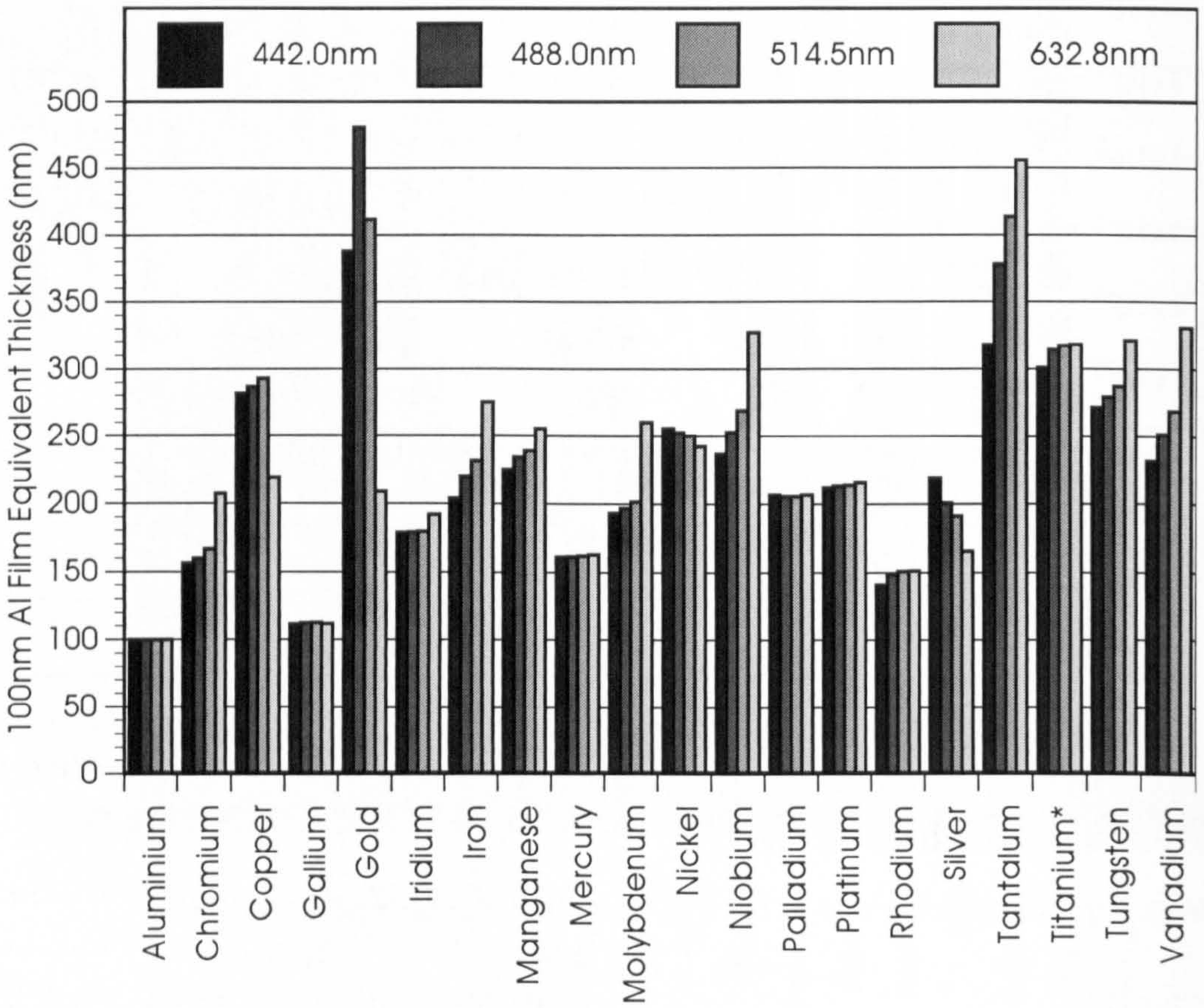


Figure 6.1.5: The thickness of various materials required to achieve results identical to 100nm of aluminium at selected wavelengths.

The intensity of light entering medium 3 after leaving conducting medium 2 (of thickness, t) is then given by:

$$I_3 \approx T_{2\to3} I_2(t) = T_{2\to3} T_{1\to2} I_1 \exp\left[-2\,t/\delta\right], t >> \delta, \tag{6.1.10}$$

where I_3 is the intensity of light in medium 3 at the interface between media 2 and 3.

In the case where medium 1 and 3 are identical:

$$Z_1 = Z_3 \Rightarrow T_{2\to3} = T_{2\to1} \Rightarrow T_{2\to3} = T_{1\to2} \Rightarrow I_3 \approx {T_{1\to2}}^2 I_1 \exp\left[-2\,t/\delta\right], t >> \delta. \tag{6.1.11}$$

Now the situation depicted in Figure 6.1.1, but with medium 1 and 3 both being vacuum can be considered. The transmittance when going from medium 1 (vacuum) to conducting medium 2 is given by:

$$T_{1\to2} = \frac{4n_2}{\left(1+n_2\right)^2 + {\kappa_2}^2}. \tag{6.1.12}$$

Then from Equations 6.1.11 and 6.1.12, the attenuation of light exiting medium 3 after passing from medium 1 through medium 2 is given by:

$$\frac{I_3}{I_1} = \frac{I_{out}}{I_{in}} \approx \left[\frac{4n_2}{\left(1+n_2\right)^2 + {\kappa_2}^2}\right]^2 \exp\left[-2\,t/\delta\right], t >> \delta. \qquad (6.1.13)$$

Figures 6.1.4 and 6.1.5 compare the ability of various materials to function as the metal shield for a SNOM probe at those wavelengths selected in Table 6.1.1. These were also calculated from the published data [Weaver, *et al.*[6] (1996)]. Figure 6.1.4 shows the transmission intensity relative to the incident intensity for a 100nm thick layer of various material choices for medium 2. It is clear that aluminium is far and away the best material followed by gallium, rhodium, silver and mercury. Therefore the most commonly used metallic shield is a 100nm thick layer of aluminium. Figure 6.1.5 shows the thickness of the other materials required to achieve identical results to a 100nm thick layer of aluminium.

6.2 DEFINING THE APERTURE

To define an aperture of dimensions ranging from 20nm to 150nm in a film 100nm to 200nm thick is not straightforward for several reasons, and so some thought had to be given to the approach to be adopted.

The most naive approach would be to attempt to directly lift-off circles in the metal shielding layer using electron-beam lithography with PMMA resist. However, from a practical standpoint, even trying to leave a resist structure of the order of 100nm in diameter standing 150nm to 200nm tall in the midst of a large exposed area using positive resist is extremely difficult [McCord, Wagner and Donohue[8] (1993)]. This is because of the intra-proximity effect: if an area of 20μm square is exposed then the backscatter dose will be sufficient to, at worst, fully expose the small area of resist defining the aperture, or at best, ruin any overhanging profile in this resist structure [see §2.7.1 and §3.5.3]. Whilst it would be possible, in principle, to compensate by increasing the region being left unexposed (proximity correction), the overhanging profile necessary for lift-off would probably be much more difficult to achieve. In addition, the pattern to be exposed to achieve a particular aperture size would have to be optimised on an individual basis, thus increasing development workload further. If a resist with lower sensitivity (or higher contrast) than PMMA could be found then this approach would have more chance of succeeding. However, relative to other resists, PMMA is generally regarded as having low sensitivity (high contrast) [Wilkinson and Beaumont[9] (1980-1993); Pease[10] (1981); Hatzakis[11] (1988)].

An alternative might be to try a negative resist, but as has been noted in §2.2, negative resists tend to suffer from low resolution, leave residues behind, are not easily removed after processing and are usually incapable of providing the overhanging profile necessary

for lift-off [Pease[10] (1981); Namaste, Obendorf and Rodriguez[12] (1988)]. Thus this approach would seem to be unfavourable also.

An obvious way to avoid these problems would seem to be to use the inverse of the resist pattern to define the aperture through a subtractive, as opposed to additive, process, *i.e.* dry etching. This would mean either carrying out the aperture definition after the release etch, as the aluminium shielding layer [see §6.1] will not survive the release etching solution [see §4.4.2], or switching to a different metal.

The former strategy is not appealing since it would mean two lithographic processes being carried out on fragile contact mode AFM cantilevers after release. This would have to be done without breaking the cantilevers and would probably mean foregoing spin-coating altogether, in addition to dealing with capillary force issues [see §4.4.2 and Kobayashi, Kim and Fujita[13] (1993); Legtenberg, Tilmans, Elders and Elwenspoek[14] (1994)] and the dry etching of these cantilevers in partial vacuum where they may be likely to oscillate.

Shield Material	100nm Aluminium Equivalent Thickness at 488.0nm	Melting Point (°C)	Boiling Point (°C)
Aluminium	100	660	2467
Gallium	112	30	2205
Rhodium	149	1965	3727
Chromium	160	1857	2672
Mercury	162	-39	357
Iridium	179	2410	4130
Molybdenum	197	2617	4612
Silver	201	962	2212
Palladium	206	1554	3140
Platinum	214	1772	3827

Table 6.2.1: Some properties of the ten best SNOM shield materials at a wavelength of 488.0nm. Melting and boiling point data from **Goodfellow**[15] (1993).

The latter strategy would be possible but for the choice of metallic shielding and etch chemistry. From Figure 6.1.5 the ten best shielding materials at a wavelength of 488.0nm are displayed in Table 6.2.1. Gallium is ruled out because its melting point is so low that there is a strong possibility illuminating the aperture with a laser will cause the shield to melt or change size by diffusion. Mercury is also ruled out because it is liquid at room temperature. Noble metals such as rhodium, iridium, palladium and platinum are all ruled out because their non-reactive nature means that they are etched by physical (sputtering) rather than chemical processes [Murad[16] (1993-1996)]. The resist, being the softer material, sputters far more readily than most metals [Cantagrel[17] (1975)] and would suffer from severe mask erosion effects [*e.g.* Karulkar and Wirzbicki[18] (1988);

Lothian, Ren and Pearton[19] (1992)]. These effects could only be overcome by using a much thicker resist layer, but this would require an unfeasibly high aspect ratio in the resist pattern. Such high aspect ratio structures are prone to collapsing due to lack of mechanical strength in the resist [McCord and Pease[20] (1987)]. This is in addition to the collimating nature of such a resist structure [Yunkin, *et al.*[21] (1996)] slowing down the sputtering and etching rate of the metal [*i.e.* 'RIE lag', *e.g.* Gottscho[22] (1993); Zachariasse and Broers[23] (1996)]. Chromium must be ruled out because it is a highly dry etch resistant material [Rangelow and Loschner[24] (1995)], due to the tough nature of its oxide. It is commonly used as a masking material for other dry etches and so etching it from resist would be extremely difficult for larger apertures, and almost impossible at dimensions of around 50nm. Silver, like aluminium, reacts with the release etch and so must also be ruled out. This then leaves molybdenum as the best choice for implementing this approach, however it is also difficult to evaporate and etch, particularly at the small dimensions required, and evaporated layers tend to be highly stressed [Thoms[25] (1993-1996)].

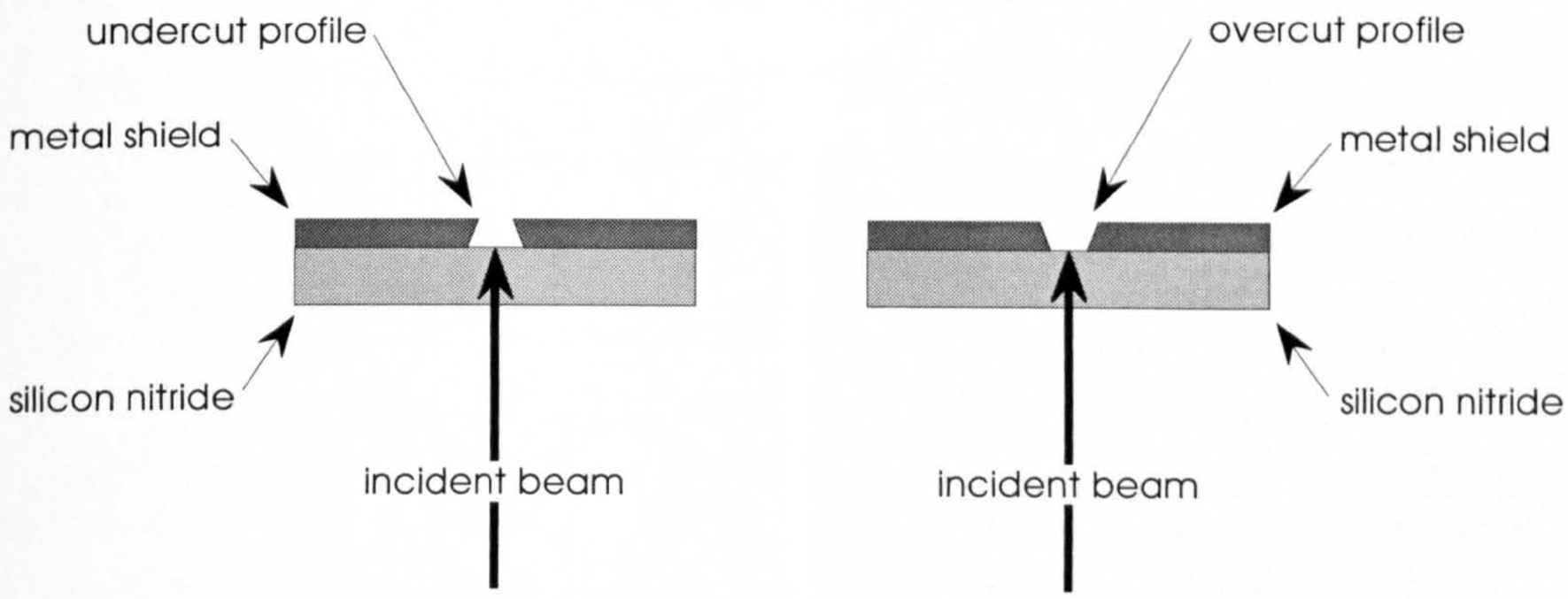


Figure 6.2.1: Diagram illustrating the different aperture profiles possible.

A further problem with dry etching the apertures directly is the need to produce a vertical or undercut profile, rather than an overcut profile. This is because an overcut profile would mean that the narrowest part of the aperture would be on the same side of the tip as the incident laser beam, with the widest part of the aperture being on the transmitted side, facing the specimen, as shown in Figure 6.2.1. Then the widest part of the aperture would principally determine the resolution obtained, while the narrowest part would primarily determine the light throughput obtained. Making the aperture diameter smaller on the transmitted side would also result in the decrease of the aperture diameter on the incident side and eventually the aperture would be totally cut-off on this side of the tip. To prevent total cut-off with an aperture size of 50nm and metal 200nm thick (molybdenum in Table 6.2.1) would require an overcut angle (angle between vertical and sidewall) of less than 7°. Therefore, even a small amount of overcut could result in a dramatic drop in light output at small aperture sizes.

This meant that a different approach was required. This approach was based on the idea that an aperture could be etched in the silicon nitride cantilever itself. This could then act as a template upon which the shield metal could be evaporated. It is common knowledge that as metal is evaporated onto an edge, the metal tends to extend the edge out further. This effect has even been used to advantage in fabricating sharp tips [Spindt, Brodie, Humphrey and Westerburg[26] (1976)]. Making use of this effect here would then give an aperture with an undercut profile as desired. This is shown schematically in Figure 6.2.2. The angle of closure will depend on various aspects of the deposition process and requires some investigation.

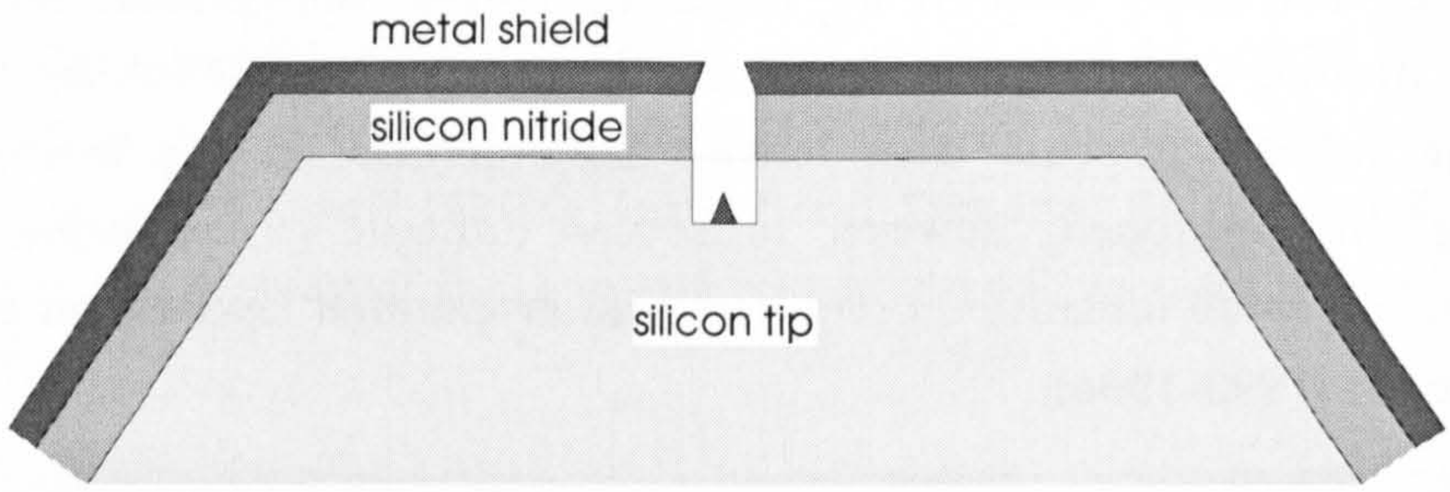


Figure 6.2.2: Diagram illustrating the function of the aperture in the shielding layer.

There are two alternative strategies that can be used for this process. The first is to etch the silicon nitride after the release etch, but as before this was ruled out because of the anticipated difficulties and high level of risk in trying to carry out lithography on fragile contact mode AFM cantilevers.

The second strategy, which consisted of defining the silicon nitride apertures prior to release etching, was chosen for development. Here also there were two possibilities. The metal could be deposited onto the silicon nitride aperture prior to release etching or after release etching. The former is satisfactory if the metal used is resistant to the release etch, but requires that the dry etching profile extend into the silicon by at least the thickness of the metal (and with an undercut profile) so that the metal deposited inside the aperture falls out during the release etch when the silicon is removed, as illustrated in Figure 6.2.2. The latter is best for depositing metals which are not resistant to the etch, such as aluminium. It may be possible to use aluminium with the former method if a protection layer is also deposited onto the shield, however previous experience of problems with pinholes, stress (thermal and physical), adhesion and the removal of any protection layer after cantilever release meant that this would be increasing process complexity with an unnecessary level of risk for little benefit [see §3.7.3].

6.2.1 Direct Aperture Definition in SiN from PMMA

First attempts concentrated on exposing an aperture in PMMA and using the PMMA as a mask for directly dry etching the silicon nitride. As the aperture in silicon nitride was to be defined prior to the release etch, there would be a solid silicon substrate beneath it (albeit with three-dimensional topography). It was therefore envisaged that exposure of

the PMMA by backscattering would be similar to that on a solid silicon substrate (with a silicon nitride coating) rather than a free silicon nitride membrane.

Initial experiments concentrated on establishing the feasibility of employing this approach. The questions to be answered were:

(a) Could small enough apertures be defined in PMMA?

(b) Could these apertures be transferred to the silicon nitride using standard, already developed, dry etching processes?

(c) Were such apertures sufficiently cleanly defined to allow the subsequent metallisation process to result in a completely clear aperture?

(d) Were there any unanticipated problems to be overcome?

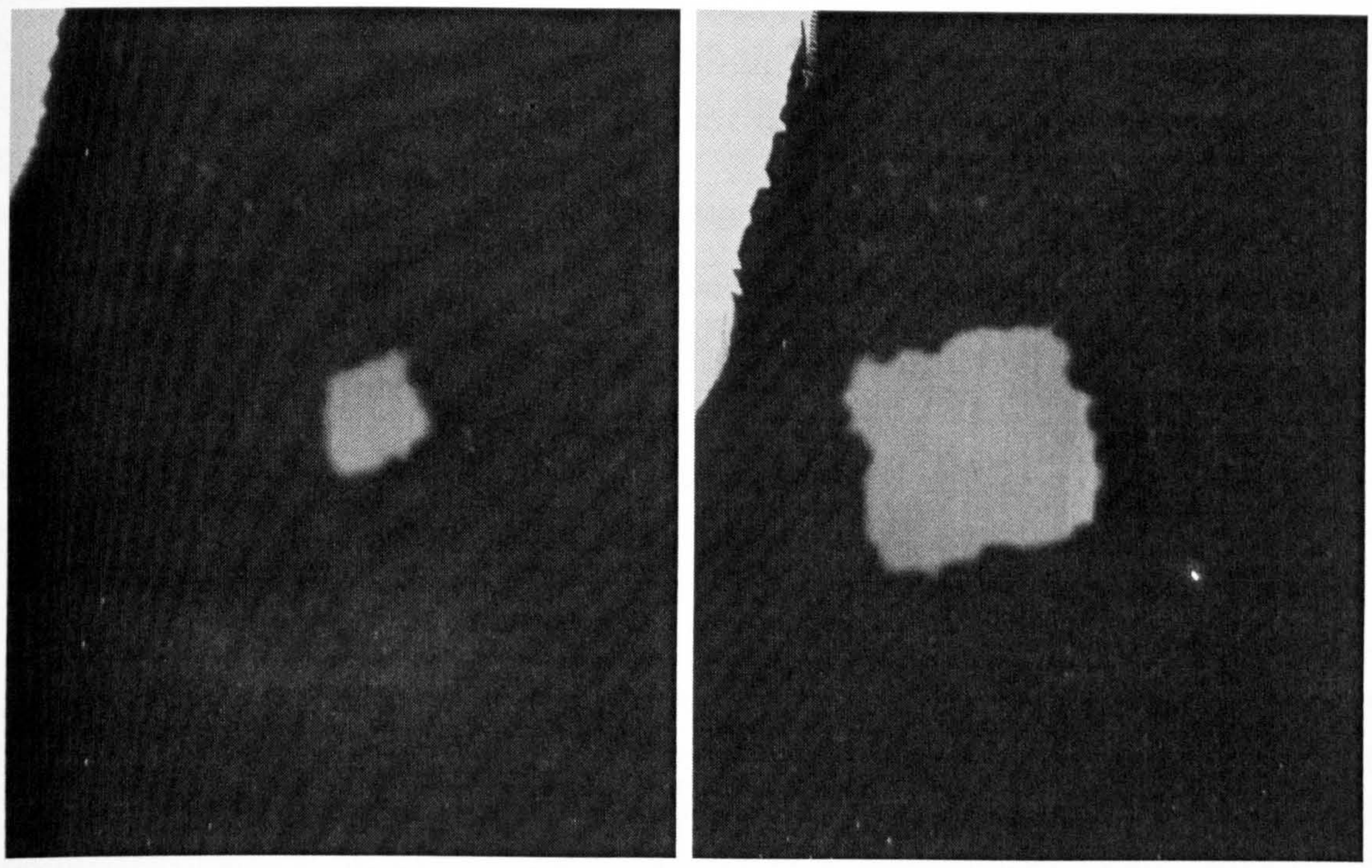


Figure 6.2.3: STEM micrographs of some square apertures in a 50nm thick silicon nitride membrane ('ama0001'). (a) Left Image: a 60nm wide aperture; (b) Right Image: a 140nm wide aperture.

To answer these questions, and given that this process was being developed in parallel with many of those discussed in Chapters 4 and 5, it was decided that the first tests should be carried out on released silicon membranes. The membrane substrates were produced during the photolithography levels, using patterns designed specifically for this purpose [see §4.2]. These membranes consisted of the original LPCVD silicon nitride coating on the silicon wafers. For the first tests these membranes were 300μm square and 50nm thick. As the test exposures were being done on membranes, it was clear that the backscatter effect would be much reduced and therefore result in apertures much smaller than would be the case with a solid substrate beneath the silicon nitride layer. However, the advantage of using these 'through-wafer' membranes was that it would be straightforward to determine whether the apertures were clearing out. If sufficiently small apertures could not be defined in a such a thin silicon nitride layer with reduced

backscatter effects, then this part of the process would require significantly greater development effort than originally anticipated.

The first attempt (ama0001) was as follows:

(a) 2.5% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(b) The resist was dried by baking at 180°C for one hour.

(c) 2.5% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(d) The resist was dried by baking it overnight at 180°C.

(e) The resist was exposed using a resolution of 5nm and a spot of 12nm at 50kV, with global alignment (100kV was not available due to a problem with the beamwriter). The pattern simply consisted of exposing square openings ranging from 70nm to 550nm in size with a dose of 3000μC/cm^2.

(f) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(g) The sample was dry etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 3 minutes in the Plasma Technology BP80 [§3.6.4].

(h) The resist was stripped off in acetone.

(i) 10nm of nichrome and 100nm of gold were evaporated onto the sample using the Plassys MEB 450.

The last step was carried out simply to gauge the effect of coating with metal; gold was conveniently available in the Plassys MEB450 evaporator [3.6.2], and at this point the wavelength being used for any light throughput measurements was 632.8nm (helium-neon laser – see Table 6.1.1).

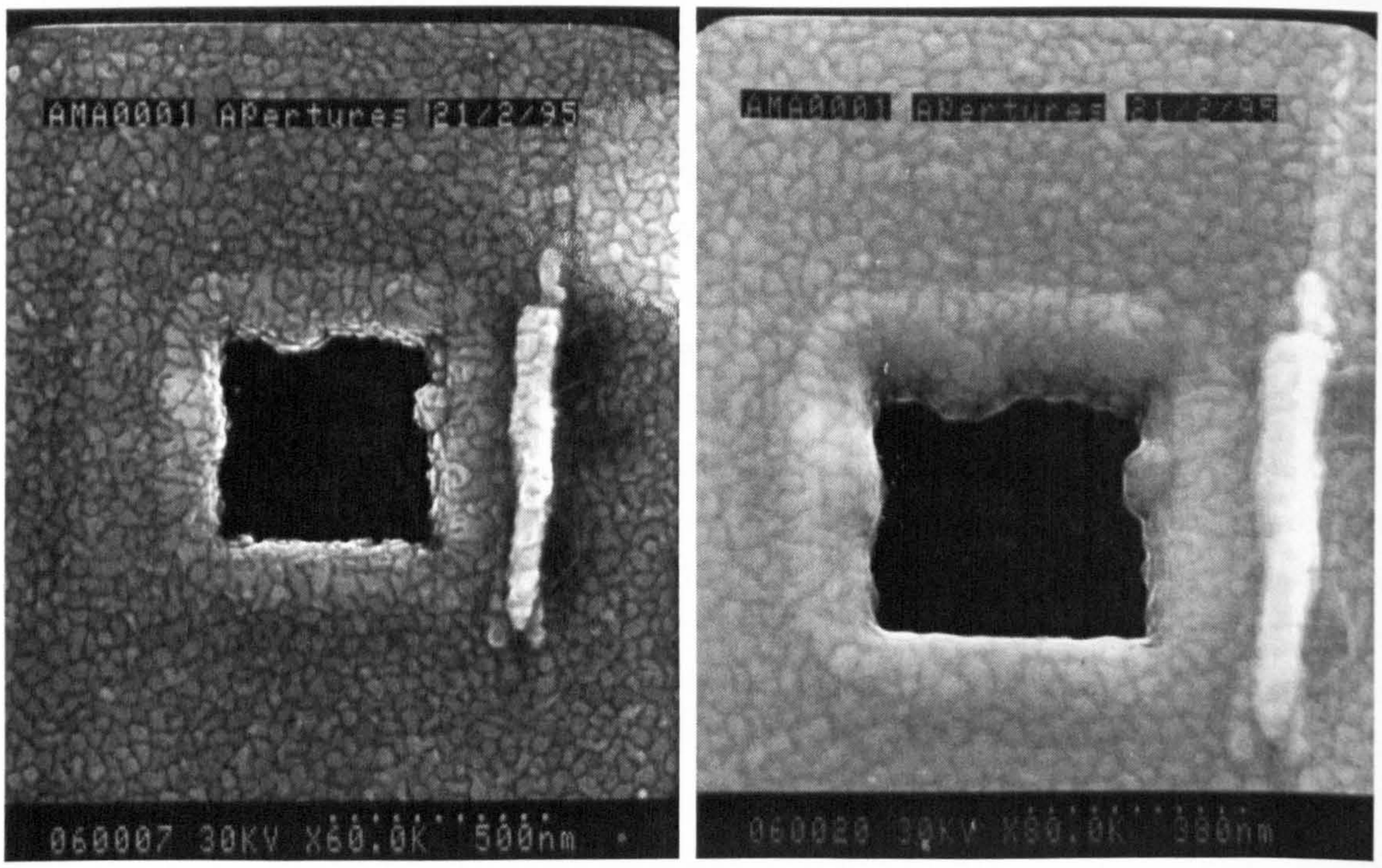


Figure 6.2.4: Micrographs showing a 500nm square aperture in a 50nm thick silicon nitride membrane ('ama0001'). (a) Left Image: flat view; (b) Right Image: 20° tilted view.

The sample was then examined [Weaver[27] (1993-1996)] in the Jeol 100 CXII STEM [Thoms, Beaumont and Wilkinson[28] (1989)] using scanning transmission electron microscopy to determine whether the apertures had cleared out. These results were quite encouraging with apertures ranging from 60nm to 500nm (square) clearing out. Figure 6.2.3 shows an STEM picture of a 60nm and 140nm square aperture. As can be seen, there is a significant degree of edge roughness. At this point it was thought that this could be attributed either to the granularity of the metal or the roughness of the dry etching process. Figure 6.2.4 shows two micrographs of a 500nm square aperture taken with the Hitachi S900 SEM [§3.6.3]. Figure 6.2.4a is a flat view while Figure 6.2.4b is taken at an angle of 20° with higher brightness. The size of the metal grains can be clearly made out in these micrographs. Also noticeable from Figure 6.2.4b is the fact that the edge profile appears to be overcut (where an undercut was expected) and that the profile does not seem to be simple due to the distinct presence of a border around the aperture. The latter effect may simply be due to a variation in the metal thickness, but the presence of such a marked boundary seems to rule this out. The cause of both these effects became clearer in subsequent tests.

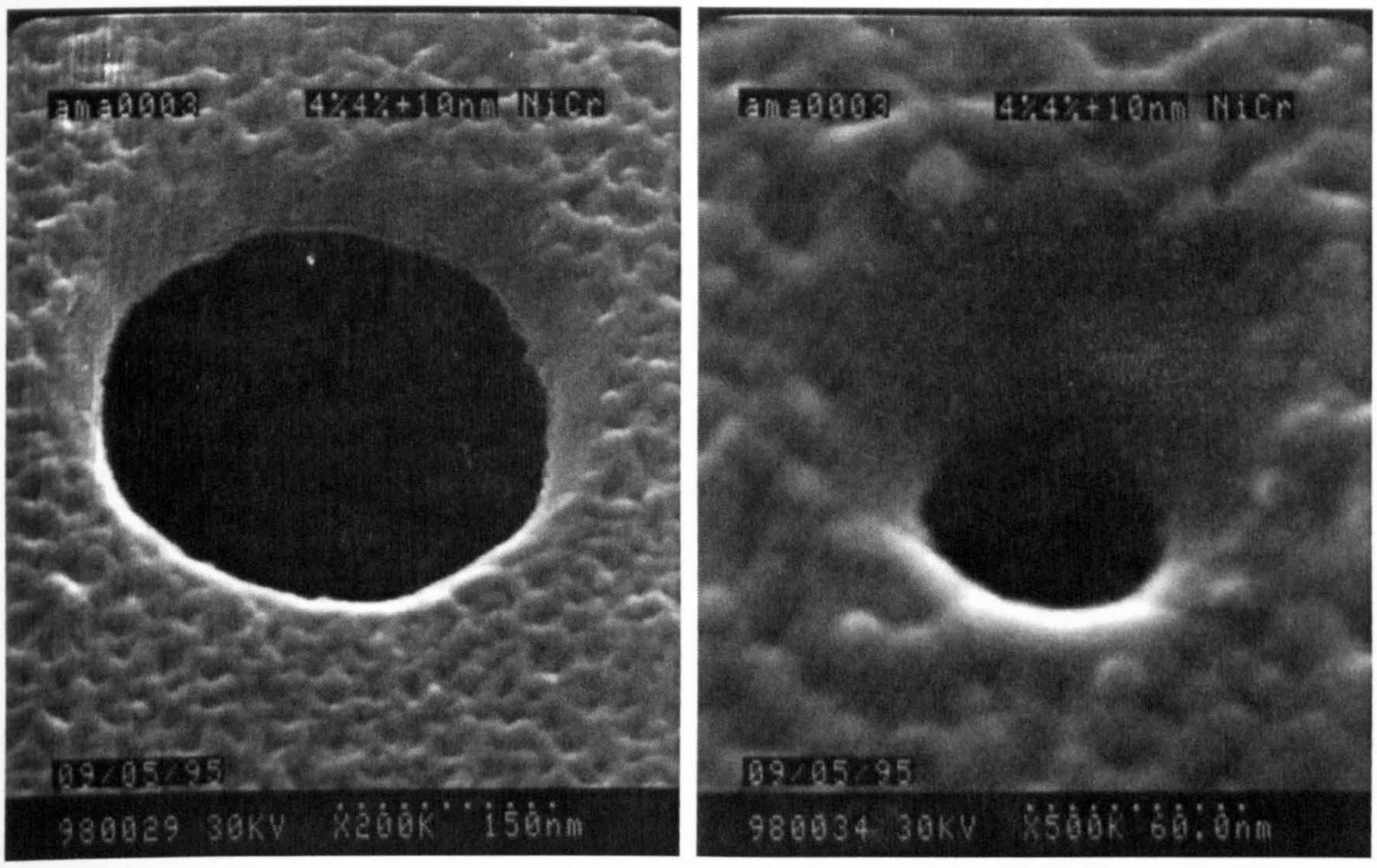


Figure 6.2.5: Micrographs of circular apertures, with resist still present, in a 50nm thick silicon nitride membrane ('ama0003'). The sample has been coated with 10nm of nichrome as a charge conduction layer. The sample has been tilted by 30° to allow the etched edges of the aperture to be viewed. (a) Left Image: a 315nm diameter aperture; (b) Right Image: a 70nm diameter aperture.

Then the evolution of the aperture from one step to the next under the proposed fabrication scheme was studied. At this point it was also decided that more representative results would be achieved by explicitly making the apertures circular. Therefore a new pattern was designed with circular apertures ranging from 30nm to 630nm in diameter (ama0003). As before the substrate consisted of 50nm thick LPCVD

silicon nitride membranes suspended by a solid silicon substrate. The resist was changed to a bi-layer of 4% BDH spun and baked as before. One possible reason for the overcut profile was erosion of the resist mask during the dry etching of the silicon nitride [Murad[16] (1993-1996)]. This being the case, the increase in resist thickness would help reduce the amount of overcut in the silicon nitride by increasing the thickness of resist to be eroded. The pattern was exposed at 100kV this time, using a 12nm spot and a dose of 2500μC/cm^2. The dose was reduced to accommodate the effect of increased exposure by forward scattered electrons [Broers[29] (1988)]. The pattern was then developed and dry etched as before.

Without stripping any remaining resist one of the samples was coated with a very thin charge conduction layer in the Plassys MEB 450 evaporator (10nm of nichrome) and then examined in the Hitachi S900 SEM. Figure 6.2.5 shows micrographs of two different apertures taken with the sample tilted 30° from the horizontal. Figure 6.2.5a shows a 315nm aperture in the 50nm thick silicon nitride layer. The surface corrugation is probably due to the evaporated charge conduction layer and underlying PMMA (if any). Figure 6.2.5b shows a 70nm aperture and from this it is clear that there has been a significant degree of overcut, despite the increased resist thickness. The smallest aperture found was 55nm in diameter although due to contamination of the specimen by oil in the SEM chamber, the micrograph obtained was not of reproducible clarity.

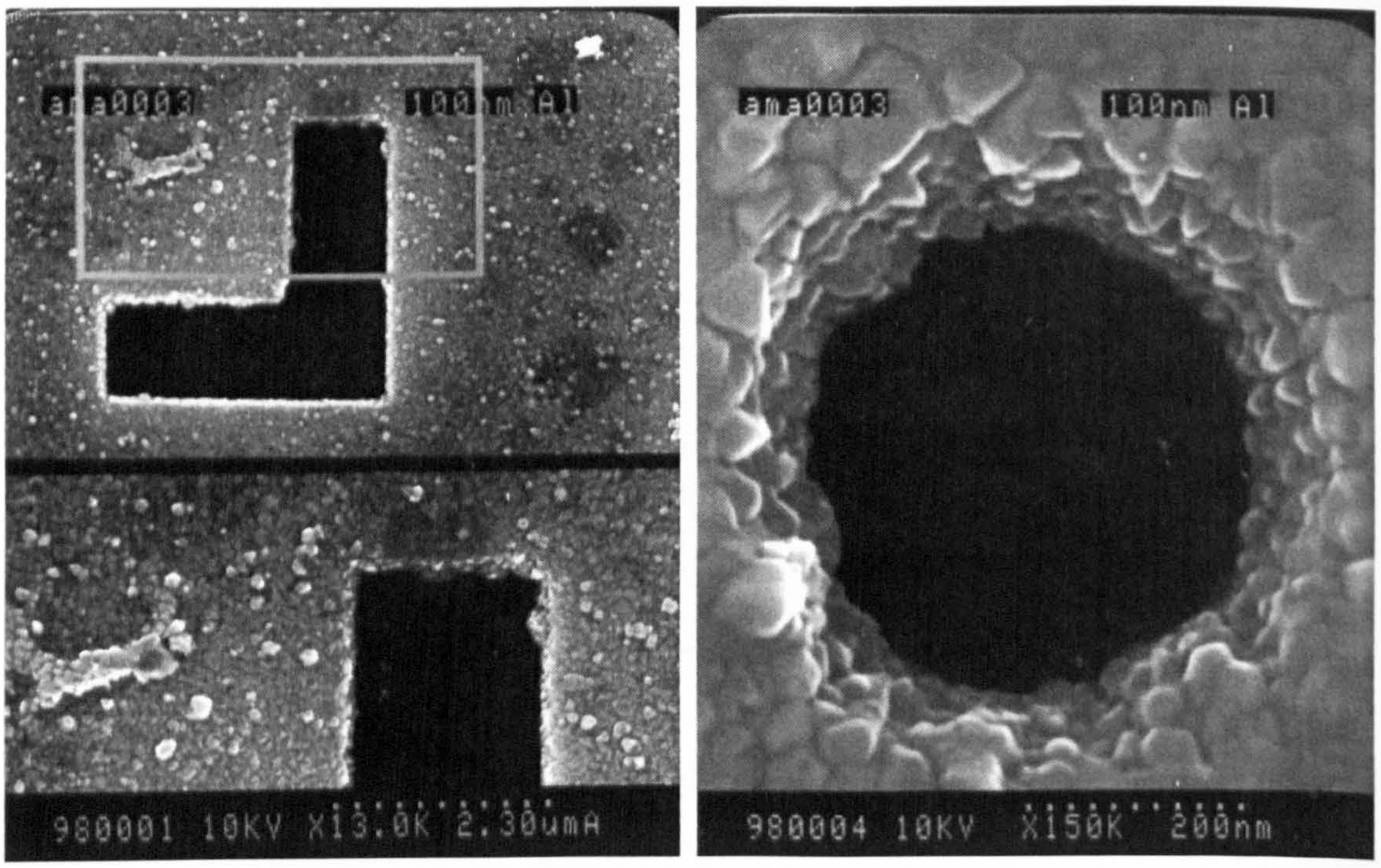


Figure 6.2.6: Micrographs of 100nm thick aluminium coated apertures in a 50nm thick silicon nitride membrane ('ama0003'). (a) Left Image: an 'L' shaped aperture used for location purposes in the SEM. The bottom half is a 2x magnification over the displayed scale; (b) Right Image: a 400nm diameter aperture.

From the preceding work it was clear that the method of pattern transfer required some development before apertures with the desired characteristics could be produced,

especially in light of the fact that the thickness of the silicon nitride of the contact mode AFM cantilevers being fabricated was likely to be around 500nm [see §4.2].

Another sample from the same batch as imaged in Figure 6.2.5, (but without a charge conduction layer) was coated with 100nm of aluminium in a modified bell-jar evaporator using resistive heating. Figure 6.2.6a is an SEM micrograph of an 'L'-shaped aperture used as an orientation and location marker and Figure 6.2.6b is an SEM micrograph of a large 400nm diameter aperture. The grain size was estimated from these as being between 40nm and 60nm. The effects of the overcut aperture profile can also be clearly seen.

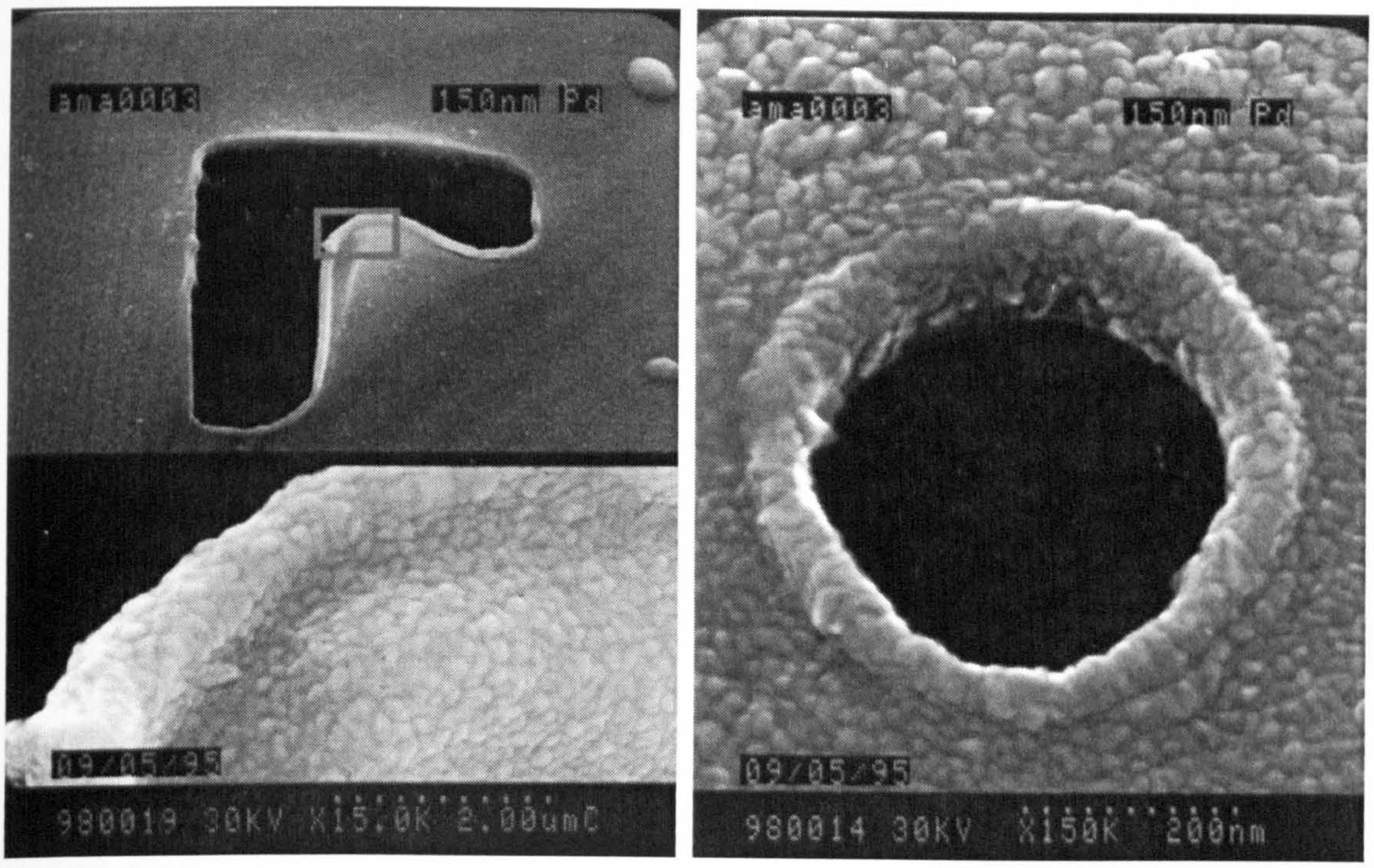


Figure 6.2.7: Micrographs of 150nm thick palladium coated apertures in a 50nm thick silicon nitride membrane ('ama0003'). Compare this with Figure 6.2.6. (a) Left Image: an 'L' shaped aperture used for location purposes in the SEM. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: a 30° tilted view of a 400nm diameter aperture.

For comparison, another sample from the same batch as used in Figure 6.2.6 was coated with 150nm of palladium in the Plassys MEB 450 evaporator. Figure 6.2.7a is an SEM micrograph of the 'same' 'L'-shaped location and orientation aperture as Figure 6.2.6a. From this it is clear that the palladium layer is far more stressed than the aluminium layer. This is further confirmed by looking at the aperture edges as shown in Figure 6.2.7b which is taken with the sample tilted 30° from horizontal. This shows a large 400nm diameter aperture. From this it can be seen that the grain size is much smaller, estimated at around 10nm to 20nm and that the edges of the aperture appear to have been pulled up by stress between the palladium and silicon nitride layers. This increase in stress may be due to a number of factors ranging from the extra thickness of the layer, the higher evaporation temperature necessary for palladium over aluminium, the deposition rate used and the smaller grain size.

Further evidence that the peculiar border round the aperture of Figure 6.2.7b is due to residual stress came from the observation that this border was not present for apertures much smaller than 200nm in diameter. Figure 6.2.8 shows SEM micrographs of two such apertures. It would be expected that smaller apertures (with smaller circumferences and areas), perturb a uniform metal film less and so a smaller the amount of energy (from residual stress) is available for the formation of this border.

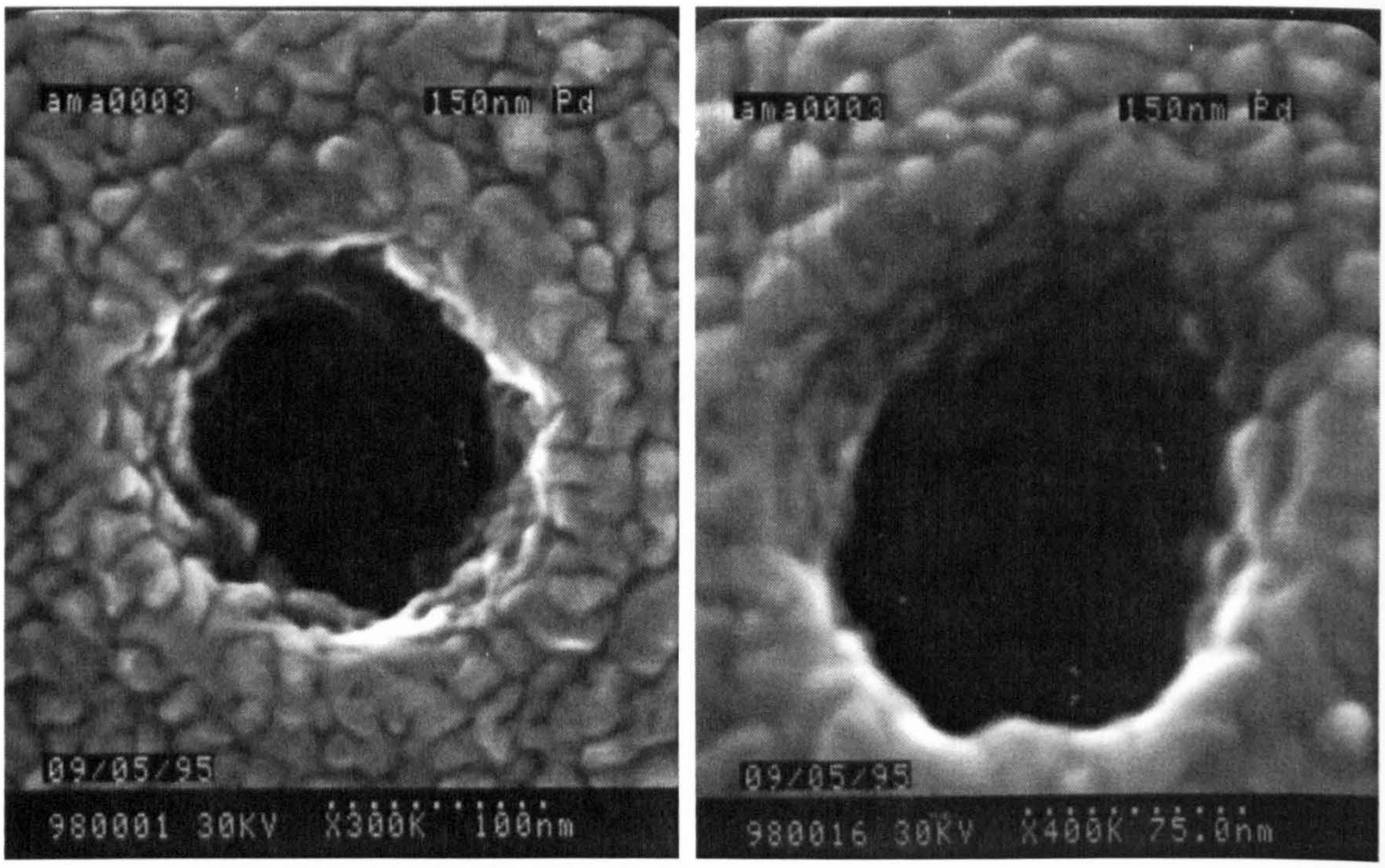


Figure 6.2.8: Micrographs of 150nm thick palladium coated circular apertures in a 50nm thick silicon nitride membrane ('ama0003'). (a) Left Image: a 150nm diameter aperture; (b) Right Image: a 30° tilted view of a 100nm diameter aperture.

Another qualitative observation from these samples was that the smaller the grain size, the better shaped the apertures, with the edge roughness of the aperture typically being estimated at around half the grain size. This meant that for the aluminium sample deposited under the conditions above, the shape of smaller apertures (less than 100nm) would become highly unpredictable as the edge roughness (around 20nm) approached that of the aperture diameter. Thus if aluminium was to be used then the deposition conditions would have to be optimised to produce smaller grains. However, this was left for future development due to time constraints, and in the interests of evaluating fabrication reproducibility, it was decided that all further development be based on 150nm to 200nm thick palladium layers. [This metal, whilst second from bottom in Table 6.2.1 in terms of the thickness required to emulate the opacity of a 100nm thick aluminium layer at a wavelength of 488.0nm, was available in the Plassys MEB 450 and would be resistant to the release etch.] The border around the apertures would not be an issue since only apertures with a final diameter of 200nm or less were of interest.

To determine the optimum doses and pattern designs required to obtain apertures of various sizes, it was necessary to move to a substrate more closely related to the

pyramidal tip and cantilever. This meant exposing the resist whilst the silicon nitride was still attached to a solid silicon substrate. However, in order to determine whether the apertures had cleared out, a more novel test pattern was devised. This consisted of opening up rectangular windows 250μm long and 50μm wide in the silicon nitride so that crystallographic etching in potassium hydroxide solution would produce an etch terminated 'v'-groove [see §3.7.3]. If the rectangular window was spanned by 10μm wide beams, oriented along the [013] axes as shown in Figure 6.2.9, then these beams would rapidly become undercut and form silicon nitride 'bridges', affording sufficient contrast to be able to determine which apertures had cleared out. These beams would be fixed at both ends to prevent capillary forces causing the beams to become stuck to the silicon substrate beneath during drying [Kobayashi, *et al.*[13] (1993); Legtenberg, *et al.*[14] (1994)]. The apertures, beams and openings could be defined in a single run using different matrices (spot sizes) to avoid the need for any alignment level.

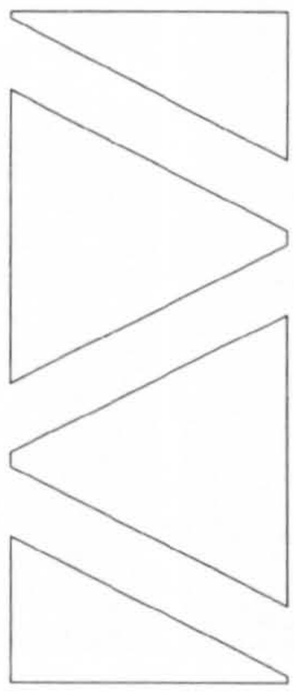


Figure 6.2.9: The design used to pattern silicon nitride beams upon which apertures could be placed, before wet etching the silicon beneath and undercutting the beams to release them from the substrate.

The first such pattern consisted of 81 apertures made up of single pixels, ranging from squares 1 pixel (5nm) to 9 pixels (45nm) across. [Here a single pixel results in a single point exposure by the electron beam. These are drawn as 5nm squares when the pattern is fractionated at 5nm resolution.] The samples were processed as follows:

(a) 8% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(b) The resist was dried by baking at 180°C for one hour.

(c) 4% Elvacite was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(d) The resist was dried by baking at 180°C for one hour.

(e) 4% Elvacite was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(f) The resist was dried by baking it overnight at 180°C.

(g) The resist was exposed using a resolution of 5nm and a spot of 12nm at 100kV, with global alignment. The pattern described above was exposed with doses of 1750μC/cm^2, 2000μC/cm^2, 2500μC/cm^2 and 3000μC/cm^2.

(h) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds (this high contrast developer ensures that features are cleared out at the possible expense of lower resolution).

(i) The samples were dry etched with 35mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 4 minutes in the Plasma Technology BP80 [§3.6.4].

(j) The resist was stripped off in acetone.

(k) The samples were wet etched in 7 molar (+5%) potassium hydroxide solution for 90 minutes at 80°C.

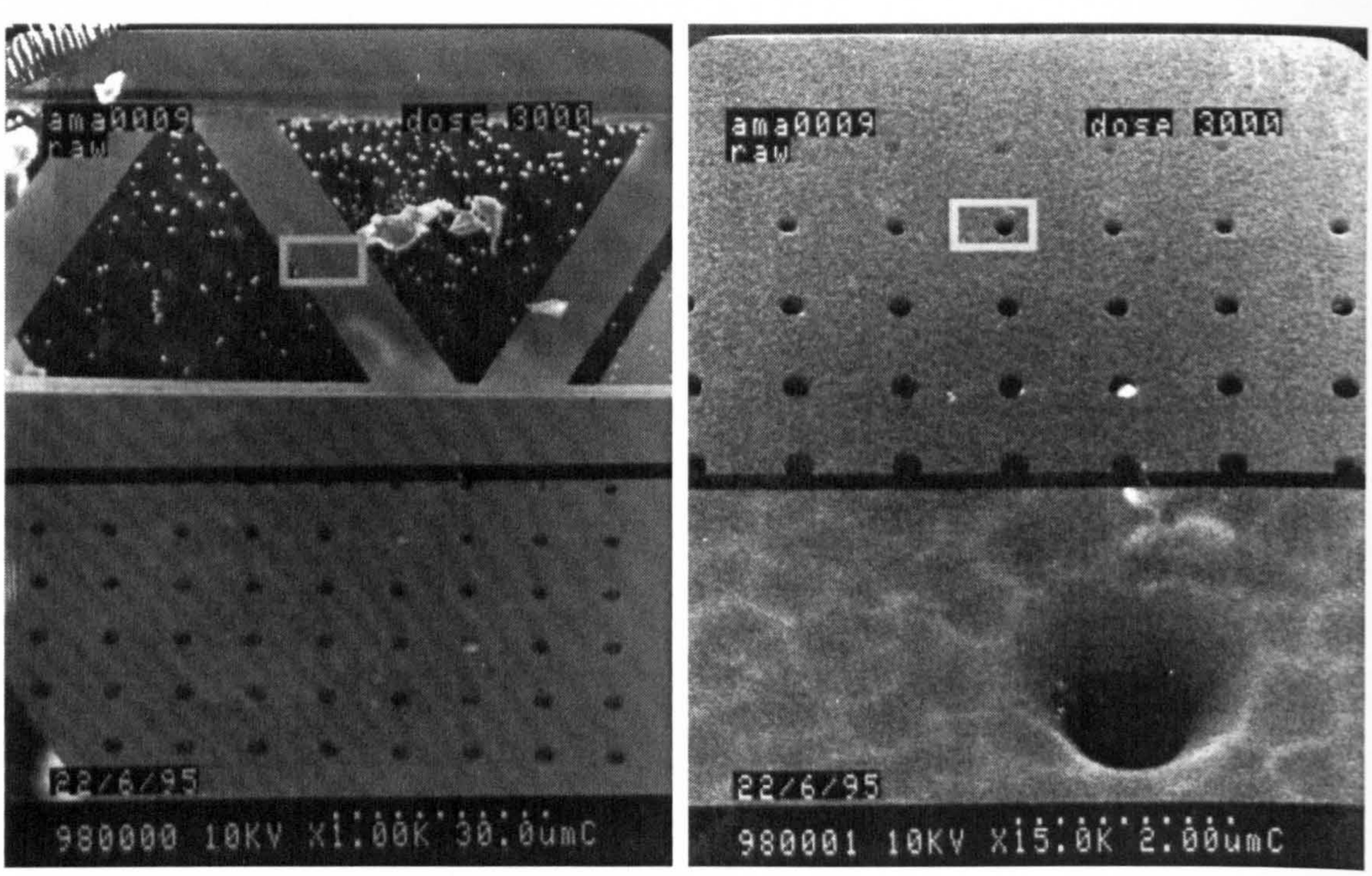


Figure 6.2.10: Micrographs of apertures fabricated on 50nm thick free-standing silicon nitride 'bridges' ('ama0009'). The sample was not coated in a charge conduction layer. (a) Left Image: aperture array in the middle of one of the 'bridges'; (b) Right Image: a close-up of one of the smaller apertures achieved. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

The use of two layers of the higher molecular weight Elvacite was intended to provide a thicker and more dry-etch resistant top layer of resist, as it was found that resist erosion rates decreased with increasing molecular weight of PMMA [Murad[16] (1993-1996)]. The operating pressure was increased because it had been suggested that if the resist was principally being eroded by physical attack from high energy ion bombardment, then this would be one way of increasing the ratio of substrate etched to resist eroded [Murad[16] (1993-1996)]. This would decrease the mean free path of the ion radicals in the chamber and hence reduce their velocity. The anisotropy of the etch would also normally be reduced, however, here the main factor determining the etch profile was the resist erosion rate rather than the directionality of the ions. This was known from results achieved by others using this dry etch with more resistant masking materials [Murad[16] (1993-1996)]. The silicon nitride etch rate would be expected to increase, due to the greater density of free radicals (up to a point, since at high ion energies, the ion collisions themselves generate more free radicals due to avalanche effects), but as the etching proceeded, the lower ion energies and reduced directionality would make it more difficult for the reactive

species to reach the bottom of the trench [Murad[16] (1993-1996)]. The etch duration was increased because the substrates being used now had a 100nm thick layer of LPCVD silicon nitride [see Appendix E (§E.1)].

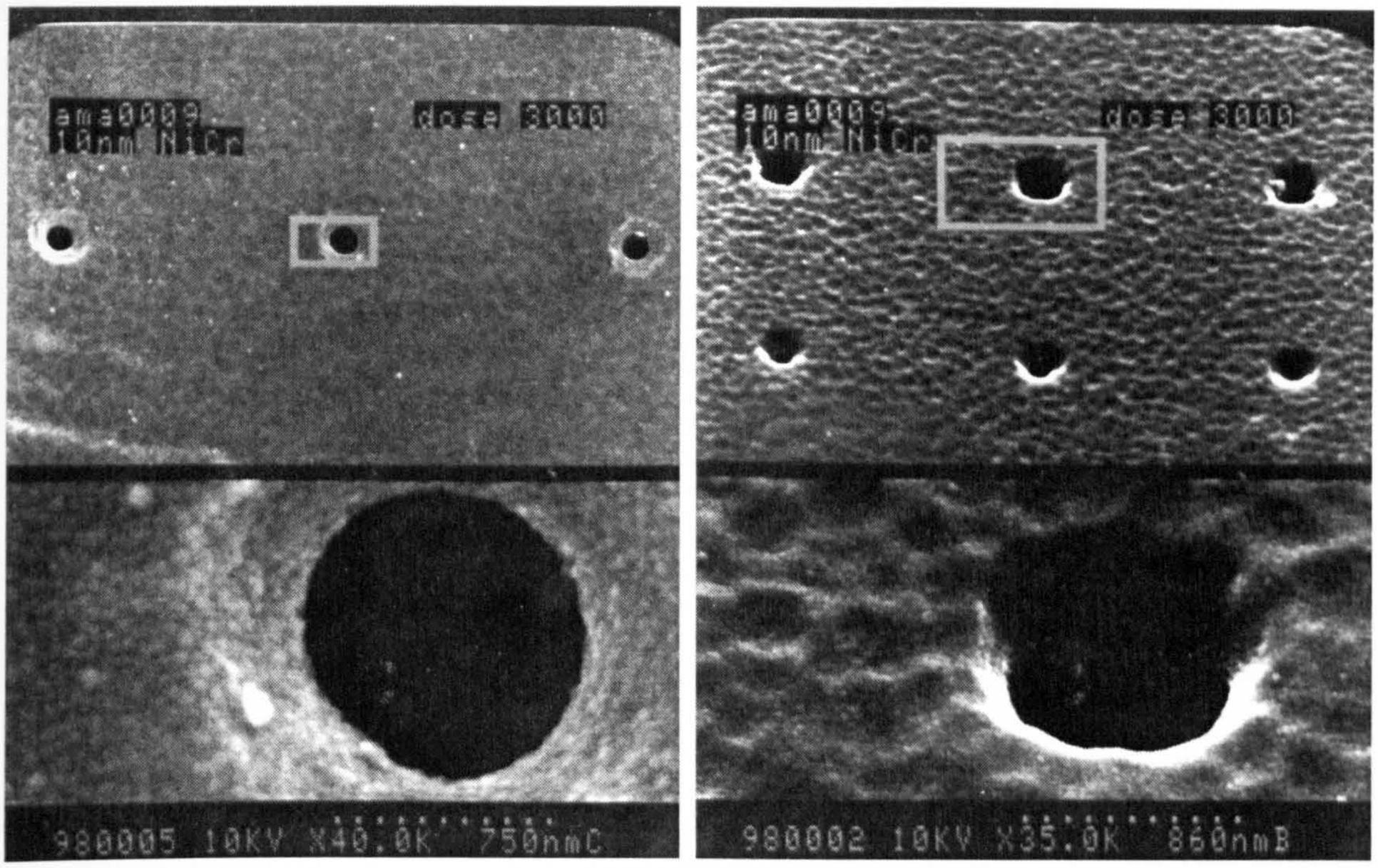


Figure 6.2.11: Micrographs of apertures fabricated on 50nm thick free-standing silicon nitride 'bridges' ('ama0009'). The sample was coated in a 10nm thick nichrome charge conduction layer. (a) Left Image: a line of apertures around 75nm in diameter. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: a 30° tilted view of (a). The bottom half is a 5x magnification over the displayed scale.

After rinsing and drying, the samples were examined in the Hitachi S900 SEM. However, the range of doses used in this test proved to be insufficient to determine the threshold doses and aperture reproducibility as a function of aperture diameter in sufficient detail, although, as can be seen from Figures 6.2.10 and 6.2.11 this method of measurement was very successful. Figure 6.2.11a even suggests good fabrication reproducibility at aperture dimensions of 75nm, although the overcut profile is very much apparent.

The same experiment was repeated but with a greater range of doses and a more considered set of patterns:

(a) single pixel squares with 9 equispaced doses from 5000 to 45000$\mu C/cm^2$
(b) double pixel squares with 9 equispaced doses from 5000 to 45000$\mu C/cm^2$
(c) triple pixel squares with 9 equispaced doses from 5000 to 45000$\mu C/cm^2$
(d) six pixel squares with 9 equispaced doses from 1000 to 9000$\mu C/cm^2$
(e) eight pixel squares with 9 equispaced doses from 1000 to 9000$\mu C/cm^2$
(f) ten pixel squares with 9 equispaced doses from 1000 to 9000$\mu C/cm^2$
(g) drawn circles with 70nm, 90nm, 110nm, 130nm, 150nm, 165nm, 190nm, 215nm and 240nm diameters using doses of 500$\mu C/cm^2$, 1000$\mu C/cm^2$ and 1500$\mu C/cm^2$.

For reproducibility assessment, nine of each of these were fabricated so that a reasonable estimate of the standard deviation could be obtained. The nine equispaced doses for the square patterns and the three equispaced doses for the circular patterns were achieved by making use of overlaps, *i.e.* by placing a copy of the pattern on top of itself from zero to eight times, doses of 1 to 9 times the basic amount can be obtained (provided the pattern is fractionated with the overlaps facility turned on in the CATS software).

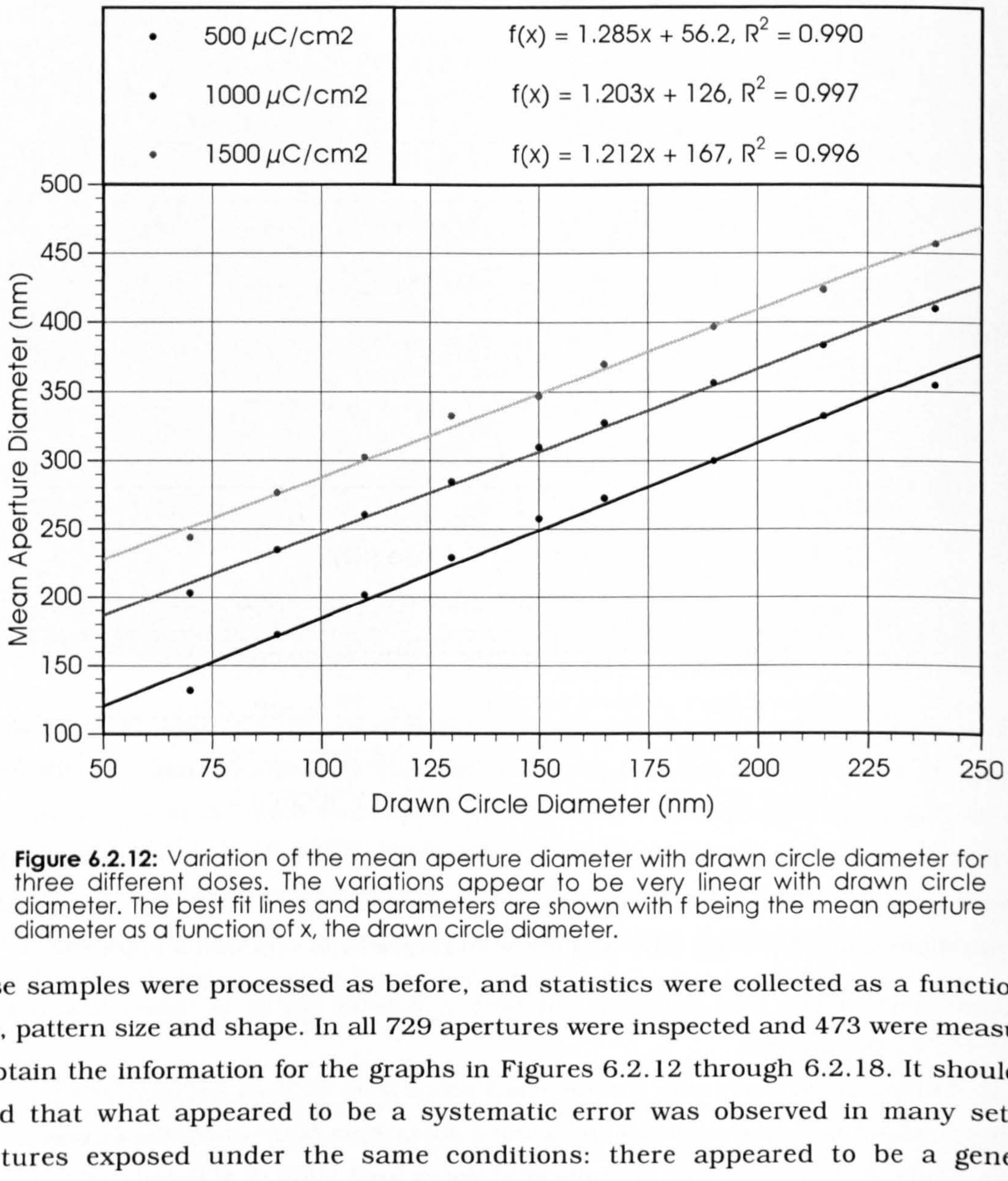


Figure 6.2.12: Variation of the mean aperture diameter with drawn circle diameter for three different doses. The variations appear to be very linear with drawn circle diameter. The best fit lines and parameters are shown with f being the mean aperture diameter as a function of x, the drawn circle diameter.

These samples were processed as before, and statistics were collected as a function of dose, pattern size and shape. In all 729 apertures were inspected and 473 were measured to obtain the information for the graphs in Figures 6.2.12 through 6.2.18. It should be noted that what appeared to be a systematic error was observed in many sets of apertures exposed under the same conditions: there appeared to be a general, approximately linear, increase in aperture size from one side of the silicon nitride beams to the other. The cause of this was not known but may have been due to effects such as charging [Liu, Ingino and Pease[30] (1995)], drift of the electron beam focal point with time or gradual defocusing due to exposure at the edge of a scan field [Thoms[25] (1993-1996)].

However, since most of these effects would not occur to such an extent when isolated apertures were being exposed on an actual device substrate, it was decided that where an adequate correlation could be established, the error be subtracted off from the

measurements. [A line was fitted by least squares to the aperture diameter as a function of position. The value given by this line at each position was then subtracted off from the aperture diameter at that position. A fixed offset, equal to the mean aperture diameter obtained from the raw data, was then added to the resulting values.]

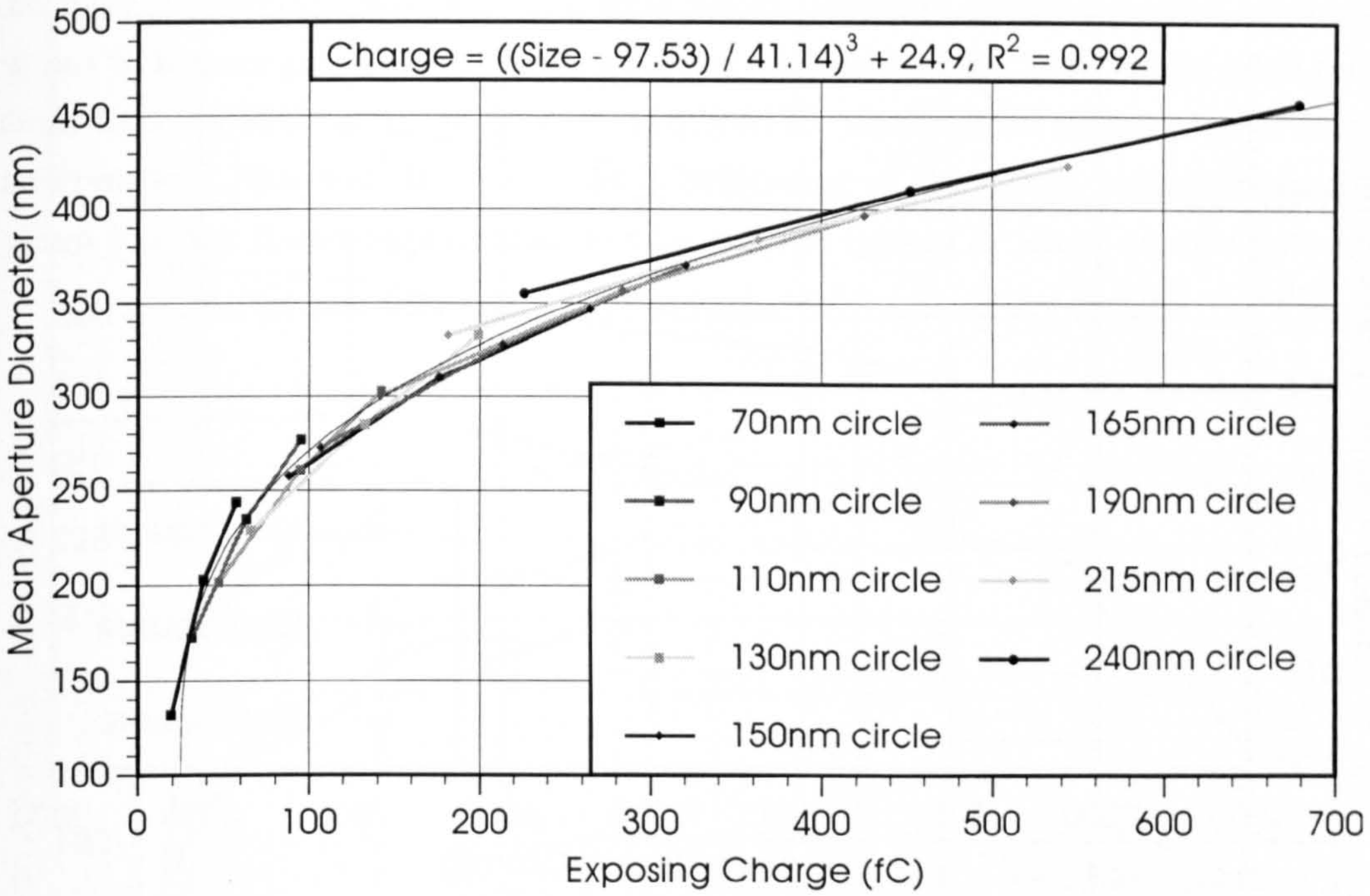


Figure 6.2.13: Variation of the mean aperture diameter with the total deposited charge and the drawn circle diameter. The deposited charge appears to depend on the cube of the mean aperture diameter. The best fit line and parameters are shown.

Figure 6.2.12 shows how the mean aperture diameter varies with the drawn diameter of the circle patterns for three different doses. As can be seen, the aperture size achieved appears to increase linearly with drawn circle diameter for a given dose.

Figure 6.2.13 shows how the mean aperture diameter varies with the total charge deposited and the drawn circle diameter. To a reasonable approximation, the total charge required appears to depend cubically on the aperture size achieved irrespective of the drawn circle diameter. However, too much should not be read into this observation in the absence of more data, as only three doses were exposed for each circular pattern diameter.

Figure 6.2.14 shows how the mean aperture diameter varies with dose and drawn square size for the three larger square sizes. As expected, the aperture size achieved increases with drawn square size and dose. Also plotted are the variabilities in aperture size. These are simply estimated as the standard deviations of the measurements.

Figure 6.2.15 shows how the mean aperture diameter varies with dose and drawn square size for the three smaller square sizes. As expected, the aperture size achieved increases with dose. However, most unexpectedly it can be seen that the 15nm drawn square pattern tends to produce smaller apertures than the 10nm drawn square pattern. The

reason for this is not known, but the author suspects that this is an anomalous result, caused by, for example, some spurious calibration error in the beamwriter during writing. Note that no apertures were formed with the 5nm square pattern for doses below 20,000 $\mu C/cm^2$, presumably, due to underexposure.

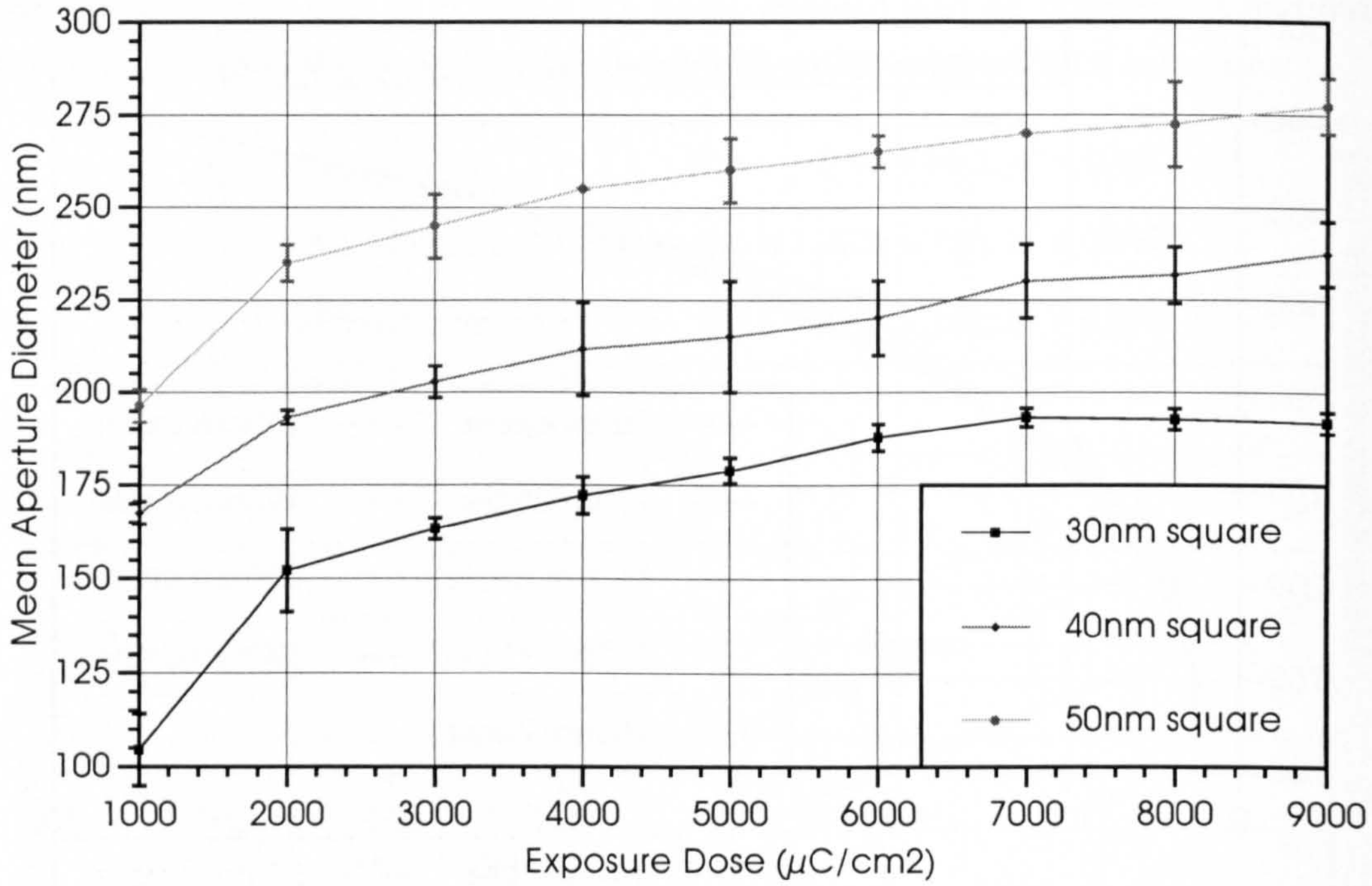


Figure 6.2.14: Variation of mean aperture diameter with dose and drawn square size for three sizes from 30nm to 50nm. The variability in aperture size is displayed as an error bar representing a one standard deviation range. Dividing a square size by 5nm gives its width in pixels.

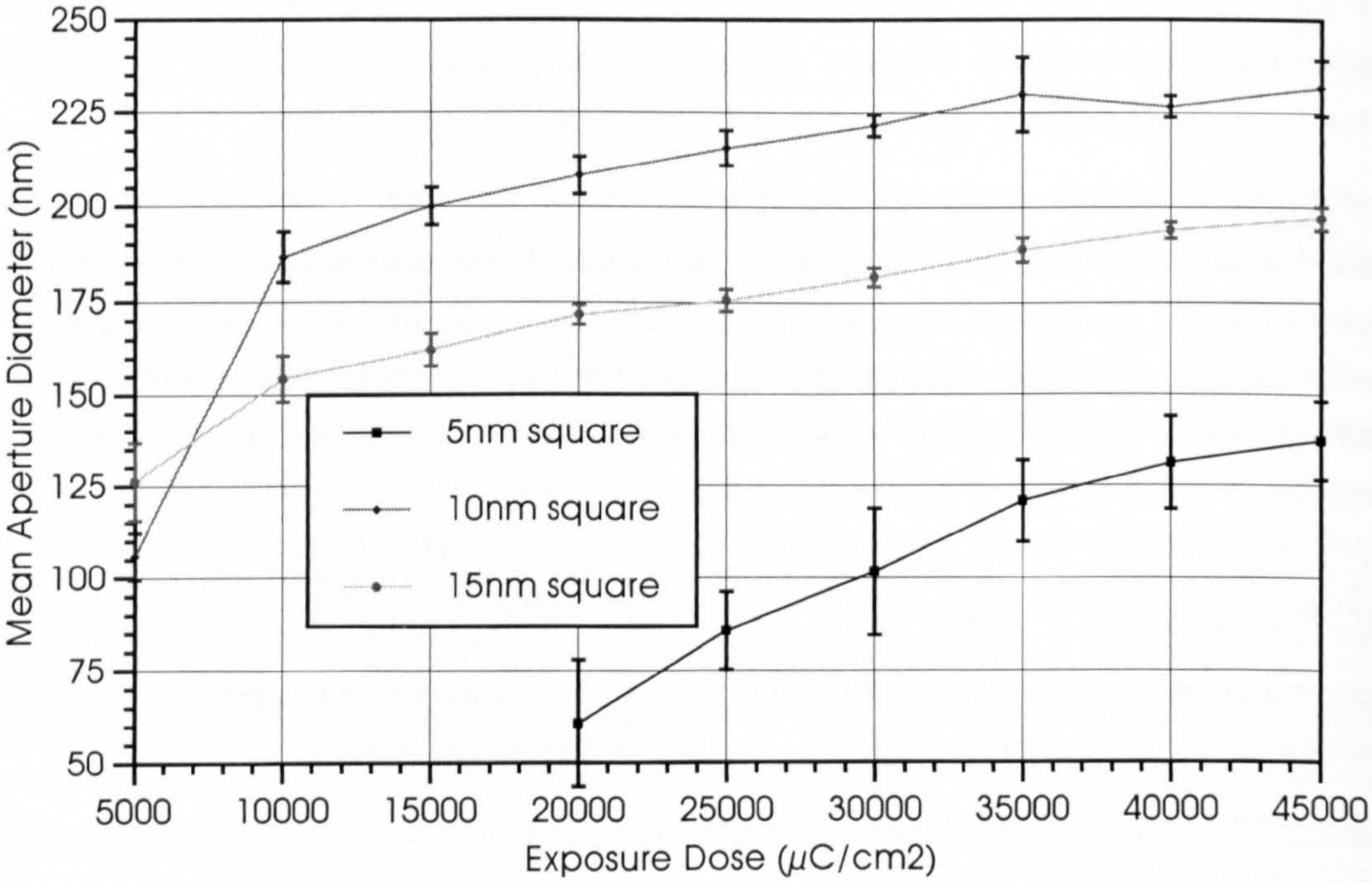


Figure 6.2.15: Variation of the mean aperture diameter with dose and drawn square size for three sizes from 5nm to 15nm. The variability in aperture size is displayed as an error bar representing a one standard deviation range. Dividing a square size by 5nm gives its width in pixels.

Figure 6.2.16 shows how the mean aperture diameter varies with the total charge deposited and the drawn square size. Here the cubic dependence of the charge deposited on the aperture size is not so clear-cut, as seen in Figure 6.2.13, and certainly does not appear to be independent of the drawn square size. With the range of doses used here, the increase in aperture size with the number of electrons deposited have all begun to flatten out to a more gradual variation, with the exception of the 5nm single pixel pattern. It is clear that a greater range of doses is required for this pattern to complete a more full characterisation. However, the anomalous behaviour of the 10nm square pattern may also mean that the flattening out observed here is not typical of 10nm square patterns.

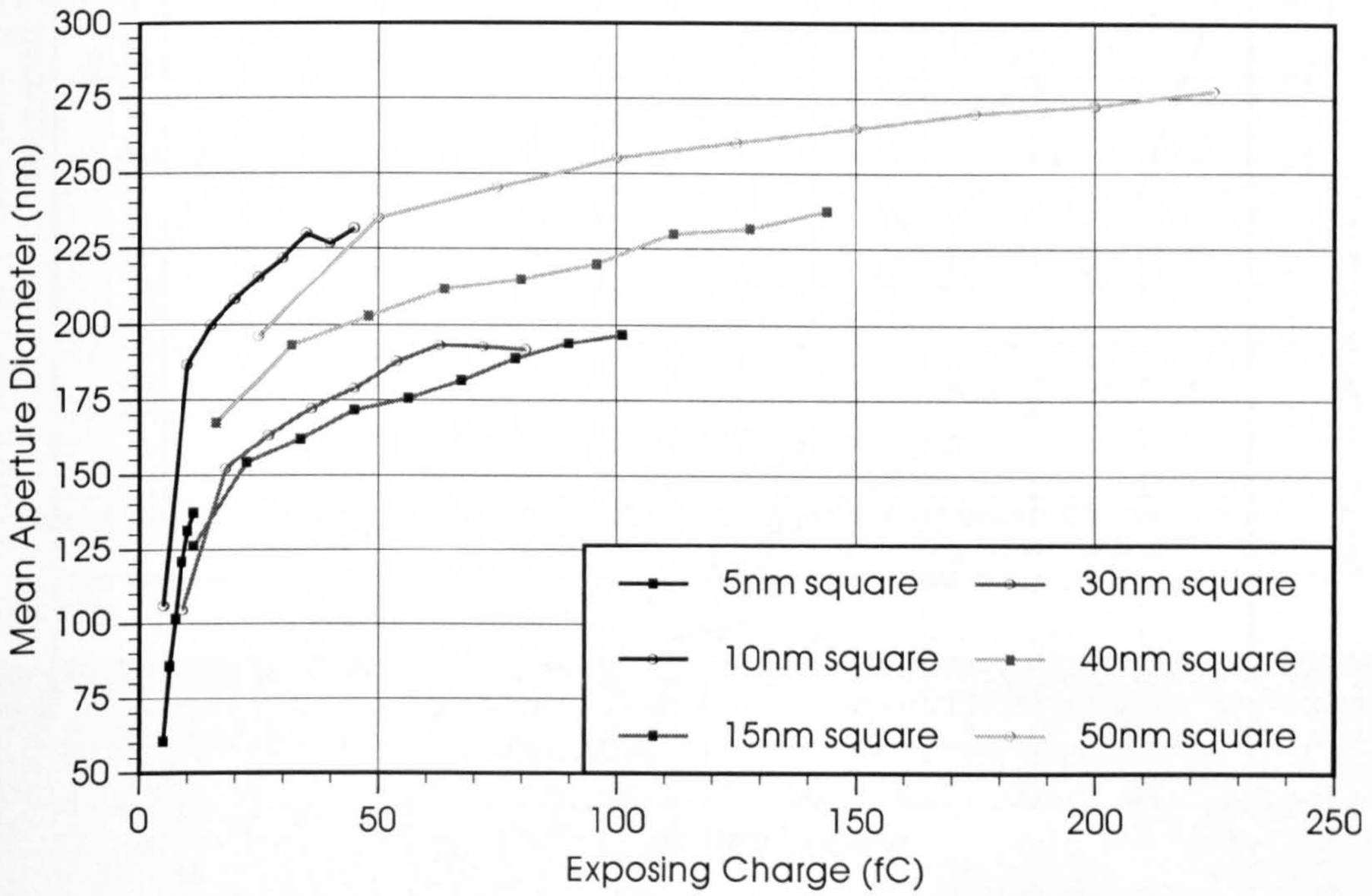


Figure 6.2.16: Variation of mean aperture diameter with total deposited charge and drawn square size. Dividing a square size by 5nm gives its width in pixels.

Figures 6.2.17 and 6.2.18 shows how the standard deviation varies with the pattern style and size used and the mean aperture diameter. It seems clear that the square style patterns and the three smallest circular patterns (Figure 6.2.17) behave in a different way to the larger circular patterns (Figure 6.2.18): the latter (with the exception of the largest circular pattern) seem to exhibit a 'peak' in the standard deviation as a function of mean aperture diameter not seen in the former. However, it is difficult to draw any conclusions without exposing the circular patterns over a wider range of doses.

These results were quite promising, despite the overcut profile, with the maximum aperture size variability being less than 30% at an aperture diameter of 60nm. The aperture size variability was more typically less than 10% for apertures 100nm or larger in diameter. At this point it was felt that if the overcut profile issue could be resolved and a better etch devised, then the aperture size variability for apertures less than 100nm in diameter would probably also improve. However, the need to develop the aperture fabrication process further was not a choice but a necessity as the overcut profile

problem would only be expected to get worse when the thicker silicon nitride layers of the actual contact mode AFM cantilevers came to be etched. The possibility of thinning the silicon nitride around the apex of the pyramid prior to defining the aperture was also briefly considered, however this would only decrease the degree of overcutting and not eliminate it, so a more fundamental solution to the problem was required.

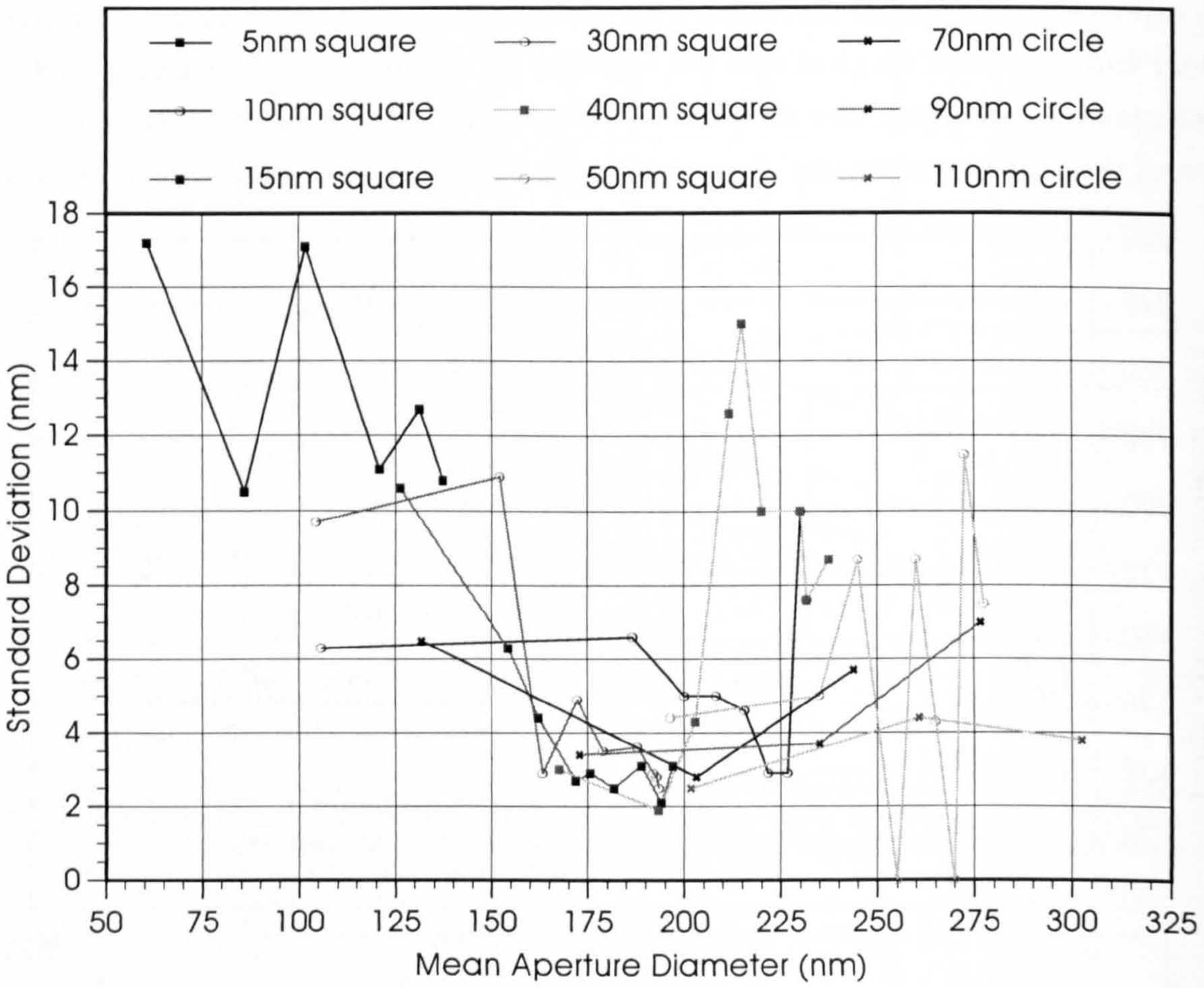


Figure 6.2.17: Variation of the standard deviation in the aperture diameter with the mean aperture diameter and the pattern used for the square and three smallest circular patterns. The points associated with each pattern style and size have been connected together by lines for clarity. Dividing a square size by 5nm gives its width in pixels.

One of the most encouraging aspects of the results achieved so far was that the fabrication of apertures at diameters around 50nm, whilst difficult, did not look implausible. Figure 6.2.19 shows two micrographs of one of the smallest apertures formed on the test sample. From Figure 6.2.19b the aperture can clearly be seen to have cleared out and be just under 60nm in diameter (this micrograph was taken under conditions of high brightness for this purpose). From Figure 6.2.19a the region over which the overcut profile is developed can be seen to extend laterally to a diameter of over 130nm which is more than twice the aperture opening. Finally, it is interesting to note that the overcut profiles achieved here in silicon nitride by etching directly from PMMA have been used to advantage by another group for the fabrication of thin film metal nanobridges [Ralls, Buhrman and Tiberio[31] (1989)].

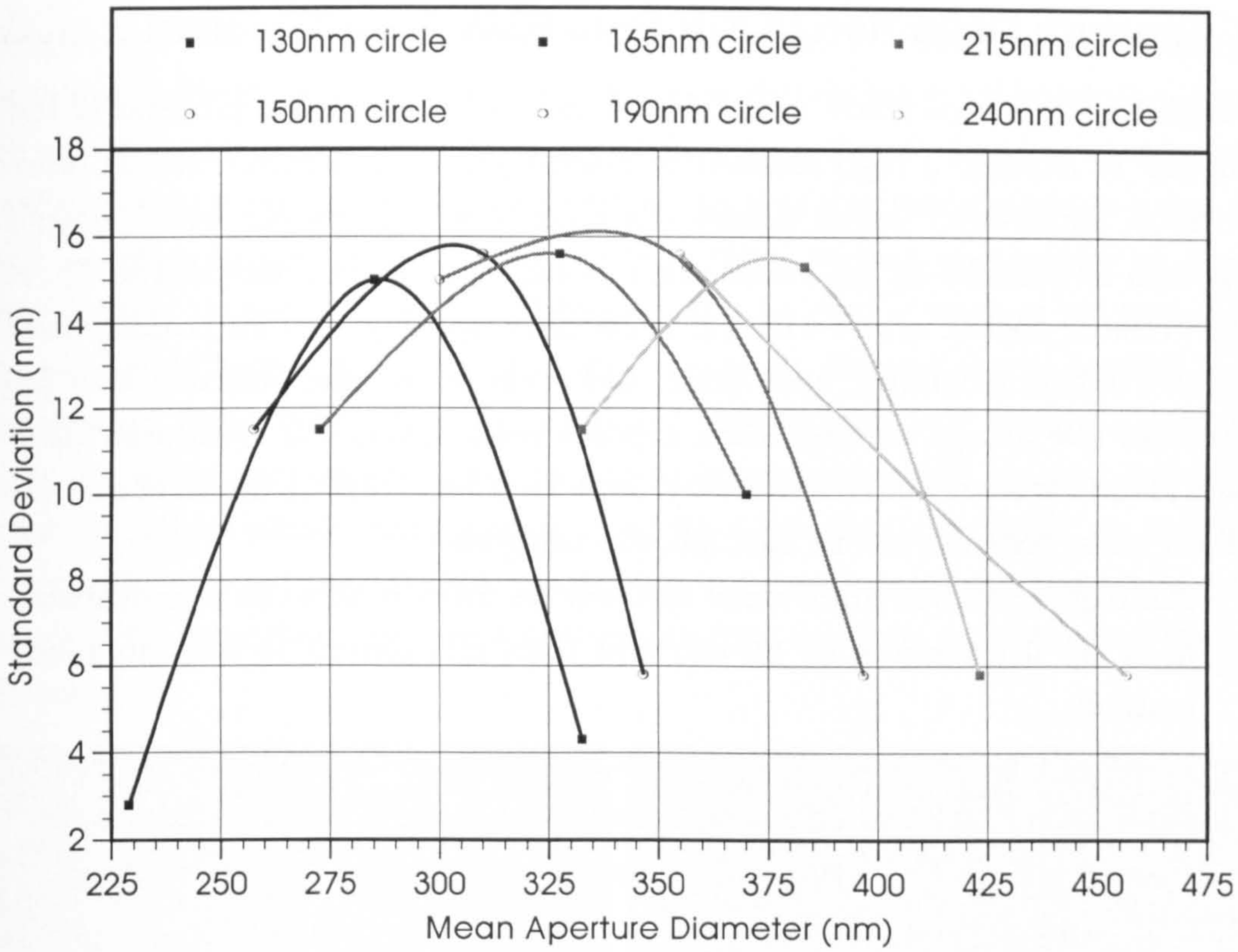


Figure 6.2.18: Variation of the standard deviation in the aperture diameter with the mean aperture diameter and the pattern used for the six largest circular patterns. The points associated with each pattern size have been connected together by lines for clarity.

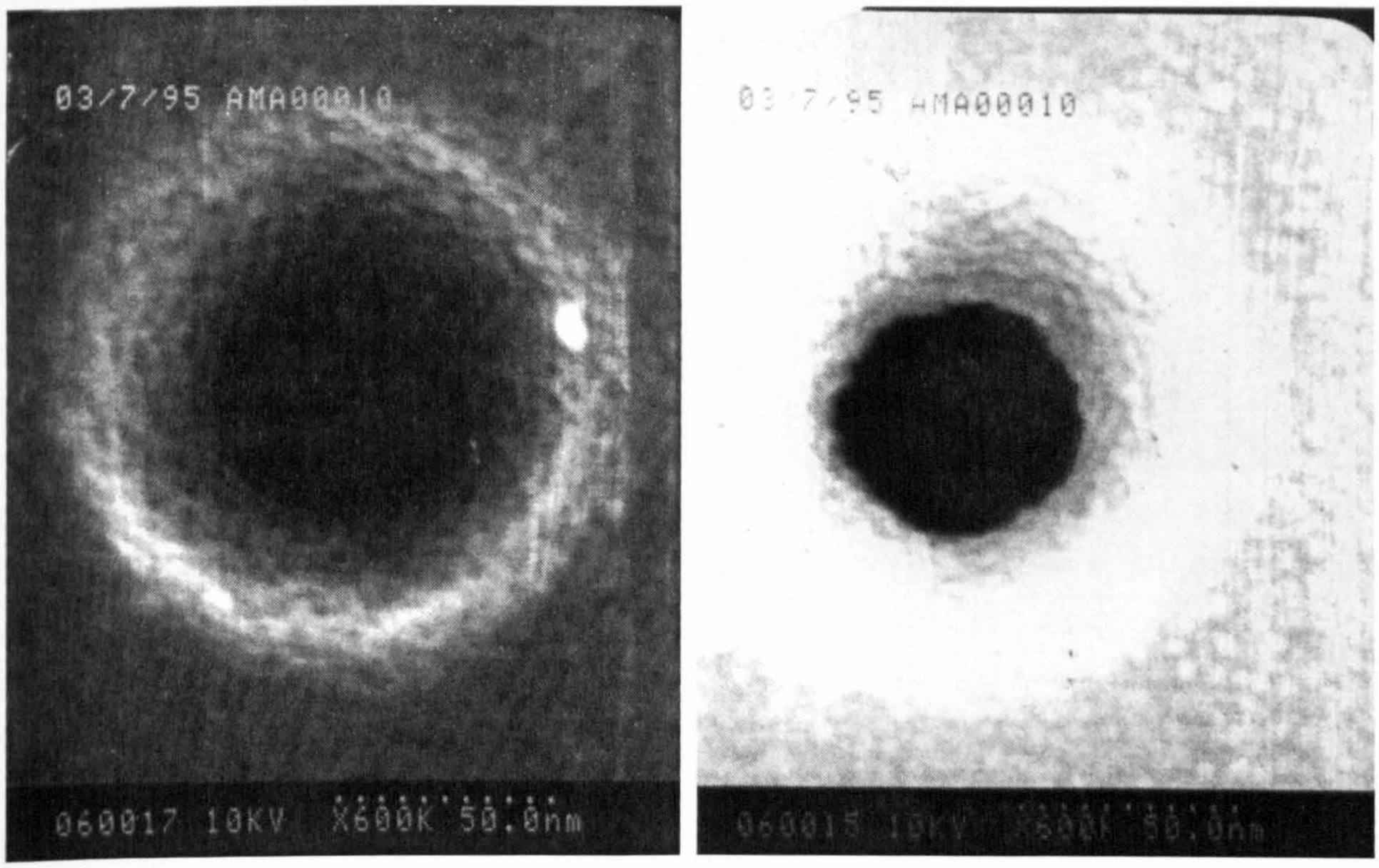


Figure 6.2.19: Micrograph of the smallest aperture fabricated on a 50nm thick free-standing silicon nitride 'bridge' by direct dry etching from PMMA ('ama0010'). The sample was coated in a 10nm thick nichrome charge conduction layer. (a) Left Image: normal brightness micrograph; (b) Right Image: high brightness micrograph showing that the aperture has cleared out.

6.2.2 Aperture Definition in SiN from PMMA via Titanium

One suggestion that was made to tackle the overcut resist profile problem of the last section was to develop a high resolution 'multi-layer resist scheme' (MLR) [Murad[16] (1993-1996)]. Multi-layer resist schemes are commonly employed where deeper dry etching than is afforded by the resist alone is required [Moran and Mayden[32] (1980); Howard[33] (1981); Jackel, *et al.*[34] (1981)]. These have typically been for processes such as bulk silicon micromachining, zone plate and *x*-ray mask manufacture, and pattern transfer from low voltage electron beam exposed resist layers [*e.g.* Matthies, David and Thieme[35] (1993); Muller, Eib and Faure[36] (1993); Lo, *et al.*[37] (1995); Stern, *et al.*[38] (1995)]. Multi-layer resist schemes can be split into two categories:

(1) The high resolution (thin) resist features are transferred to an intermediate layer by dry etching or lift-off, and then transferred to the intended material.

(2) The high resolution (thin) resist features are transferred to an intermediate layer by dry etching, then transferred to a thicker and/or more resilient resist layer by dry etching with oxygen, before finally being transferred to the intended material.

These schemes work by optimising the etch chemistry and conditions for each material independently. For category (2) schemes, titanium is commonly used as an intermediate layer due to the low sputtering yield of its oxide in oxygen etches – aluminium and nickel have been reported to offer between two and five times less resistance than titanium [Cantagrel[17] (1975); Unger, *et al.*[39] (1988); Juan and Pang[40] (1994)]. Several hundred to one selectivities of resist over titanium can be achieved [Juan, *et al.*[40] (1994)].

However, because of the poor etching resistance of PMMA, or any similar high resolution resist, as a masking layer for dry etching titanium, the thickness of the titanium layer has been limited to between 5nm and 15nm [*e.g.* Unger, *et al.*[39] (1988); Schneider, Schliebe and Aschoff[41] (1995)]. This then limits the maximum thickness of the second resist layer and leads to the problem of pinholes in the thin metal layer being transferred to the resist beneath. In many cases, the use of PMMA has been avoided by compromising the resolution in order to be able to use a more resistant electron-beam resist, or else, the more easily patternable, but generally less etch-resistant, germanium has been used as a substitute for titanium. Obviously, where a more resistant resist can be used to etch the titanium, a thicker layer of titanium can be used. However, for reasons of resolution and development time, a move away from PMMA was undesirable.

Therefore when it was suggested that there might be a way of devising an etch to etch titanium directly from PMMA with reasonable selectivity [Murad[16] (1993-1996)], it was decided that development proceed along these lines. The titanium layer would be used to etch a thicker resist layer beneath, which would then be used as a mask for etching the silicon nitride layer.

Photoresist and polyimide are commonly used as the second resist layer in category (2) schemes due to their good etch resistance to many dry etches (high resolution features can be transferred into these resists by dry etching even if such features could never be exposed directly). Polyimide had to be ruled out as it is a negative electron-beam resist [Patrick, Mackie, Beaumont and Wilkinson[42] (1985)], leaving photoresist as the only possible choice. However, then the problem would be that a method for float coating photoresist would have to be devised and time was short. Instead, it was decided that the thicker second resist layer simply be a thicker, higher molecular weight PMMA layer. However, for initial process development, the titanium layer could be used directly as a mask for the silicon nitride etch, turning this into a category (1) scheme. As titanium is more dry etch resistant than PMMA, conversion to a category (2) scheme would only be necessary if the maximum thickness of titanium that could be etched from PMMA was insufficient.

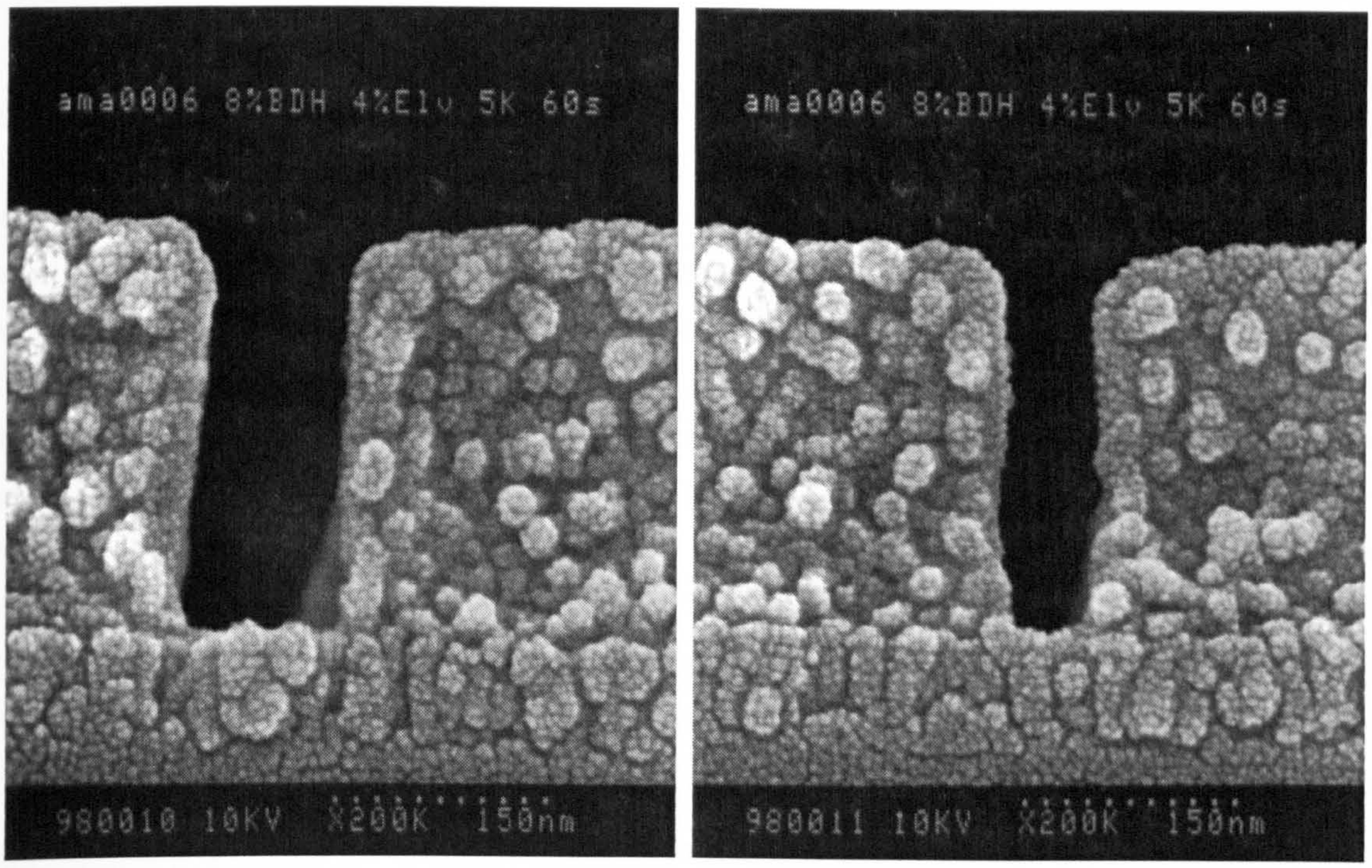


Figure 6.2.20: Micrographs of the smallest resist lines used for dry etch development ('ama0006'). The sample was cleaved through before being sputtered with about 20nm of gold/palladium as a charge conduction layer. The samples consisted of a resist layer about 300nm thick on a 75nm titanium layer, which in turn was on top of a 100nm thick LPCVD silicon nitride layer. (a) Left Image: resist line around 100nm wide; (b) Right Image: resist line around 50nm wide.

For development of dry etching processes, it is normal to measure etching progress by looking at trenches which have been cleaved through edge-on using the Hitachi S900 SEM [§3.6.3]. A sample, based on *n*-type silicon wafers with 100nm of LPCVD silicon nitride as used for the fabrication of the AFM probes [§4.1.1], was devised. The pattern used consisted of 500μm long wires, one to five pixels wide (based on 10nm squares, fractionated with 10nm resolution), and repeated ten times each at 1μm pitch, with a short grating of 250nm wide wires at either side to enable these lines to be found easily

in the SEM. [It should be noted that the dry etching aspects of the following development work was closely supervised by Dr. S. K. Murad.]

The first test samples were prepared in the following manner:

(a) 8% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(b) The resist was dried by baking at 180°C for one hour.

(c) 4% Elvacite was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(d) The resist was dried by baking it overnight at 180°C.

(e) The fine line patterns were exposed using a resolution of 10nm and a spot of 12nm at 100kV, with no alignment. The dose used was 1500μC/cm^2.

(f) The grating patterns were exposed using a resolution of 25nm and a spot of 56nm at 100kV, with no alignment. The dose used was 750μC/cm^2.

(g) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds (this high contrast developer ensures that features are cleared out at the possible expense of lower resolution).

The designs were placed at the centre of 5 x 6mm areas, so that cleaving through the lines would give two pieces, each 5 x 3mm. [For the Hitachi S900 SEM, with the holder used for mounting samples vertically, the sample can only be focused over a narrow range of about 0.5mm so that the sample width must be controlled. A width of 3mm is usually optimal.] The substrate was 23 x 24mm so that an array of 4 x 3 test samples could be exposed at one time. When cleaving, the sample was scribed midway along and perpendicular to one of the 6mm edges for a short distance from that edge (typically 1mm to 1.5mm), and then placed at the edge of a glass slide before pressure was applied along the scribed line using the rear end of a pair of metal tweezers. Another pair of tweezers was used to clamp the sample firmly onto the glass slide. In this way, the scriber does not damage the resist or etched lines by being dragged through it. When resist was imaged, the sample was typically sputter coated with 10nm to 20nm of gold/palladium along its edge after cleaving, to act as a charge conduction layer.

The pattern design, resist, exposure and development conditions used resulted in PMMA lines ranging from about 50nm to 250nm in width. Figure 6.2.20 shows micrographs of the two smallest width lines in resist. Here, prior to the resist coating, 75nm of titanium had been evaporated onto the silicon nitride surface using the Plassys MEB 450 evaporator.

It is well known that titanium can be etched in both fluoride [Dagostino, Fracassi, Pacifico and Capezzuto[43] (1992)] and chloride [Dagostino, Fracassi and Pacifico[44] (1992)] chemistries. For thin layers, some typical chloride etches include BCl_3, CCl_4, $SiCl_4$ and various mixtures of these with O_2, N_2 and Cl_2, while some typical fluoride etches include CF_4 and SF_6 as well as mixtures of these with O_2 and N_2. Fluoride chemistries tend to erode resist rapidly, while chloride chemistries are slow at breaking down the titanium oxide layer on the surface [Dagostino, *et al.*[44] (1992)].

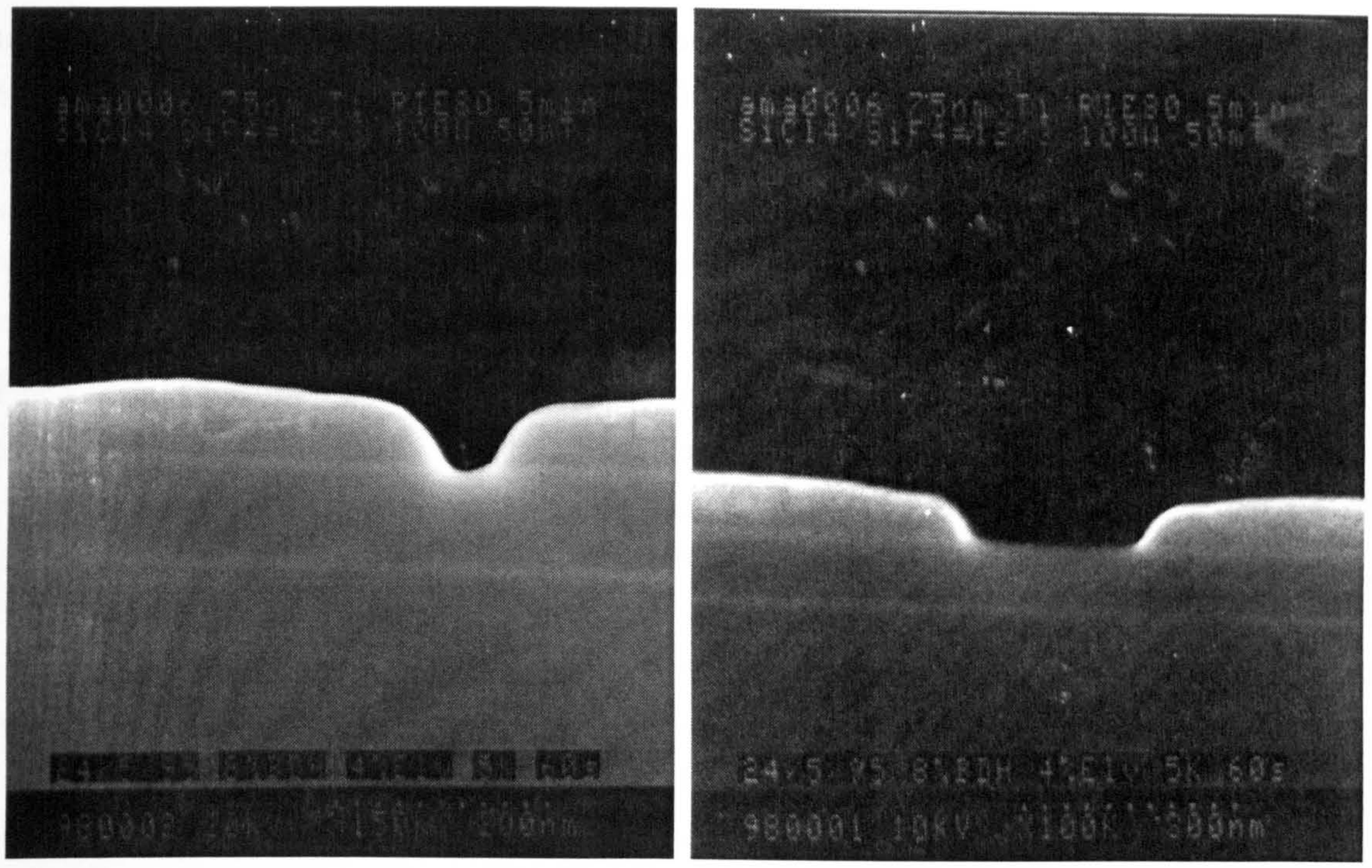


Figure 6.2.21: Micrographs of etched titanium profiles during dry etch development ('ama0006'). The sample was not coated with a charge conduction layer. The PMMA has all been etched away and the titanium profiles are slightly overcut. The silicon nitride layer beneath can also be made out.

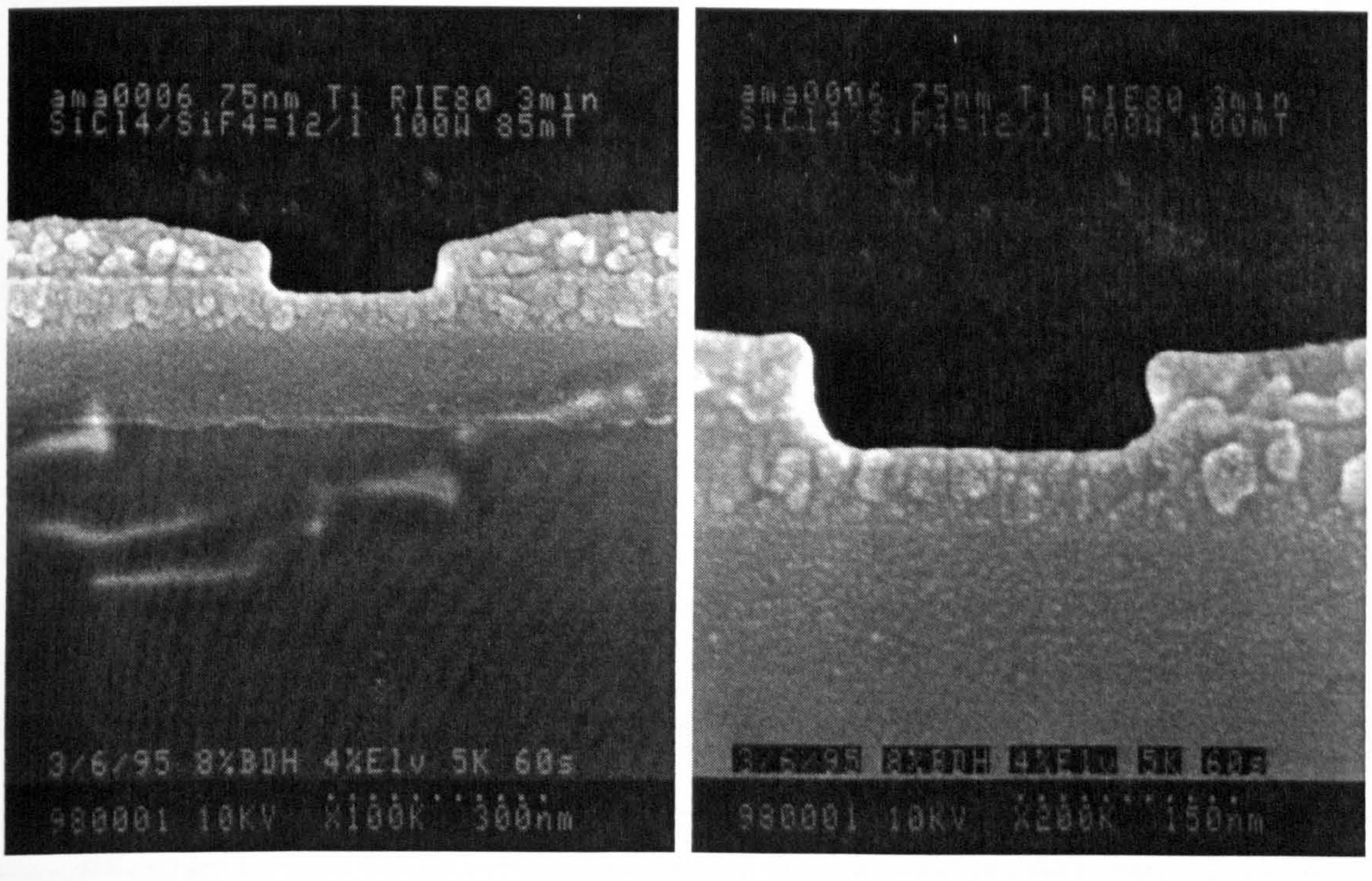


Figure 6.2.22: Micrographs of etched titanium profiles during dry etch development ('ama0006'). The sample was coated with a 10nm thick gold/palladium charge conduction layer. Some PMMA is still present and the titanium profiles are less overcut than in Figure 6.2.21. The silicon nitride layer beneath can also be made out. (a) Left Image: line etched at a pressure of 85mT; (b) Right Image: line etched at a pressure of 100mT.

The first suggestion [Murad[16] (1993-1996)] was to try a mixture of $SiCl_4$ and SiF_4 in the ratio 12 sccm to 3 sccm at a base pressure of 50mT in the Plasma Technology RIE80 etcher [§3.6.4] with an r.f. power of 100W and a bias of -150V for 5 minutes. These

samples were examined and it was found that no resist was left at the end of the etch, while the titanium had been etched by the full 75nm. However, the etch profile of the titanium appeared to be overcut, probably due to erosion of the resist [Murad[16] (1993-1996)]. Figure 6.2.21 shows micrographs of the results at two different linewidths. Reducing the etch time to 3 minutes, showed that the resist had all gone by this time while the titanium had only been etched about halfway through.

In an effort to reduce the resist erosion rate, a reduction in the SiF_4 content to a ratio of 1 sccm to 12 sccm $SiCl_4$ was suggested [Murad[16] (1993-1996)]. This resulted in the resist lasting for a full five minutes with the titanium etched to a depth of 70nm in this time. Again mask erosion resulted in an overcut etch profile in the titanium. Increasing the base pressure to 75, 85 and 100mT seemed to result in an increasing resist etch rate and better anisotropy but no increase in etch selectivity. Further experiments seemed to suggest that there was an induction time, probably due to the time it took to break down the titanium oxide layer on the surface, of between 1 and 2 minutes [Murad[16] (1993-1996)]. Figure 6.2.22 shows the etch depth at two different pressures after 3 minutes. The resist is still present in these images.

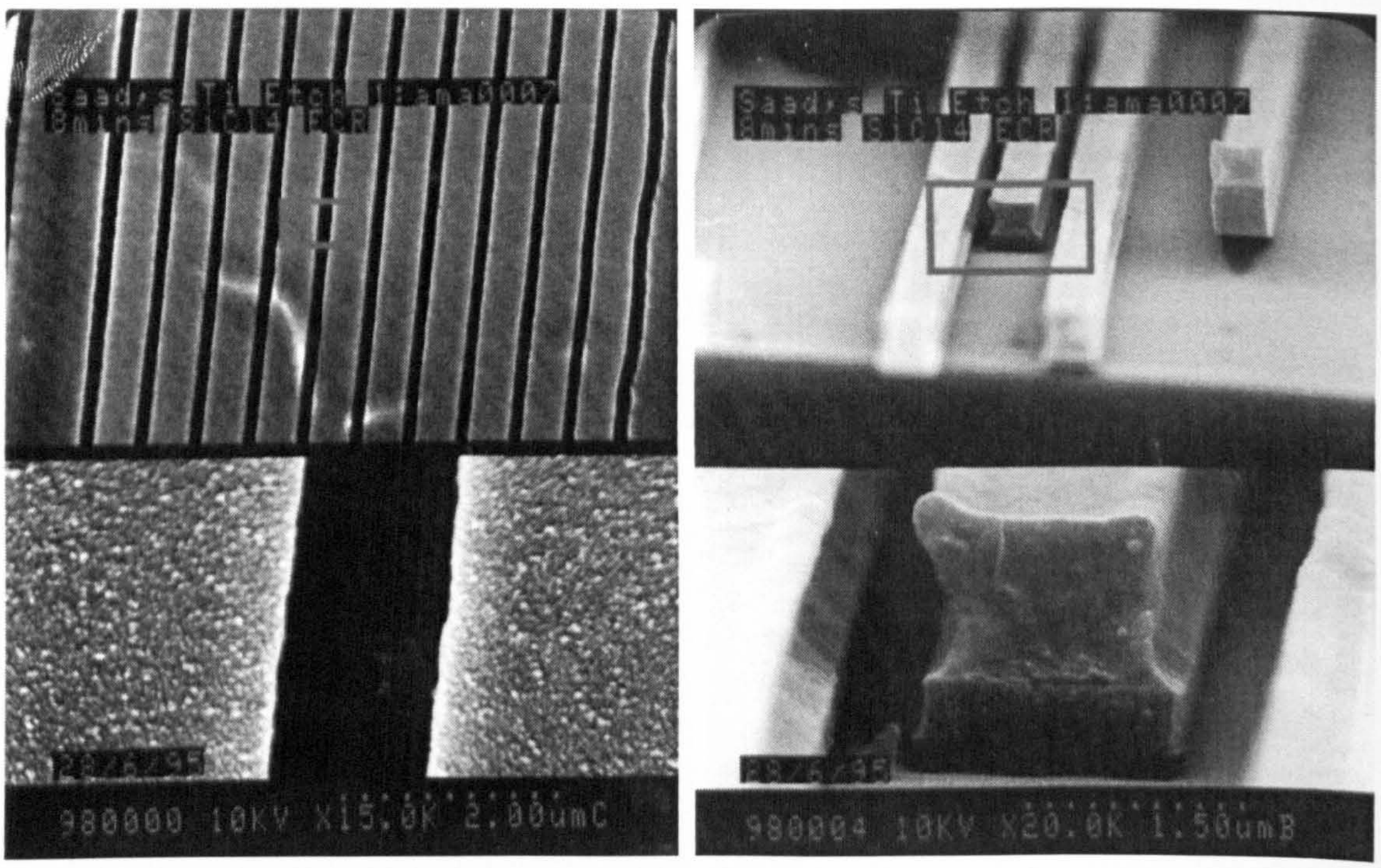


Figure 6.2.23: Micrographs of the first successfully etched titanium wires during dry etch development ('ama0007'). The sample was not coated with a charge conduction layer. (a) Left Image: top view of the etched lines. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: oblique view of the etched lines after cleaving through the sample. The PMMA resist can be clearly seen. The titanium wires lost adhesion with the silicon nitride beneath during the cleaving process. The bottom half is a 5x magnification over the displayed scale.

At this point it was decided that better results could be achieved in the ECR machine since operation at low pressure and bias would reduce resist erosion while the microwave power could be used to independently control ion energy [Murad[16] (1993-1996)]. New samples were prepared with three layers of resist instead of two, the extra layer being an

additional layer of 4% Elvacite (higher molecular weight than BDH and hence more dry etch resistant). As before, this layer was spun on at 5000 rpm for 60 seconds and a one hour, 180°C bake separated this layer from the previous layers with an overnight 180°C bake following this layer. To accommodate the increased resist thickness, the exposure dose of the fine lines was increased to 1700μC/cm^2. This time the thickness of titanium used was 70nm instead of 75nm. The development process and pattern design remained as before.

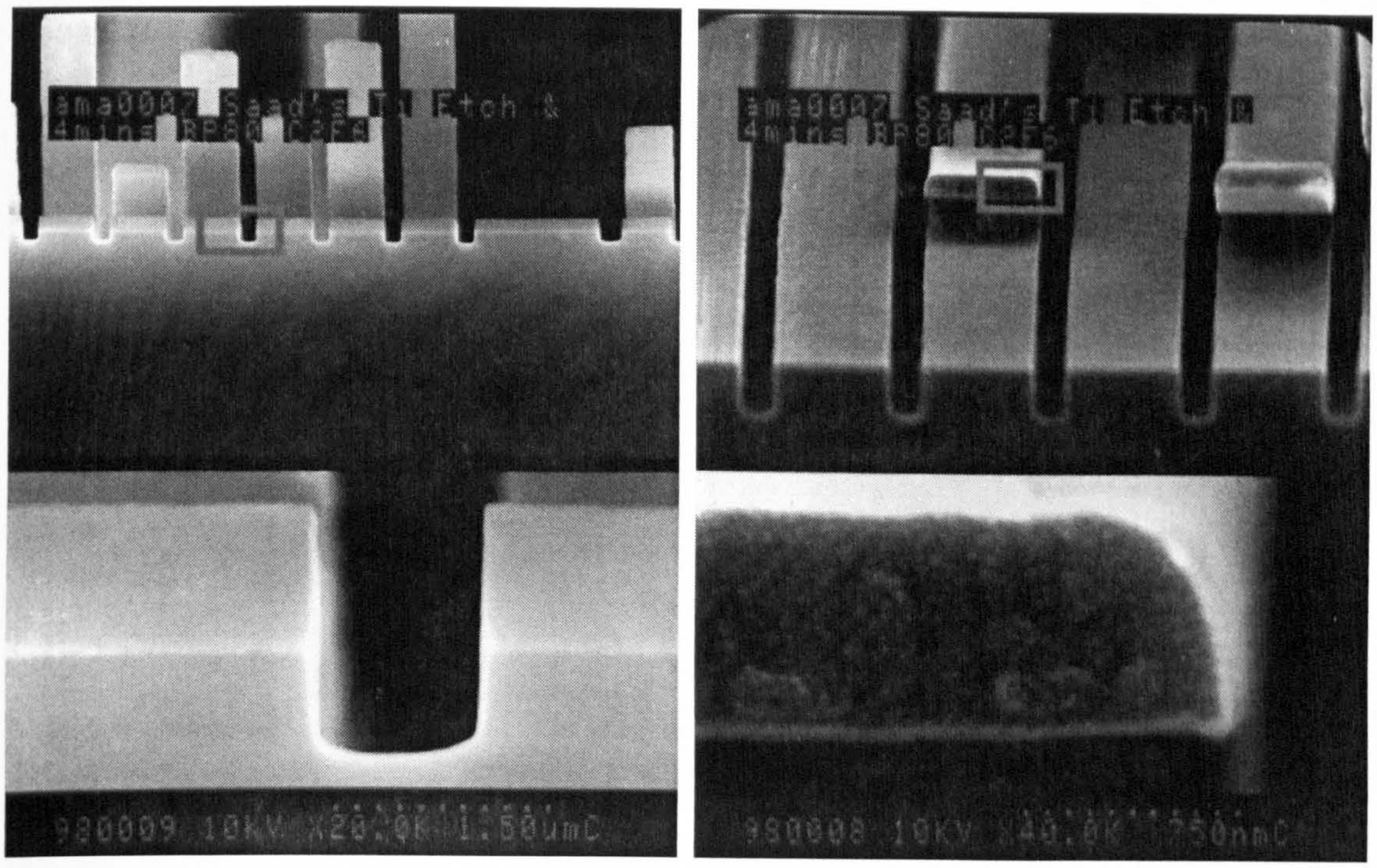


Figure 6.2.24: Micrographs of the first successfully dry etched silicon nitride wires using a titanium mask ('ama0007'). The sample was coated with a 10nm thick gold/palladium charge conduction layer. (a) Left Image: the titanium masking has lost adhesion with the substrate during cleaving but the silicon nitride layer has clearly been etched through along with some of the silicon beneath. The etch appears to be highly anisotropic; (b) Right Image: oblique view and close-up of one of the titanium wires used to mask the etch. The titanium has become rounded but otherwise does not appear to have been attacked by the silicon nitride etch. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

A first attempt was made using pure $SiCl_4$ at a pressure of 3mT, with 100W of microwave power, 25W of r.f. power and a sample bias of -80V. The titanium was found to have etched 35nm deep in 5 minutes of etching but the surface was extremely rough. This was presumed to be due to inhomogeneities in the thickness of the titanium oxide layer resulting in different areas of the titanium layer starting to etch at different times [Murad[16] (1993-1996)]. However, resist erosion was less than 70nm (shrinkage of the resist under the electron beam of the SEM, even after sputter coating with gold/palladium, only allows an upper limit to be placed on the resist erosion rate).

In an effort to break down the titanium oxide layer more quickly, it was decided [Murad[16] (1993-1996)] that the base pressure be increased to 4mT and the sample be etched with 150W of microwave power and 30W of r.f. power at a sample bias of -85V for 1 minute, followed by 7 minutes of etching with 20W of r.f. power and a sample bias of -57V. These

etching conditions proved to be extremely successful. Figure 6.2.23 shows the results of this etch. [Note that while cleaving through these titanium strips, the titanium wires tend to lose their adhesion with the silicon nitride surface beneath.] As can be seen from Figure 6.2.23a, the etching is very clean and from Figure 6.2.23b, the etch can be seen to be very anisotropic, with the resist having been eroded by less than 100nm for a 70nm titanium etch depth. This is a PMMA:Titanium etch ratio of better than 1.5:1.

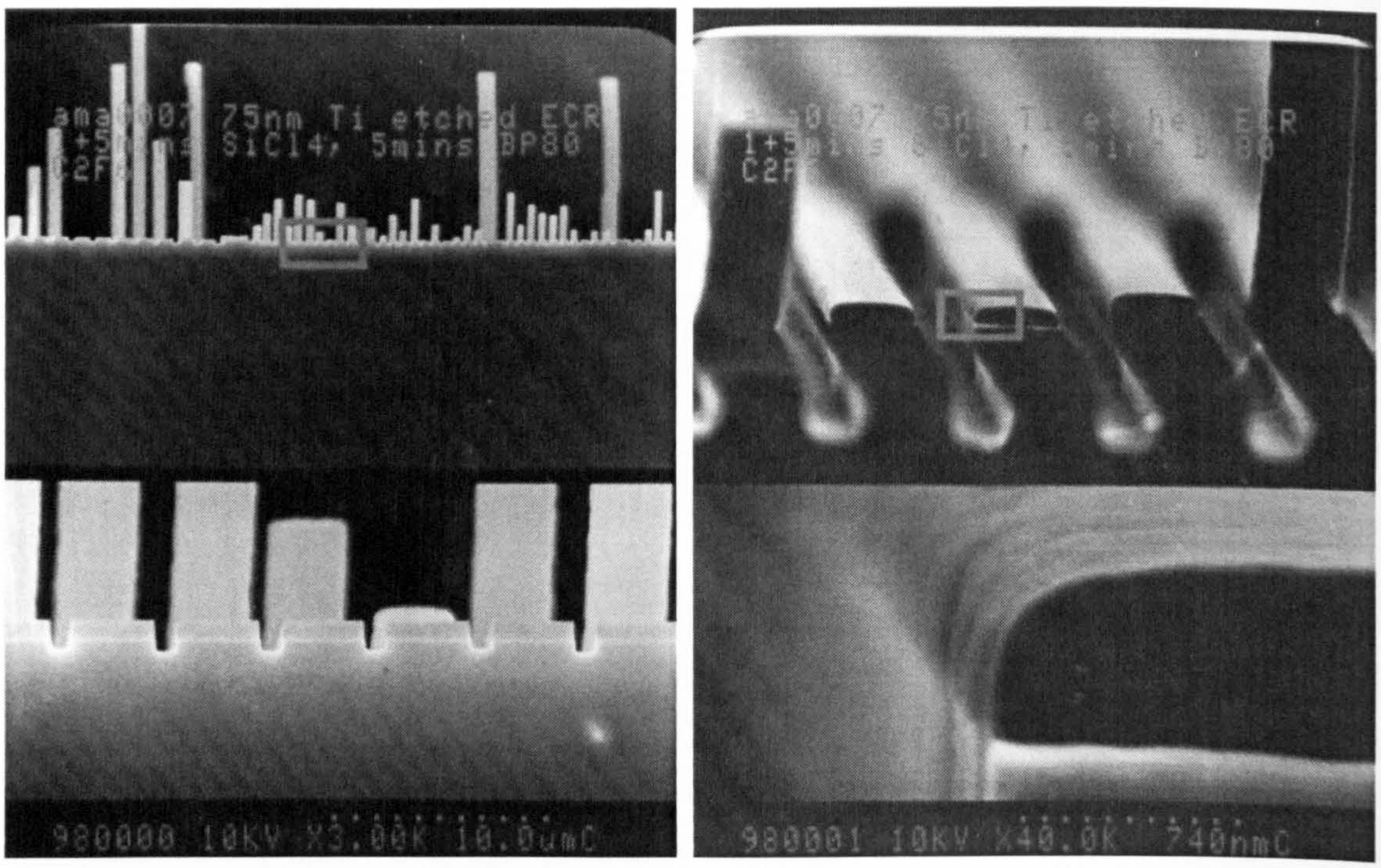


Figure 6.2.25: Micrographs of the first successfully dry etched silicon nitride wires using a titanium mask ('ama0007'). The sample was not coated with a charge conduction layer. The titanium etching time was reduced by 2 minutes. (a) Left Image: the titanium masking has lost adhesion with the substrate during cleaving but the silicon nitride layer has again been etched through along with some of the silicon beneath; (b) Right Image: oblique view and close-up of one of the titanium wires used to mask the etch. The titanium has again become rounded but otherwise does not appear to have been attacked by the silicon nitride etch. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

As mentioned previously, a category (2) scheme would have used the titanium mask to anisotropically etch a hardbaked resist layer beneath with oxygen before using this resist layer to etch the silicon nitride beneath in CF_4 or SF_6 (two common silicon nitride etches). However, since it is known that free fluorine radicals produced during etching attack titanium much more than CF_x radicals [Dagostino, *et al.*[43] (1992)] an attempt to directly etch the silicon nitride from titanium was made using C_2F_6 which should produce less free fluorine than either CF_4 or SF_6 [Turban[45] (1984)]. The sample was therefore etched in the RIE80 etcher with an r.f. power of 100W, a pressure of 15mT, a flow of 20 sccm and a self bias of -400V for 4 minutes. The results of this are shown in Figure 6.2.24. As can be seen the silicon nitride has been completely etched through along with about 75nm of silicon beneath and with very high anisotropy. From Figure 6.2.24b the effects of the silicon nitride etch on the titanium masking layer can be determined. The edges of the mask have become rounded, probably due to electrostatic effects acting so as to focus ion

bombardment onto the titanium vertices [Ingram[46] (1989)], but the thickness of the titanium has not detectably changed.

The experiment was repeated, but the second part of the ECR etching was reduced from 7 minutes to 5 minutes and the silicon nitride was etched under standard conditions in the BP80 etcher (15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 5 minutes). Figure 6.2.25 show the results of this etch. Again the titanium masking appears to have been barely touched and the silicon nitride has been etched all the way through.

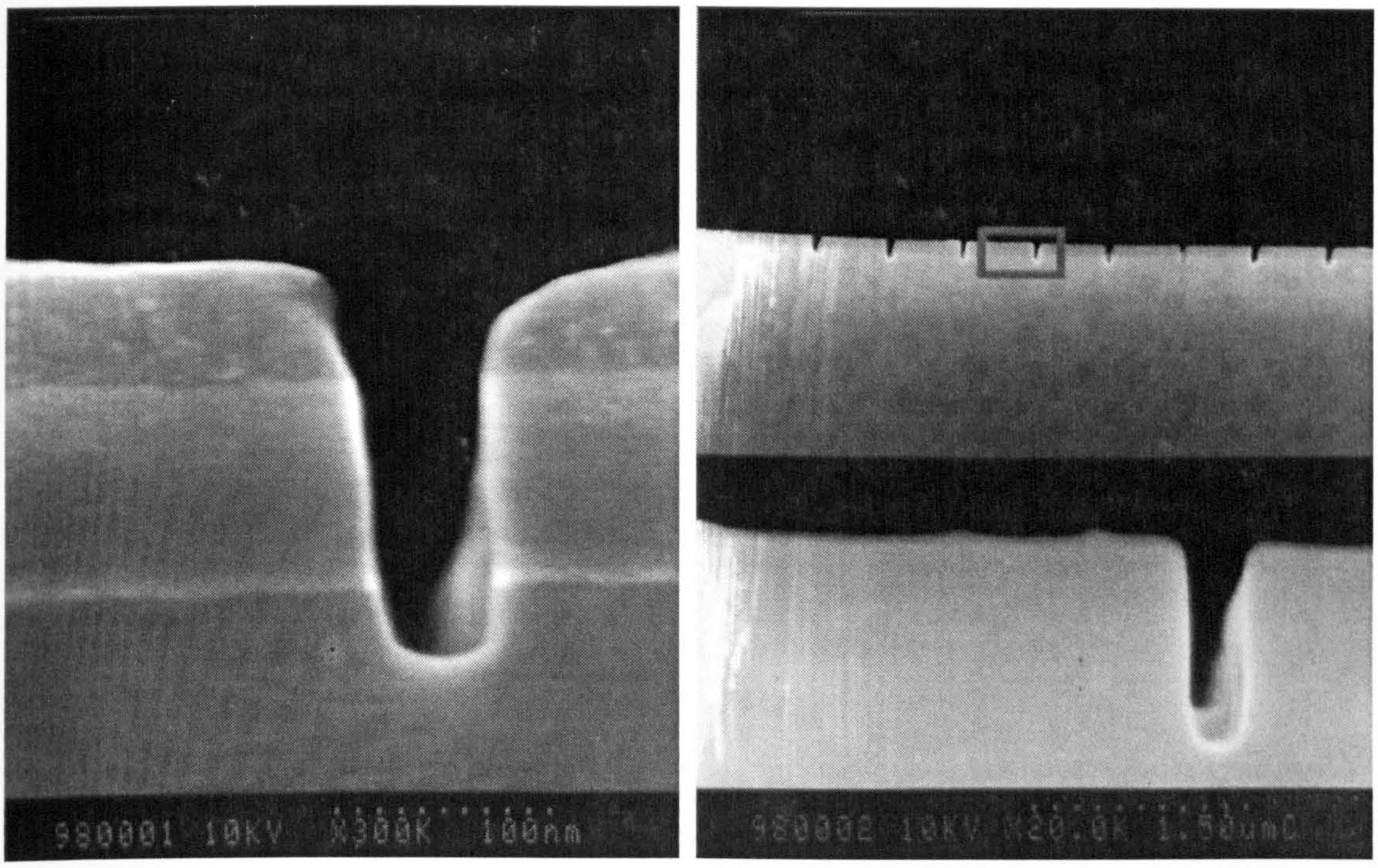


Figure 6.2.26: Micrographs of the narrowest silicon nitride wires achieved by dry etching through a titanium mask ('ama0007'). The titanium masking layer was reduced to 50nm thick and the resist was reduced to two layers of 4% Elvacite. (a) Left Image: this sample was coated with a 10nm thick gold/palladium charge conduction layer. The titanium mask can still be seen and the silicon nitride layer has been etched through along with some of the silicon beneath at a width of around 55nm; (b) Right Image: this sample was not coated with a charge conduction layer. An even narrower silicon nitride wire with a width of about 30nm. The titanium was stripped off prior to imaging. The bottom half is a 10x magnification over the displayed scale.

The experiment was repeated one more time, but this time using 50nm of titanium and reducing the resist layer used to simply two layers of 4% Elvacite. As before the second part of the ECR etching was of 5 minutes duration and the silicon nitride etching was carried out in the BP80 etcher, but this time only for 4 minutes. Figure 6.2.26 shows the results of this attempt. Figure 6.2.26a shows a 55nm line etched through 100nm of LPCVD silicon nitride and 35nm of silicon, while Figure 6.2.26b shows a 30nm line etched 150nm deep into the LPCVD silicon nitride and silicon beneath, giving an aspect ratio of 5:1.

From these results it was concluded that this new etching procedure was of good reproducibility and capable of adequate resolution for the purpose of SNOM probe

fabrication. That the silicon beneath is etched is not a problem for this process, and is actually desirable if the metal shielding is evaporated prior to release, since any metal deposited within the aperture would block the aperture unless removed during the release etch. If a sufficiently deep groove is made in the silicon, the metal deposited within the aperture will be attached to silicon and will be removed during the release etch when the silicon itself is removed, as illustrated by Figure 6.2.2.

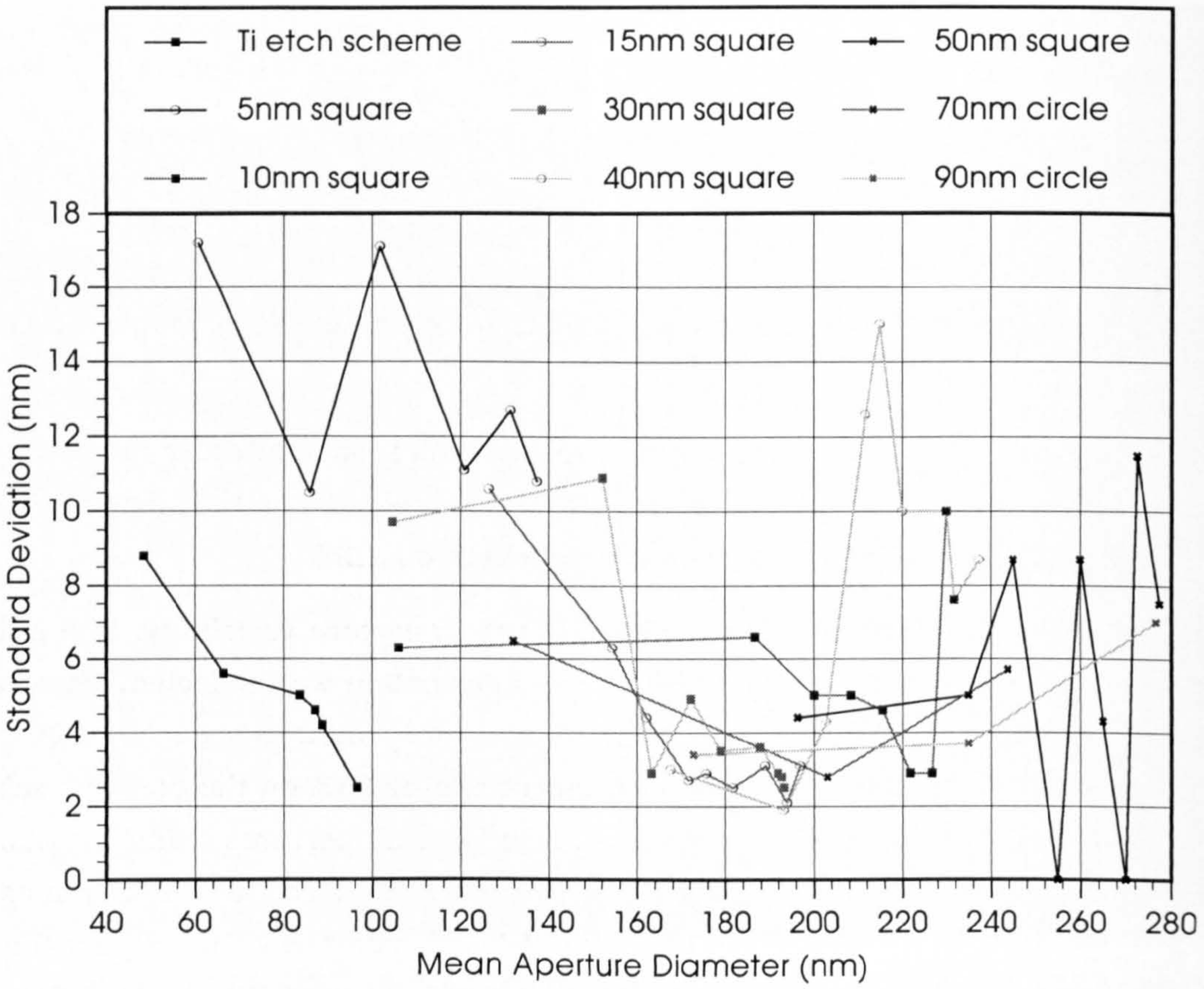


Figure 6.2.27: Comparison of the aperture diameter variability with the mean aperture diameter between the titanium etching scheme using a single pixel pattern and the direct etching scheme using the square patterns and two smallest circular patterns of Figure 6.2.17. The points associated with each pattern style and size have been connected together by lines for clarity.

Then it was decided that some statistics be collected for the variability in aperture size at the smaller aperture sizes made possible with this technique. Unfortunately, at this point the beamwriter machine was unavailable due to technical problems and so Dr. J. M. R. Weaver offered to expose a pattern using the Jeol 100 CXII STEM. As the Jeol 100 CXII is capable of higher resolution than the beamwriter [Thoms, *et al.*[28] (1989)], it would have been of interest to find out how small an aperture could be produced using this machine. As before the substrate was silicon coated with 100nm of LPCVD silicon nitride. The sample was processed as follows:

(a) 30nm of titanium was evaporated onto the sample using the Plassys MEB 450.

(b) 4% Elvacite was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(c) The resist was dried by baking it overnight at 180°C.

(d) The resist was exposed in the Jeol 100 CXII STEM at 100kV, with no alignment.

(e) The pattern was developed in MIBK:IPA 1:3 at 23°C for 60 seconds (this high contrast developer ensures that features are cleared out at the possible expense of lower resolution).

(f) The sample was dry etched in the ECR using the new titanium etch for 1 + 3 minutes (instead of 1 + 5 minutes for 50nm of titanium).

(g) The sample was dry etched again, this time with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 5 minutes in the Plasma Technology BP80.

(h) The sample was cleaned using the inorganic clean [App. D], with the titanium being stripped off during the nanostrip phase of this procedure.

(i) The sample was wet etched [App. G.1] in 7 molar (+5%) potassium hydroxide solution at 80°C for 90 minutes.

The changes in resist coating and titanium thickness were made in order to ensure that the first attempt succeeded and because a thinner resist would lead to higher resolution. These samples were then examined using the Hitachi S900 SEM.

Figure 6.2.27 shows the effect of this scheme on the aperture variability. The pattern consisted of single pixels. Unfortunately no dose information was collected. However, it can be seen that the variation of aperture size has been improved to better than 20% at a diameter of 50nm. This represents an improvement of over 33% on the previous scheme at a diameter of 60nm. Figure 6.2.28 shows some typical apertures from this sample. From Figure 6.2.28a the reproducibility of apertures at dimensions around 50nm can be qualitatively judged. Compare this with Figure 6.2.11. However, it would appear that the profile has once again become overcut. A close examination of Figure 6.2.28b which shows a tilted view of a 100nm diameter aperture reveals that the profile is vertical for the lower half (approximately) of the silicon nitride layer and overcut for the top half. That this had not been seen in the previous dry etch testing suggested that the cause was associated with the change in titanium thickness and resist thickness. A number of explanations were possible:

(a) It was noted earlier that the silicon nitride dry etching causes the edge of the titanium masking layer to become rounded. Thus, it may simply have been that this rounding extended to the surface of the silicon nitride much quicker due to the reduction in titanium thickness from 50nm to 30nm.

(b) The use of a single layer of resist may have meant that the resist had become sufficiently eroded during the titanium etching that an overcut profile had been formed part-way through the titanium. Then during the silicon nitride etching, the rounding effect of (a) had done the rest.

(c) In all the dry etching tests, the resist was never removed after the titanium etching, but left on even during the silicon nitride etching. This resist acted

so as to protect the titanium, albeit for only a limited time, from the full force of the silicon nitride dry etch. In this case the resist layer used had been halved and so only offered half the protection. This may not have been sufficient to prevent the titanium oxide layer breaking down part way through the silicon nitride etching, resulting in part of the silicon nitride profile being vertical, while the rest became overcut.

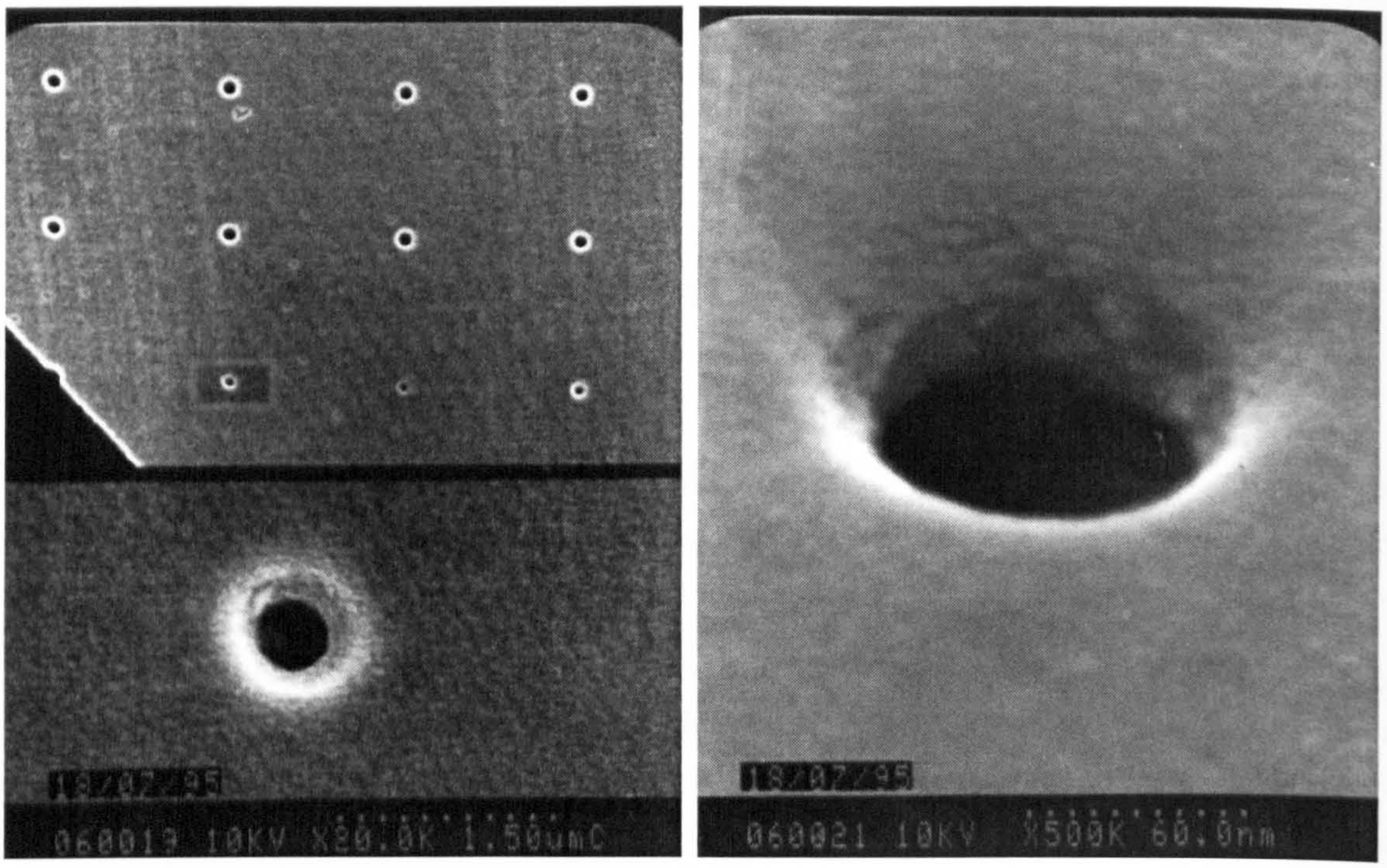


Figure 6.2.28: Micrographs of apertures fabricated on 100nm thick free-standing silicon nitride 'bridges' using the titanium etching scheme ('JEOL'). The titanium masking layer was reduced to 30nm thick and the resist was reduced to a single layer of 4% Elvacite. The sample was coated in a 10nm thick nichrome charge conduction layer. (a) Left Image: reproducibility of apertures around 50nm in diameter. Compare this with Figure 6.2.11. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: a 30° tilted view of a 100nm diameter aperture. It can be seen that the aperture profile is vertical for only part of the way through the silicon nitride layer.

Of these, the best explanation is (c), as the same effect has been seen before in a multi-layer resist scheme consisting of PMMA and aluminium for etching silicon when the aluminium has been too thin, or the etch has been too long [Lo, *et al.*[37] (1995)]. It is also likely that the titanium erodes quite rapidly in the silicon nitride etch once its oxide layer has been breached, and this leads to another possibility for extending the lifetime of the titanium mask: periodically during the silicon nitride etch, particularly where much deeper etching is required, the etch could be stopped and the chamber flooded with ordinary (non-ionised) oxygen gas to rebuild the titanium oxide layer. The silicon nitride etch could then be resumed. This procedure may have proven useful in the etching of the thicker silicon nitride associated with the contact mode AFM cantilevers being used.

The other, more obvious solution, would be to simply thin the silicon nitride around the apex of the pyramid, *e.g.* by partially dry etching this layer, prior to aperture definition. In any case, the thickness of the resist and titanium could be increased to that used in the

final dry etch test (two layers of 4% Elvacite, 50nm of titanium) which were so successful in achieving etch depths of up to 150nm with good verticality.

Finally, Figure 6.2.29 shows two micrographs comparing the smallest aperture on this last sample and that used to obtain statistics with the direct silicon nitride etching scheme. Both micrographs are on the same scale. It can be seen that the titanium etching scheme has resulted in an aperture over 20nm smaller in diameter.

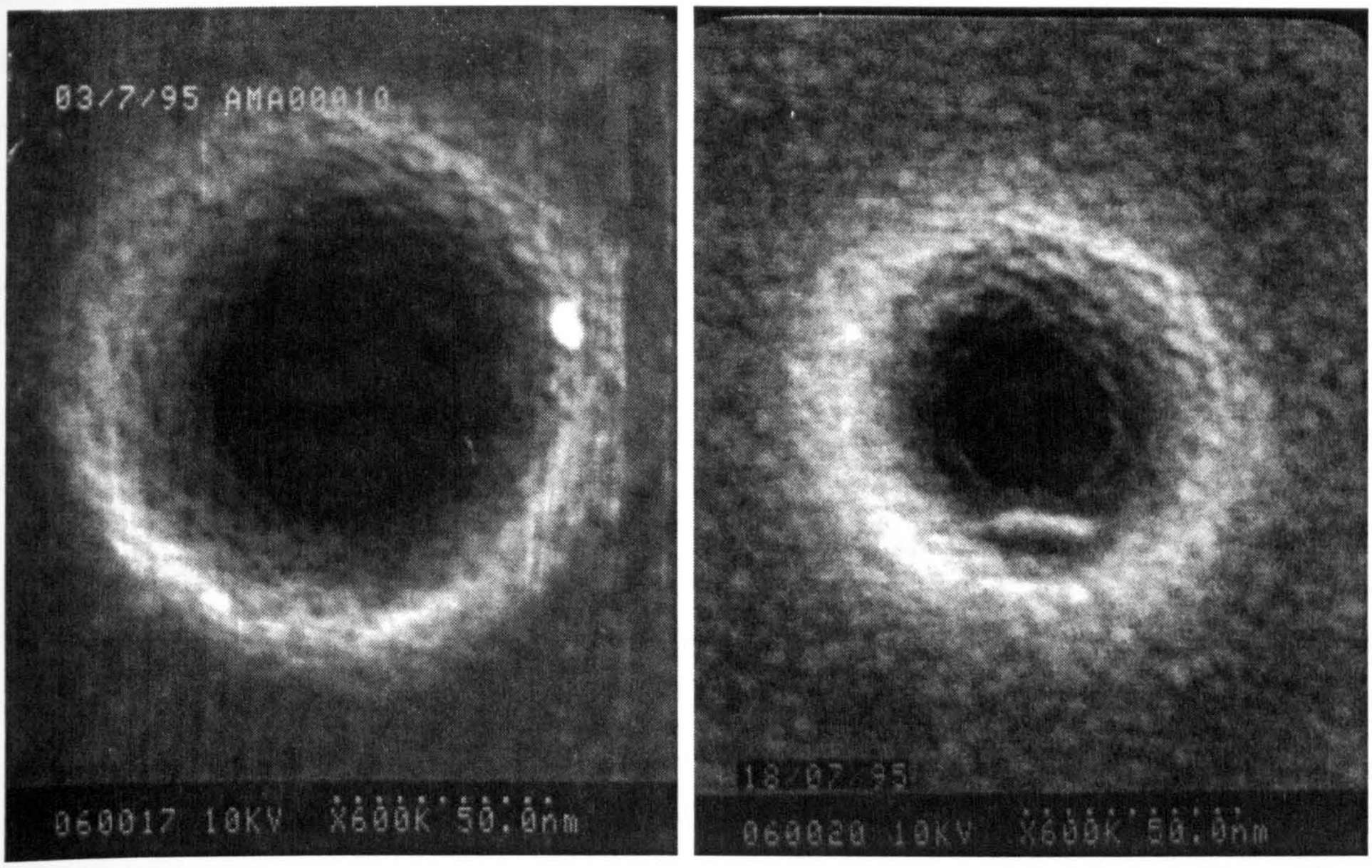


Figure 6.2.29: Micrographs comparing the smallest apertures produced directly from PMMA and via an intermediate titanium layer ('JEOL'). The titanium masking layer was reduced to 30nm thick and the resist was reduced to a single layer of 4% Elvacite. Both samples were coated with a 10nm thick nichrome charge conduction layer. (a) Left Image: a 60nm diameter aperture in 50nm thick LPCVD silicon nitride produced by dry etching directly from PMMA; (b) Right Image: a 35nm diameter aperture in 100nm thick LPCVD silicon nitride produced by dry etching from PMMA via a 30nm thick intermediate titanium layer. The titanium was stripped off in nanostrip prior to imaging.

Unfortunately, at this point disaster struck. The ECR machine, of which there was only one in the department, developed a computer problem. The hard disk had failed. Being an extremely old system, the hard disk was formatted using a proprietary format with proprietary software. Unfortunately, because the machine was no longer supported, fixing the machine would have required funds beyond the worth of the machine and so a quick solution was not envisaged. Eventually it was decided that a consultant be brought in to write completely new software, based on a standard PC running a popular operating system. However, the timescale for the completion of this work was well over six months and time was short.

Therefore, having found a reasonable solution to forming apertures for the SNOM probe, another solution now had to be found. This solution turned out to be even more radical than this one.

6.2.3 Aperture Definition in SiN from PMMA via Nichrome

Early on, the author had experimented with cross-linked PMMA. This work had been inspired by that of Dr. J. M. R. Weaver, who had demonstrated features down to about 5nm using cross-linked PMMA. Cross-linking, which can be thought of as a polymerisation process, can be induced chemically by the addition of a cross-linking agent to the resist and the use of an initiator [*e.g.* Schneider, *et al.*[41] (1995); Schliebe, Schneider and Aschoff[47] (1996)], or, by exposure to, for example, an electron beam [*e.g.* Broers, Harper and Molzen[48] (1978); Lo, *et al.*[37] (1995); Tada and Kanayama[49] (1995)] when an extremely high dose of electrons, typically ranging from 100 to 1000 times greater than the normal exposing dose, is required. [This means that the technique is quite useless for any large area lithography (as exposure times scale with exposure dose). On the beamwriter, at 100kV and with a 12nm spot, the exposure rate is about 100 seconds per micron square of area per million $\mu C/cm^2$ dosage.]

When the cross-linking is induced chemically, PMMA still behaves as a positive resist, but offers higher contrast (lower sensitivity) and greater dry etch resistance than ordinary PMMA [Schneider, *et al.*[41] (1995)]. When the cross-linking is induced by electron-beam exposure, PMMA behaves more like a high resolution negative resist with enhanced dry etch resistance, the high resolution once again as a result of higher contrast (lower sensitivity) than ordinary PMMA [Tada, *et al.*[49] (1995)].

Due to the high doses involved in the production of electron-beam induced cross-linked PMMA, the question always arises as to whether the resist that is seen is simply the contamination 'resist' commonly observed when an electron beam is focused on a particular spot for a lengthy time [see §2.3.3]. This contamination is as a result of the cracking of hydrocarbons present in the vacuum chamber of the machine (the origin of which is most likely to be the pump oil used in diffusion and mechanical rotary pumps), which is naturally focused at the most intense energy concentration of the beam of electrons, *i.e.* at the spot where the electron beam is focused. In this case the cross-linked polymer is believed to be a carbon rich compound [Broers[29] (1988)].

However, some tests carried out by the author had suggested that this explanation was not correct: two samples were exposed identically, but only one sample had a PMMA layer on it. If the origin of the cross-linked polymer was simply contamination in the chamber, then features should have appeared on both samples. This did not rule out the possibility that during the PMMA exposure, some short-chain hydrocarbons of a gaseous nature were evolved and that under the intense electron beam these were then cracked and subsequently deposited at the focal point of the electron beam. Subsequent tests revealed that the cross-linked polymer was always of virtually the same thickness as the original PMMA layer, which also suggested that this was not contamination. This was the same conclusion reached by another group [Tada, *et al.*[49] (1995)].

Then if it is actually the case that PMMA can be cross-linked by a high dose of electrons, the question as to the mechanism by which it is formed arises. This may be due to a combination of resist heating [Yasuda, *et al.*[50] (1994)] and radiation damage [Choi, *et al.*[51] (1988)]. However an analysis carried out by one group has shown that high exposure results in the removal of all the side chains from the PMMA molecule, leaving it with a carbon-like composition which accounts for its enhanced dry etch resistance [Tada, *et al.*[49] (1995)].

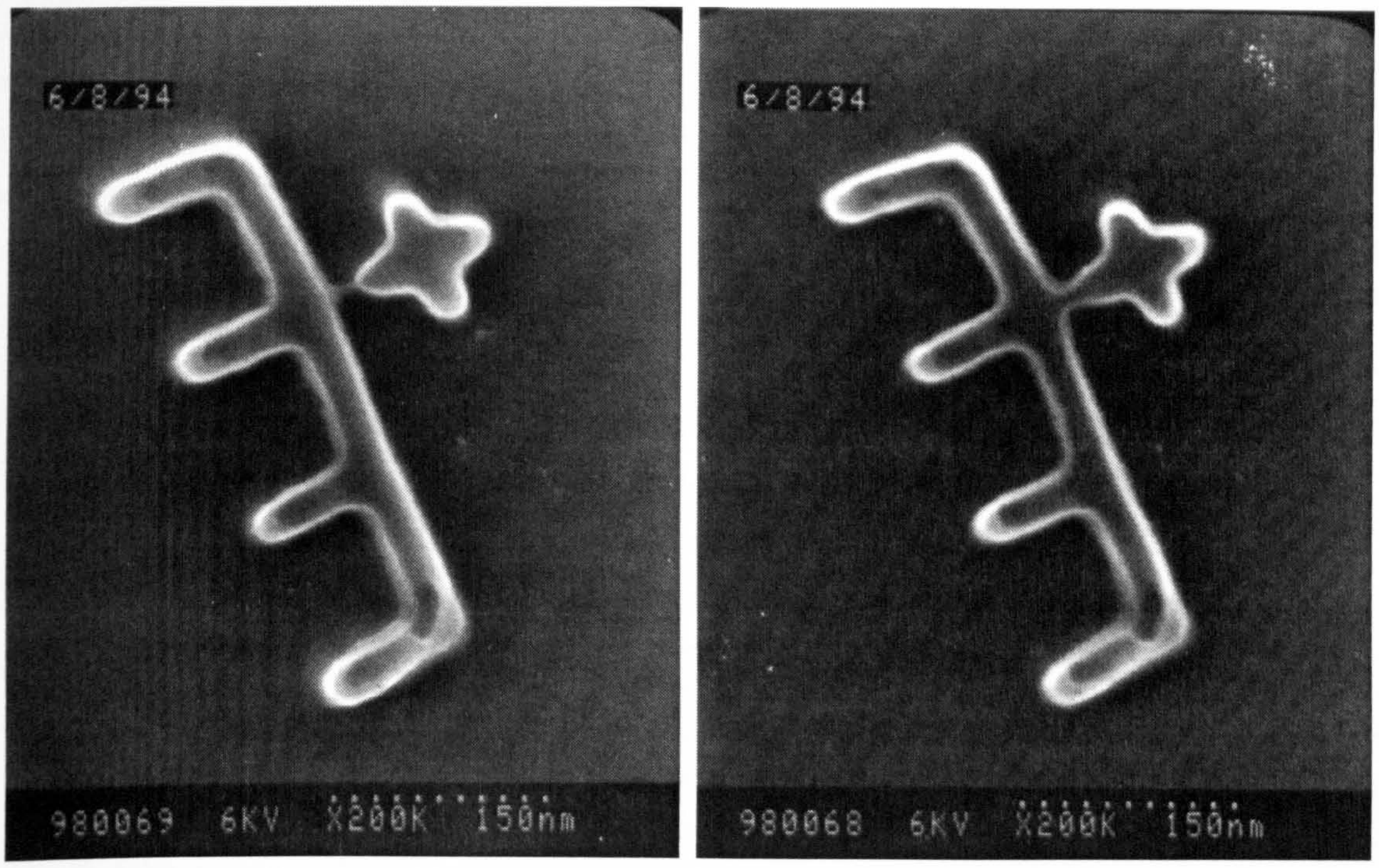


Figure 6.2.30: Micrographs illustrating the properties of cross-linked PMMA as a negative resist ('amh0003'). The sample was not coated with a charge conduction layer. (a) Left Image: the threshold separation less than which a separation between the cross and 'line' cannot be maintained; (b) Right Image: the effect of capillary forces between the cross and 'line' despite the presence of reinforcing buttresses. Notice how the 'line' is no longer straight, but bulges out towards the cross.

All this then means that the big difference between PMMA and an ordinary negative resist is that it can behave as both a high resolution positive resist and negative resist simultaneously simply by exposing different areas with different doses. As would be expected, there are some complications to using this feature of PMMA, *e.g.* if an area is cross-linked, then the high dose required means that the resulting backscatter electrons expose the PMMA positively for some area around the cross-linked PMMA feature, so that each cross-linked feature is surrounded by a positively exposed border. Unlike an ordinary negative resist, it is not only the areas which are exposed negative that remain after development, but, like a positive resist, also the regions which have been underexposed. However, since cross-linked PMMA is not soluble in acetone [Tada, *et al.*[49] (1995)], if acetone is used as a developer all the unexposed and moderately exposed PMMA is removed and the cross-linked PMMA can be made to behave in an identical manner to a conventional negative resist.

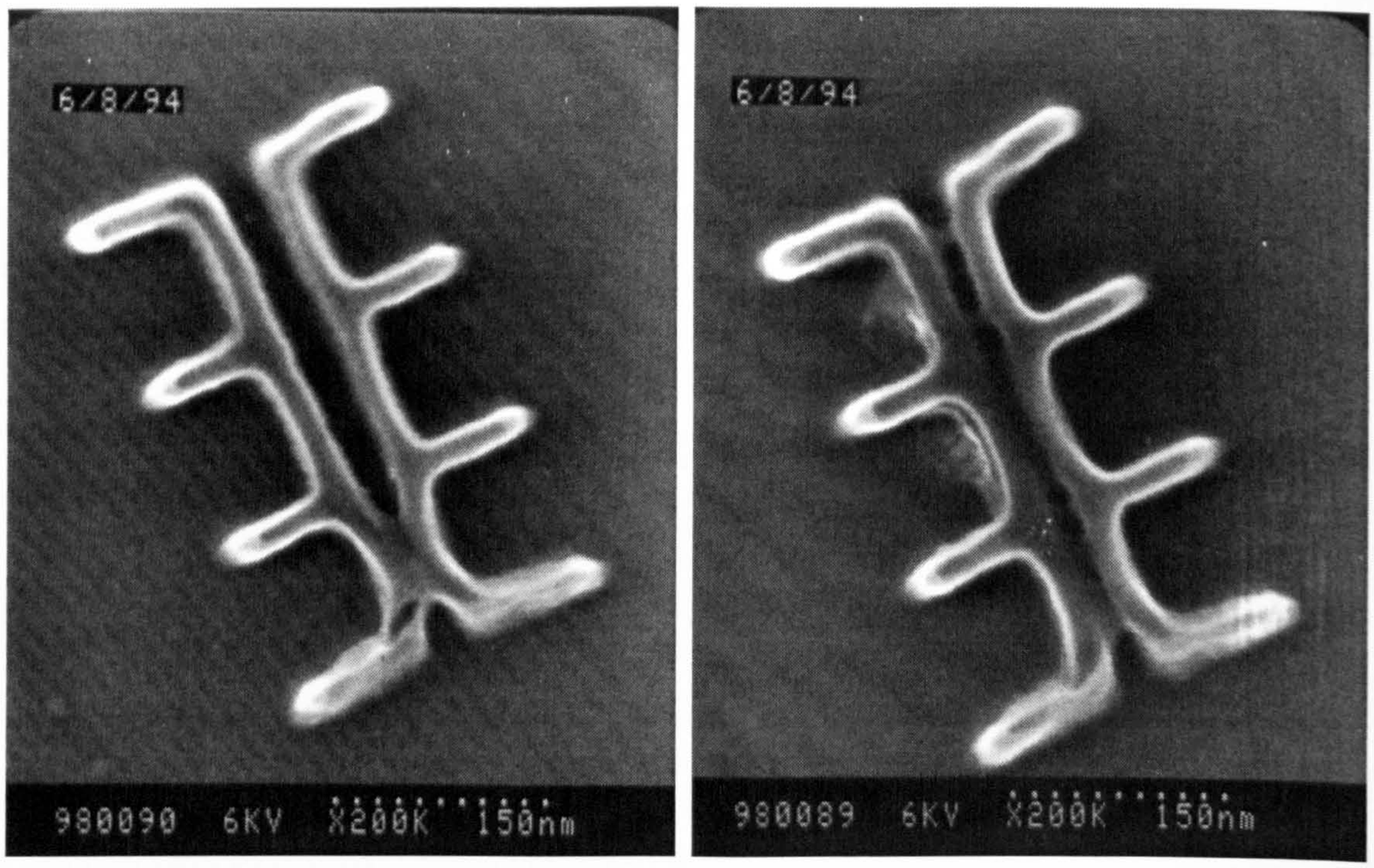


Figure 6.2.31: Micrographs illustrating the properties of cross-linked PMMA as a negative resist ('amh0003'). The sample was not coated with a charge conduction layer. (a) Left Image: the effect of capillary forces between the two 'lines' leads to resist collapse at the least reinforced parts of the structure; (b) Right Image: smaller separations than in (a) lead to a more comprehensive collapse of the resist features.

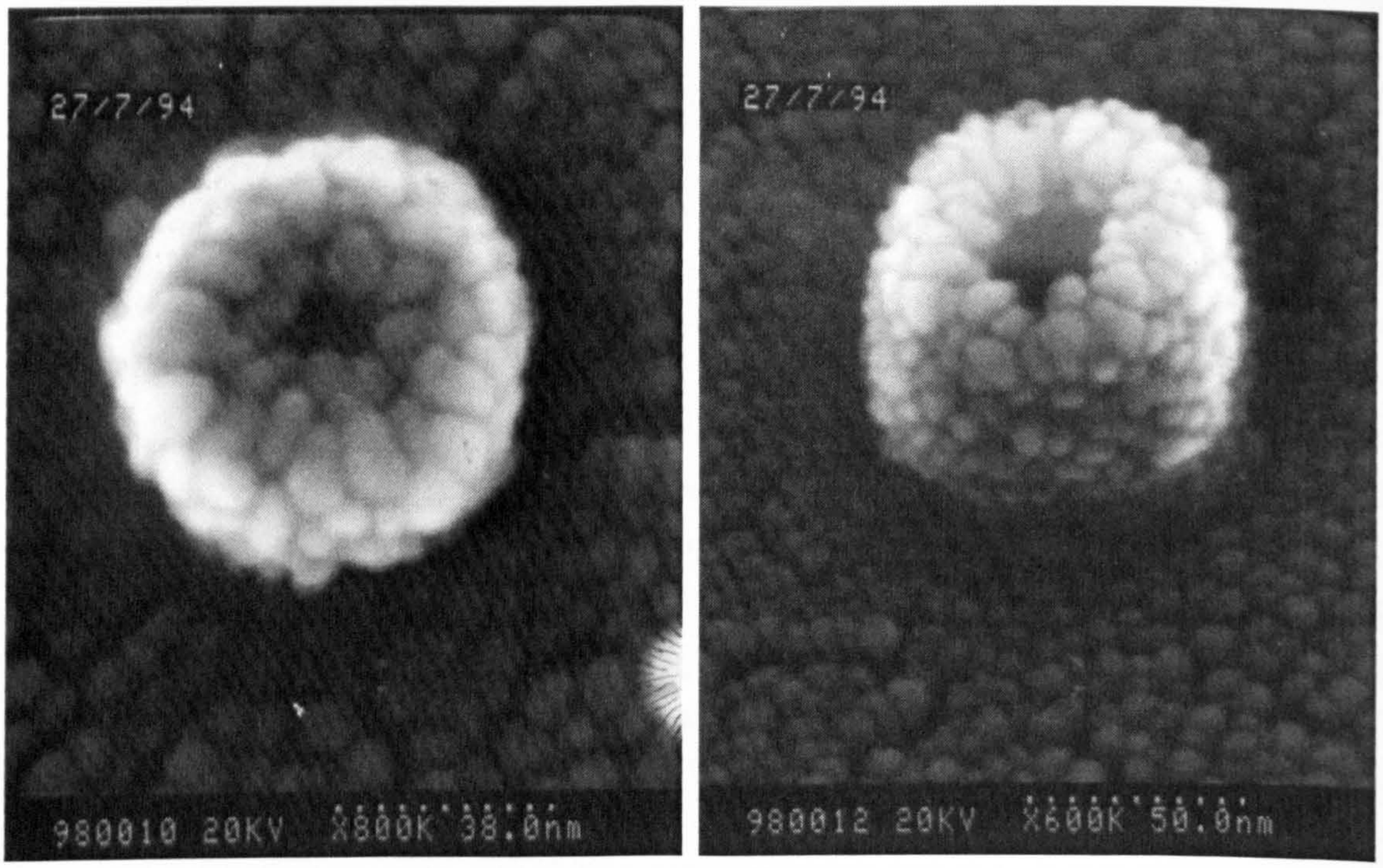


Figure 6.2.32: Micrographs illustrating the properties of cross-linked PMMA as a negative resist ('amh0002'). The sample was coated with 20nm of gold/palladium as a charge conduction layer. An annular ring with what appears to be a sub-10nm diameter aperture. (a) Left Image: flat view; (b) Right Image: 30° tilted view.

The exact nature of the formation process and composition of cross-linked PMMA are of less significance than the ability to use it for pattern transfer. The principal hurdle to using this polymer in pattern transfer was found to be that of its removal prior to

subsequent processing. Negative resists are generally very resistant to most forms of chemical attack, however, cross-linked PMMA appears to be more resistant still.

Figures 6.2.30 to 6.2.34 illustrate some of the salient points in the use of cross-linked PMMA. Figures 6.2.30 and 6.2.31 are micrographs of uncoated resist which illustrate the fact that whether features appear or not is not so much limited by proximity as by the capillary forces generated during development when the substrate is dried [McCord, *et al.*[8] (1993)]. [Since the PMMA has already become cross-linked due to over exposure by an electron beam, the attentions of the SEM does not present the usual shrinkage problems associated with viewing uncoated resist.] In Figure 6.2.30 a cross has been placed close to a line. The line is supported by buttresses to help reinforce it against capillary forces. Even so, when the cross is close enough to the line, although the cross has been placed opposite a directly reinforced part of the line, capillary forces are strong enough to pull the cross and line together. Notice how the line is pulled towards the cross in Figure 6.2.30b. Figure 6.2.30a shows the threshold separation between cross and line at which the capillary forces are not quite strong enough to pull the two together.

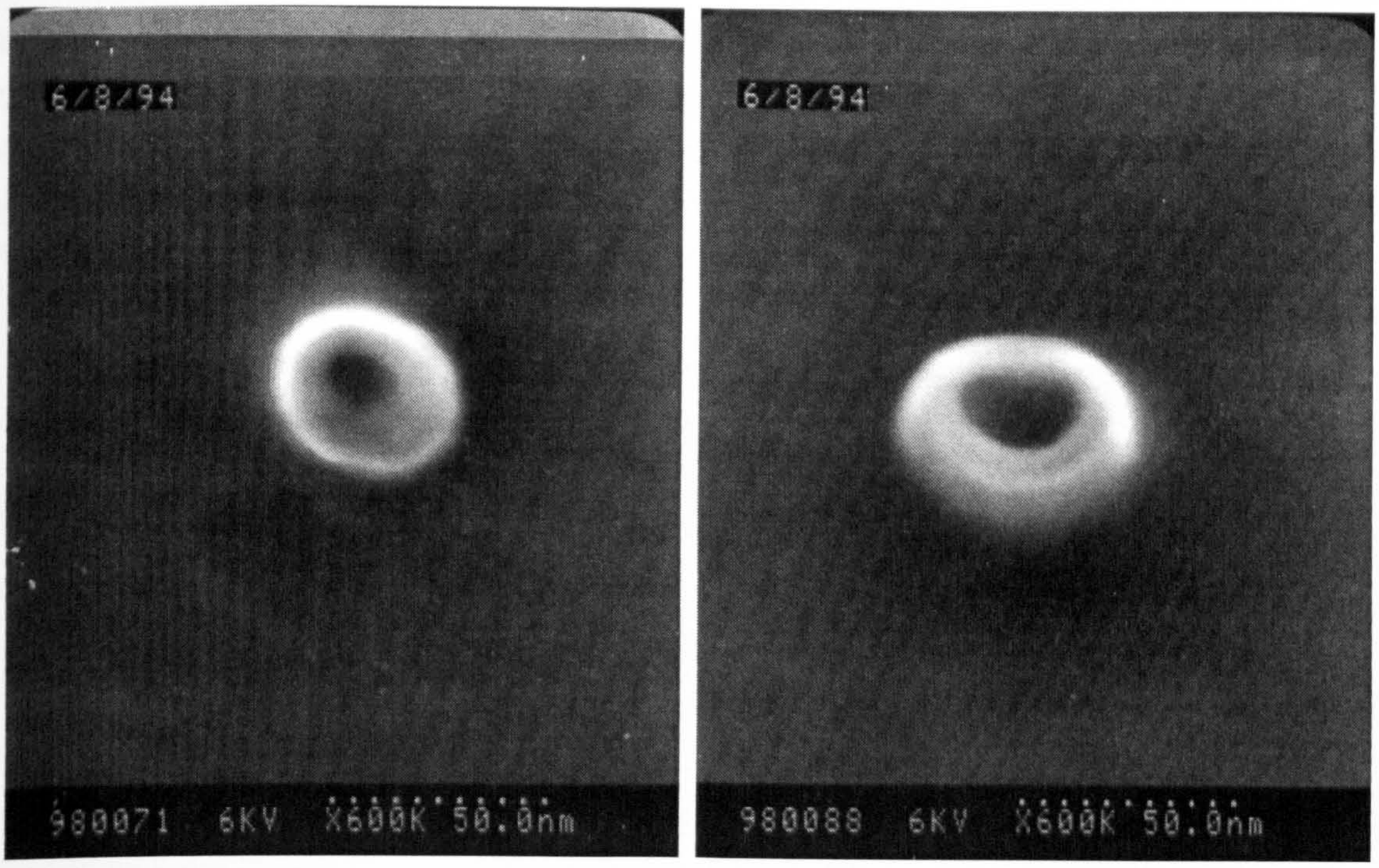


Figure 6.2.33: Micrographs illustrating the properties of cross-linked PMMA as a negative resist ('amh0003'). The sample was not coated with a charge conduction layer. Annular rings with small central apertures. (a) Left Image: flat view of an annular ring with a 15nm diameter aperture; (b) Right Image: 30° tilted view of an annular ring with a 25nm diameter aperture.

In Figure 6.2.31 two buttressed lines are placed opposite each other, the separation between them being varied. As can be seen from Figure 6.2.31a when the capillary forces do become strong enough, the two lines are pulled together at the weakest point, midway between the buttresses. With shorter and shorter separations a more comprehensive collapse of the support mechanism occurs and both lines become more and more solidly welded together, as demonstrated in Figure 6.2.31b. This again reinforces the fact that

capillary forces play a major role in the generation of small resist features, since the low sensitivity of the cross-linked PMMA allows proximity effects to be virtually ruled out.

One of the most rigid self supporting structures that can be made is an annular ring. Figure 6.2.32 shows such a cross-linked PMMA structure sputtered with about 20nm of gold/palladium. The exposure pattern consists of a square in which the central region has been left unexposed. Figure 6.2.32a suggests that the central aperture is less than 10nm in diameter. Figure 6.2.32b shows a tilted sample view of a similar annular ring. However, the sputtering process may have made the central hole smaller than it was before being coated. Even so, this structure is impossible to achieve with positively exposed PMMA due to the intraproximity effect, *i.e.* it would normally be impossible to expose a central hole and surrounding region without also exposing the ring itself through backscattering. Figure 6.2.33 shows similar features in uncoated resist. Figure 6.2.33a shows an annular ring with a central aperture around 15nm in diameter. Figure 6.2.33b shows a tilted sample view of a slightly larger annular ring.

Figure 6.2.34 shows another feature difficult to achieve with positively exposed PMMA. These are arrays of dots of resist. At sufficiently small dot separations, capillary forces start causing the dots to collapse towards one another (not shown). With positively exposed PMMA it may just be possible to achieve the reverse of the pattern in Figure 6.2.34a, although that of Figure 6.2.34b, which are 30nm dots placed at a centre-to-centre separation of less than 60nm, would be difficult to achieve due to the interproximity effect.

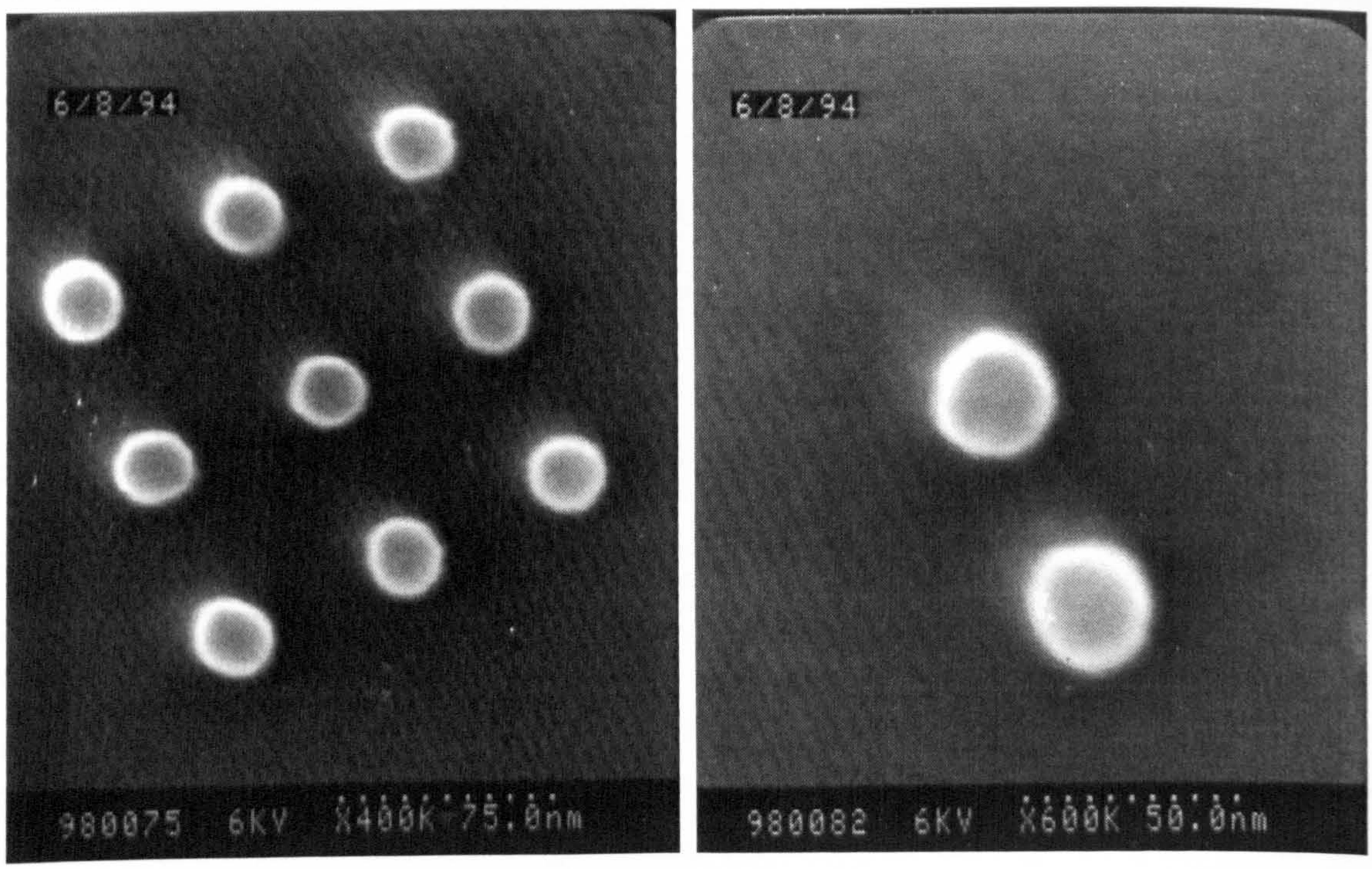


Figure 6.2.34: Micrographs illustrating the properties of cross-linked PMMA as a negative resist ('amh0003'). The sample was not coated with a charge conduction layer. (a) Left Image: an array of dots at a centre-to-centre spacing of about 65nm; (b) Right Image: two dots at a centre-to-centre spacing of about 50nm.

All the above features were made up of double pixel lines based on 5nm pixels exposed at 100kV with a 12nm spot. The samples used a bi-layer of PMMA consisting of a 2.5% BDH spun at 5000 rpm for 60 seconds and 2.5% Elvacite also spun at 5000 rpm for 60 seconds with a one hour, 180°C bake between layers and an overnight 180°C bake after the final layer [see Table 3.5.1]. The dose used for the line and cross combination and annular rings was about 415,000μC/cm^2, for the double buttressed lines was about 240,000μC/cm^2 and for the dot arrays was about 830,000μC/cm^2. Obviously had the designs been reduced to single pixel patterns, the doses required would increase, probably by a factor of two or more. Such large doses would normally cause the beamwriter to generate an out of frequency error, since the lowest rate at which the beam blanking can operate is 0.5kHz. This can be overcome by setting the CATS software to include overlaps and by designing the patterns with several copies superimposed on each other. During exposure, the beam will be blanked by the number of overlaps but not deflected from its current position, thereby having the same effect as if the dose had been multiplied up by the number of overlaps. At 100kV and with a 12nm spot and 5nm resolution, typically the maximum practical dose that can be used is about 400,000μC/cm^2 per overlap. This corresponds to a frequency of just below 1kHz, and taking into account that the beam current at 100kV may change by a factor of two between when the filament is new and towards the end of its lifetime, frequencies any lower than this would risk rejection by the beamwriter.

Thus, the problem was to find a way of using cross-linked PMMA to fabricate apertures for use in SNOM. Since the apertures would be small and isolated, it would be advantageous if the large doses required could be compensated for by a small exposure area. This would be the case if the aperture were to be exposed directly, assuming this was possible.

The first thought along these lines was that the annular rings could be used as a mask for directly dry etching the silicon nitride (or even titanium using the earlier RIE80 etching process – see §6.2.2), given its superior dry etch resistance when compared to ordinary PMMA. However the problem would be that the region outside the annular ring would also be etched. One possibility was to cover the whole sample with nichrome [from the Plassys MEB 450 – see §3.6.2], then to open up 1μm wide windows in this layer at the apex of the pyramidal tips using chrome etch (and one level of electron-beam lithography), before forming the cross-linked PMMA annulus over the apex of the tip (using a second level of electron-beam lithography), by extending the exposure region out one micron on each side, so as to leave a small aperture free of resist at the centre. This could then be used as a mask to dry etch the silicon nitride beneath, with the nichrome layer protecting the rest of the substrate. However, exposing a solid 2μm square region (allowing for some overlay error) with a dose of 250,000μC/cm^2 would require around 100 seconds per aperture, which translates to over an hour for 50 devices. Furthermore this process would require two reasonably accurate alignments to each device.

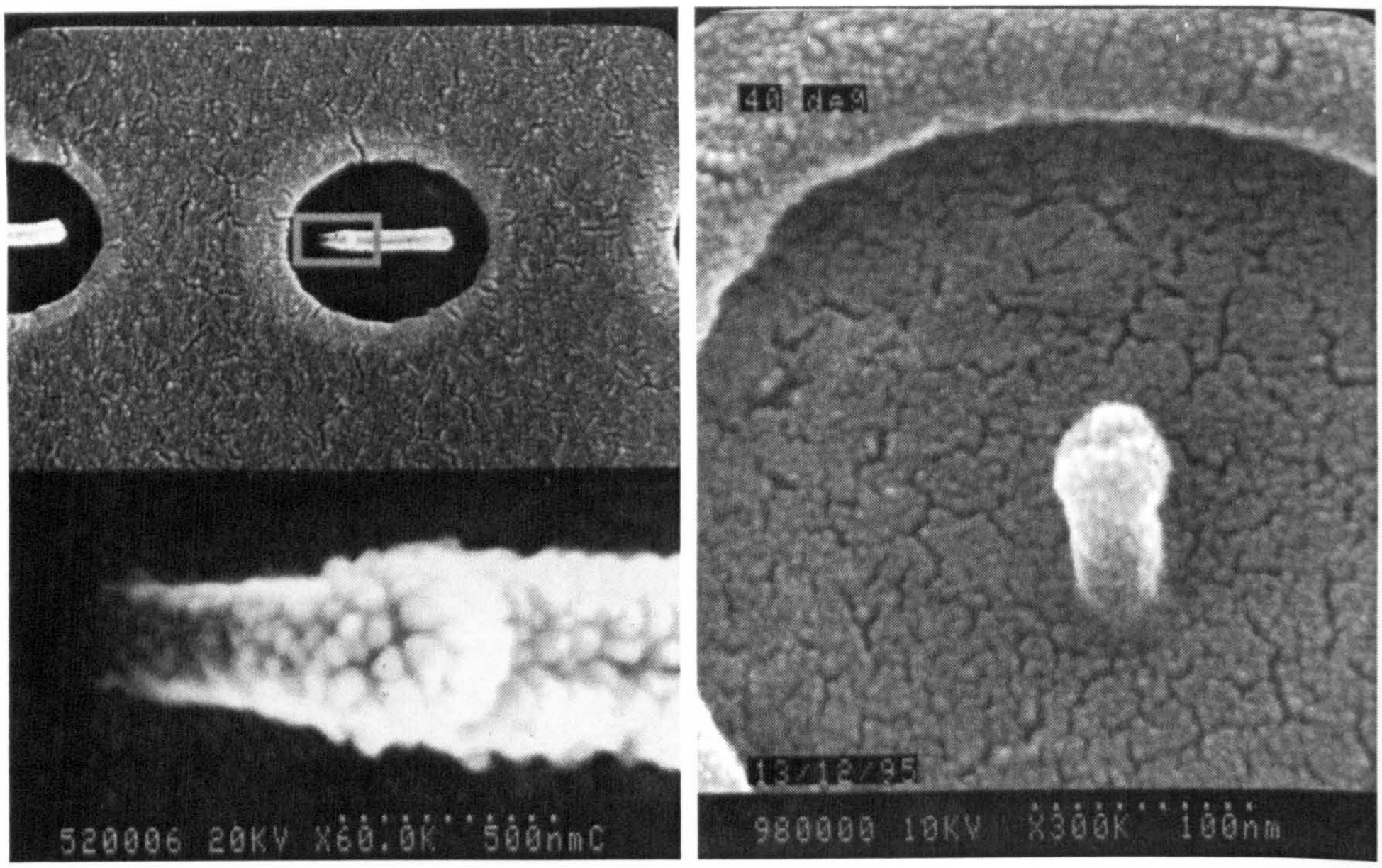


Figure 6.2.35: Micrographs illustrating an undercut profile in cross-linked PMMA. The samples were coated with 20nm of gold/palladium as a charge conduction layer. (a) Left Image: horizontally tilted view of an exposed short line. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: 30° tilted view of a single dot.

Within the department it was known that nichrome was one of the best dry etch masks available [Murad[16] (1993-1996)], being extremely immune to even chemically reactive fluorine based etches. If nichrome could be used as a masking layer for the etching of silicon nitride then there would be no problems with mask erosion or thickness of silicon nitride. The difficult part would be defining a small aperture in a nichrome film. The notion that it might be possible to use the single cross-linked PMMA dots to define such an aperture was explored. This would essentially constitute a lift-off process using a negative resist. As noted in §2.7.1 [also see Hatzakis[11] (1988)], this would only ever be possible if the metal deposited on the resist-free part of the substrate was physically detached from that deposited on top of the resist. This meant that the resist would have to have an undercut profile. At this point it was speculated that if the size of the cross-linked PMMA structure in some way reflected the lengths of the original PMMA chains, then it may be possible to obtain this profile by simply using a bi-layer of resist of different chain lengths. Some old evidence and another quick test, as shown in Figure 6.2.35, seemed to confirm this.

The problem then was how to remove the PMMA so as to achieve lift-off. One idea was to simply oxygen ash the sample (*i.e.* use a barrel etcher [see §3.6.4] to subject the resist to an oxygen plasma) so as to 'burn' the resist away. Typically an hour of ashing is used to remove stubborn resist [Cameron[52] (1993-1996)], so an attempt was made to see if cross-linked PMMA was similarly removed. After an hour of ashing it appeared that the resist had withered to become extremely thin and many of the cross-linked dots had fallen over, however, none had actually completely disappeared. [Had this ashing procedure failed

outright, *i.e.* had the cross-linked PMMA not been affected at all by the ashing process, the result would have been significant from a lithographic point of view as it would have opened up the possibility of using ashing to carry out 'dry development', that is, the non-crosslinked PMMA could have been removed using a development process free of capillary forces and so point the way to higher resolution lithography. Unfortunately the result was midway between outright success and outright failure.]

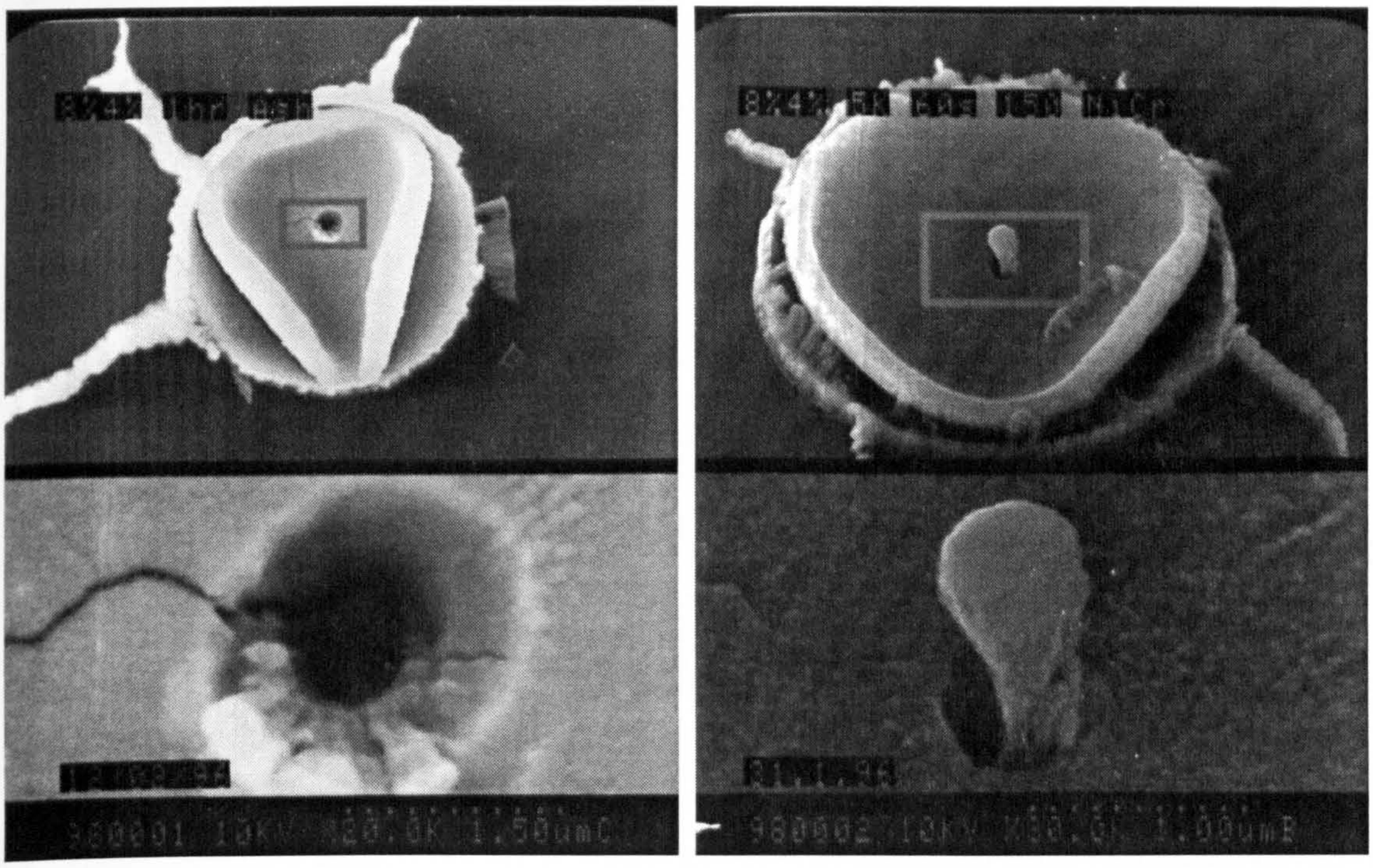


Figure 6.2.36: Micrographs illustrating the attempt to lift-off apertures in 150nm thick nichrome directly from cross-linked PMMA. The samples were not coated with a charge conduction layer. Stress in the metal layer has caused it to curl up at the edges. (a) Left Image: a successful lift-off, displaying an overcut aperture profile. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: a failed lift-off caused by the lack of rigidity of the PMMA pillar resulting in it keeling over during either development or evaporation. The bottom half is a 5x magnification over the displayed scale.

In order to see whether this withering was sufficient to achieve lift-off another sample upon which nichrome had been evaporated after the sample was developed was put through the same process. In this case the PMMA appeared not to have been affected at all. This was attributed to some peculiar charging effect, possibly caused by the fact that the metal on top of the cross-linked PMMA dots was electrically isolated from the rest of the substrate: then it was possible that the plasma had ionised these metal caps to a level at which they repelled the attentions of additional ions, thereby keeping the active chemical ions away from the resist.

At this point another idea emerged based on experience obtained during the development of the photolithography levels. There it had been noted that nichrome was not attacked by the hot nanostrip clean [see Appendix D] and that nanostrip was capable of removing even extremely hardbaked photoresist (as used when the hot orthophosphoric acid based wet etching of silicon nitride was being carried out – see Appendix E (§E.1)). The

possibility that hot nanostrip could remove the cross-linked PMMA had to be explored. After a few tests it was found that two ten minute dips in hot nanostrip:water 1:1 did dissolve away the cross-linked PMMA whilst leaving the nichrome unscathed.

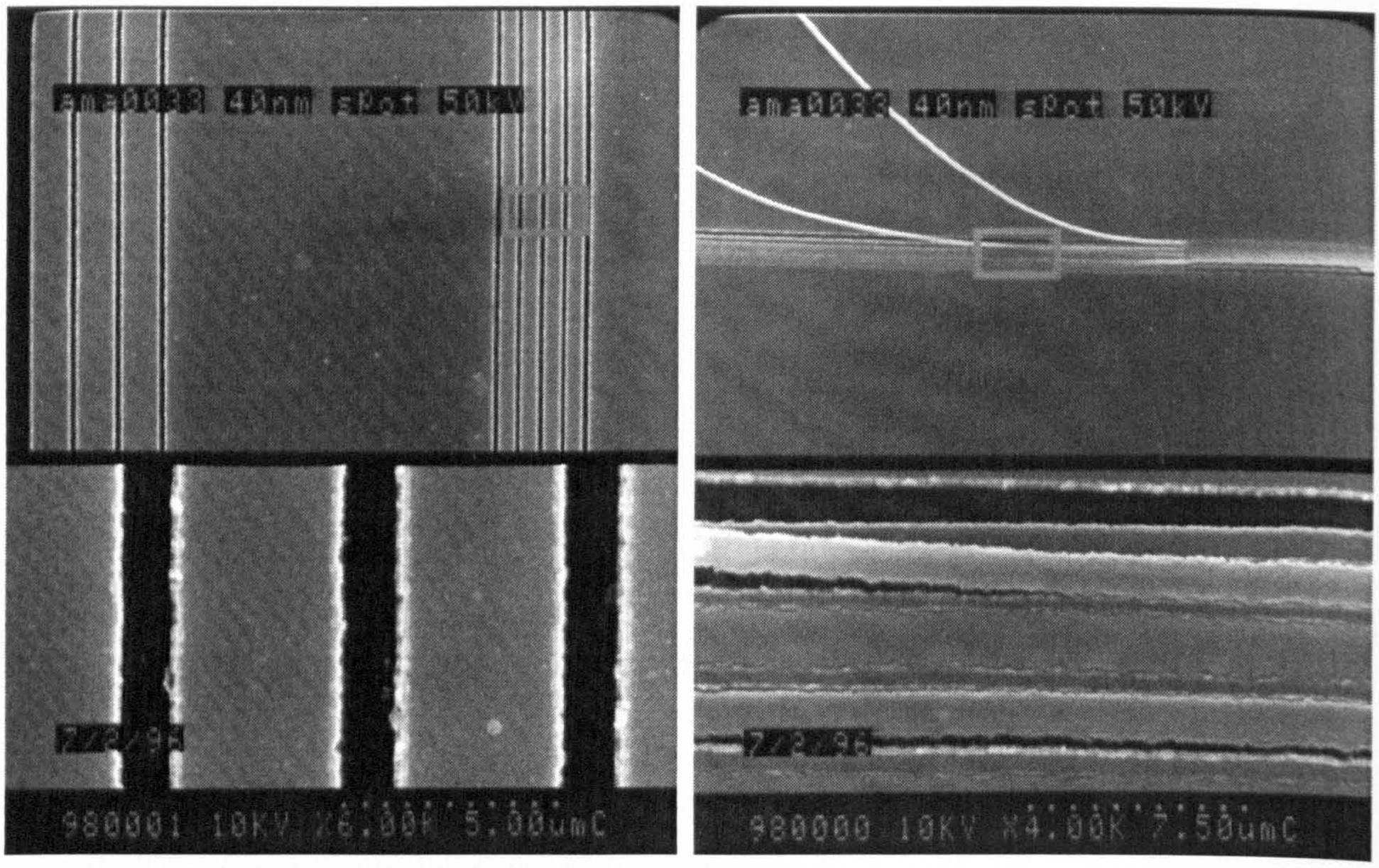


Figure 6.2.37: Micrographs illustrating an attempt to lift-off 50nm thick nichrome wires using cross-linked PMMA ('ama0033'). The samples were not coated with a charge conduction layer. (a) Left Image: a successful lift-off; (b) Right Image: a failed lift-off caused by a stitching error on the part of the beamwriter. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

This success led to the suggestion that it might be possible to define the aperture directly in the SNOM metal shield, provided that a metal shield was chosen which could survive the resist stripping and release etch. To test this possibility a sample was prepared with a bi-layer of resist consisting of 8% BDH spun at 5000 rpm for 60 seconds and 4% Elvacite also spun at 5000 rpm for 60 seconds with a one hour 180°C bake between layers and an overnight 180°C bake after the final layer. The resist was developed in acetone after exposure and a layer of nichrome 150nm thick was evaporated onto the sample. The specimen was then examined after the nanostrip lift-off process had been carried out.

Figure 6.2.36 shows two apertures formed in this way. [Note that the residual stress in the nichrome has caused the metal ring to curl up off the substrate.] Most had failed, but two other observations were made which suggested that direct definition was not a good idea. The first was that the aperture profile was overcut, as shown in Figure 6.2.36a, due to the overhanging profile of the original resist pillar. This would not matter if the nichrome was used as a dry etch mask as its low erosion rate means that the profile would not be transferred to the silicon nitride layer. The second observation was that the rigidity of the pillar depended on the dose used, and that the thicker the resist the higher the dose required. If the pillar was not rigid enough, then it would either keel over during development or it would lean over as metal was deposited upon it, as shown in

Figure 6.2.36b. Clearly with a thick metal layer, this problem would be made worse. Overall it was felt that it was better to overexpose than underexpose, since features did not seem to grow appreciably with increasing dose, but underexposure left the features unable to support themselves. This lack of rigidity due to underexposure has been noted in other, conventional negative resists, and has been attributed to how impervious the resist becomes to the developer [Chiong, Wind and Seeger[53] (1990)].

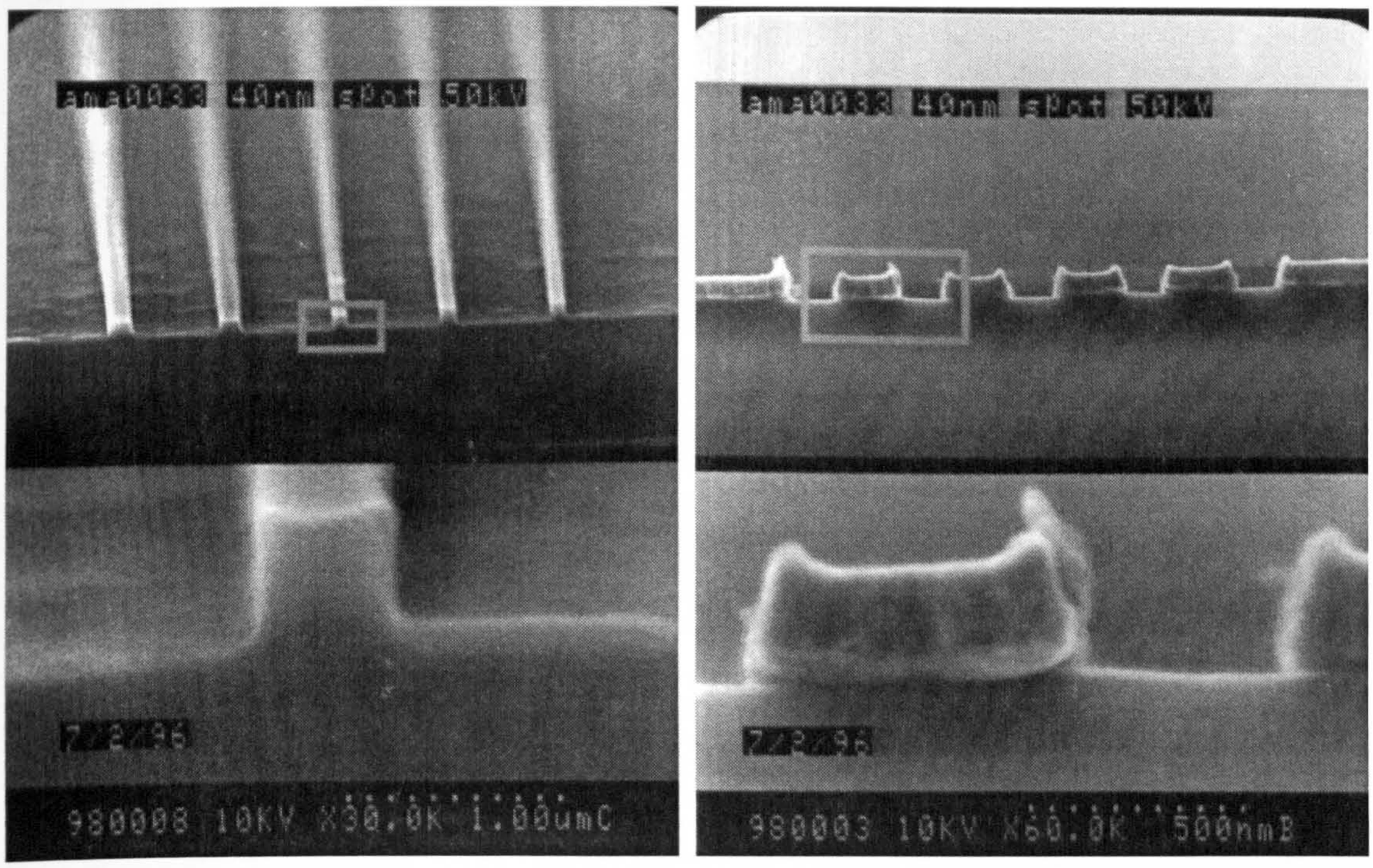


Figure 6.2.38: Micrographs illustrating an attempt to lift-off 50nm thick nichrome wires using cross-linked PMMA ('ama0033'). The samples were not coated with a charge conduction layer. (a) Left Image: a cleaved-through view of a cross-linked PMMA line. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: a cleaved-through view of lifted off wires. The lack of any residue is encouraging. The bottom half is a 5x magnification over the displayed scale.

Another test was performed to determine whether the nanostrip process left any residue behind. Negative resists are notorious for leaving residues behind [Chiong, *et al.*[53] (1990)] and this can be a problem when dry etching is used as the residue acts as a mask. Some 500μm long, single pixel lines were exposed at 50kV with a 40nm spot on a sample with a bi-layer of resist consisting of 4% BDH and 2.5% Elvacite both spun at 5000 rpm for 60 seconds with a one hour 180°C bake between them and an overnight 180°C bake following the final layer. The substrate consisted of silicon upon which 500nm of PECVD silicon nitride had been deposited. The sample was developed in acetone for 35 seconds. As the pattern had been repeated several times on the sample, the sample was split into several parts and 50nm of nichrome was evaporated onto one part using the Plassys MEB 450 evaporator. Nanostrip lift-off was performed on this sample, before it was examined briefly in the Hitachi S900 SEM. On the whole these wires appeared to have lifted off successfully, as illustrated by Figure 6.2.37a, although in one area where there appeared to be a stitching error on the part of the beamwriter, the wires did not lift-off properly, as illustrated in Figure 6.2.37b. Two samples were then cleaved so as to cut

through the lines and viewed edge-on in the Hitachi S900 SEM. Figure 6.2.38 shows the result. There was no obvious signs of any large scale residue which was encouraging.

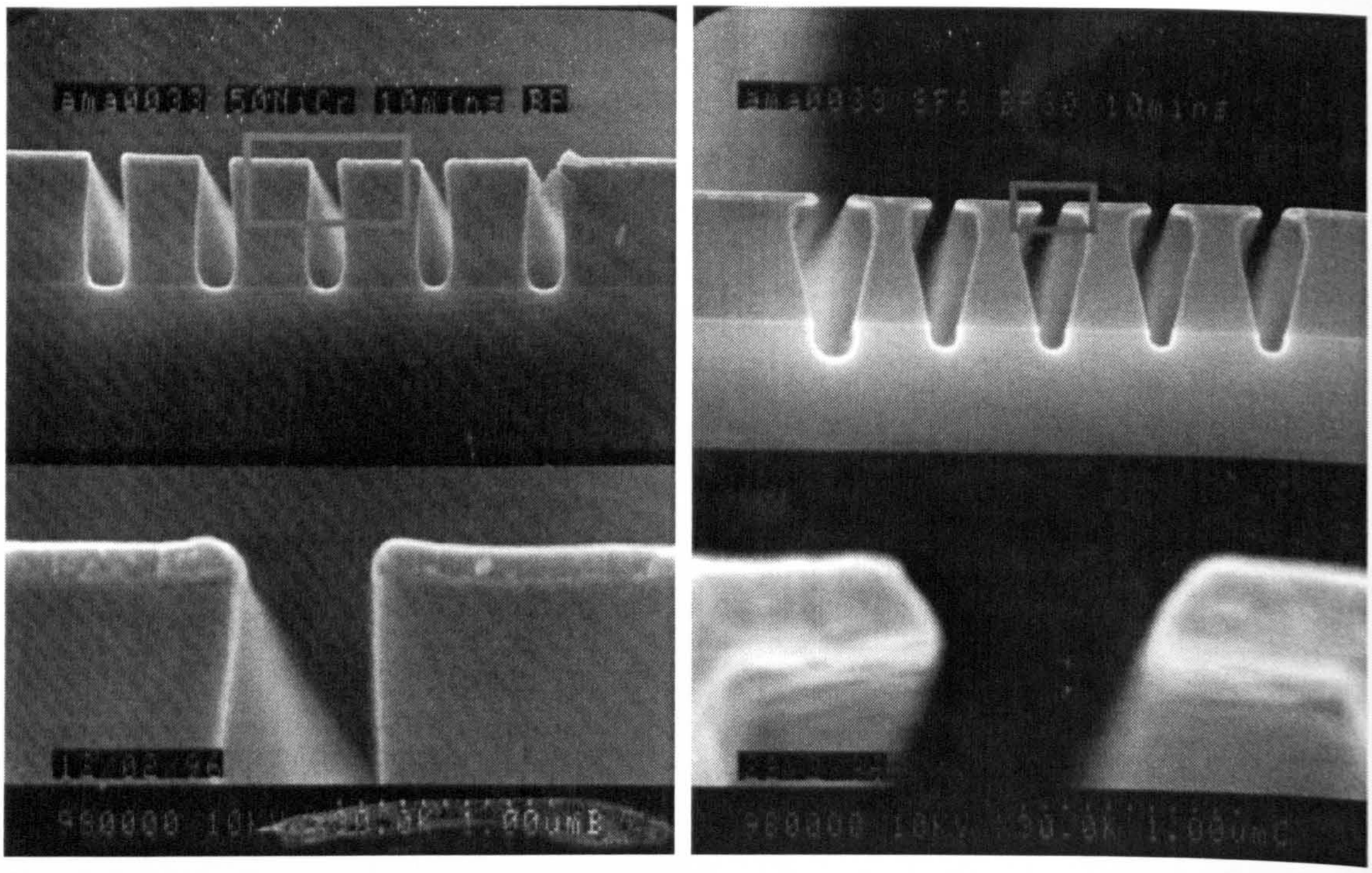


Figure 6.2.39: Micrographs illustrating the dry etching of 500nm of PECVD silicon nitride from cross-linked PMMA via nichrome ('ama0033'). The samples were not coated with a charge conduction layer. The resistance of the 50nm thick nichrome masking layer is clearly demonstrated. (a) Left Image: etched using C_2F_6. The etching profile is undercut but the silicon beneath has not started etching. The bottom half is a 5x magnification over the displayed scale of 1μm; (b) Right Image: etched using SF_6. The etching profile is overcut but the mask has been undercut and the silicon beneath has been partly etched. The bottom half is a 10x magnification over the displayed scale.

This did not, however, rule out some residue on the level of a few monolayers. To make sure that there were no other hidden problems with using this process to fabricate apertures for SNOM, two such lifted-off samples (uncleaved) were dry etched using the standard C_2F_6 and SF_6 based silicon nitride etches for ten minutes. These samples were then cleaved and viewed edge-on in the Hitachi S900 SEM. The results are shown in Figure 6.2.39.

It is clear that both dry etches can be regarded as successful., with the nichrome masking being hardly affected at all. However, while the SF_6 based etch undercuts the mask as desired, the profile produced is overcut. It also appears to etch the silicon much more readily than the C_2F_6 etch. However, the C_2F_6 etch produces the right kind of undercut profile, but does not etch the silicon as readily. The most probable reason for the delay in the silicon etching in this case is that as the etch occurs deeper and deeper within the surface, access to the etch front by the reactive species becomes more limited and transport issues associated with the diffusion of waste products from the etch front to the surface and active species to the bottom of the trench start to have a greater effect [Murad[16] (1993-1996)]. The net result is that the etch slows down. Thus this meant that the C_2F_6 etch was the best choice, but that the etch time would have to be increased to

allow sufficient time for the silicon to be etched. As overetching is not a problem in this process it was acceptable to double or triple the etch time.

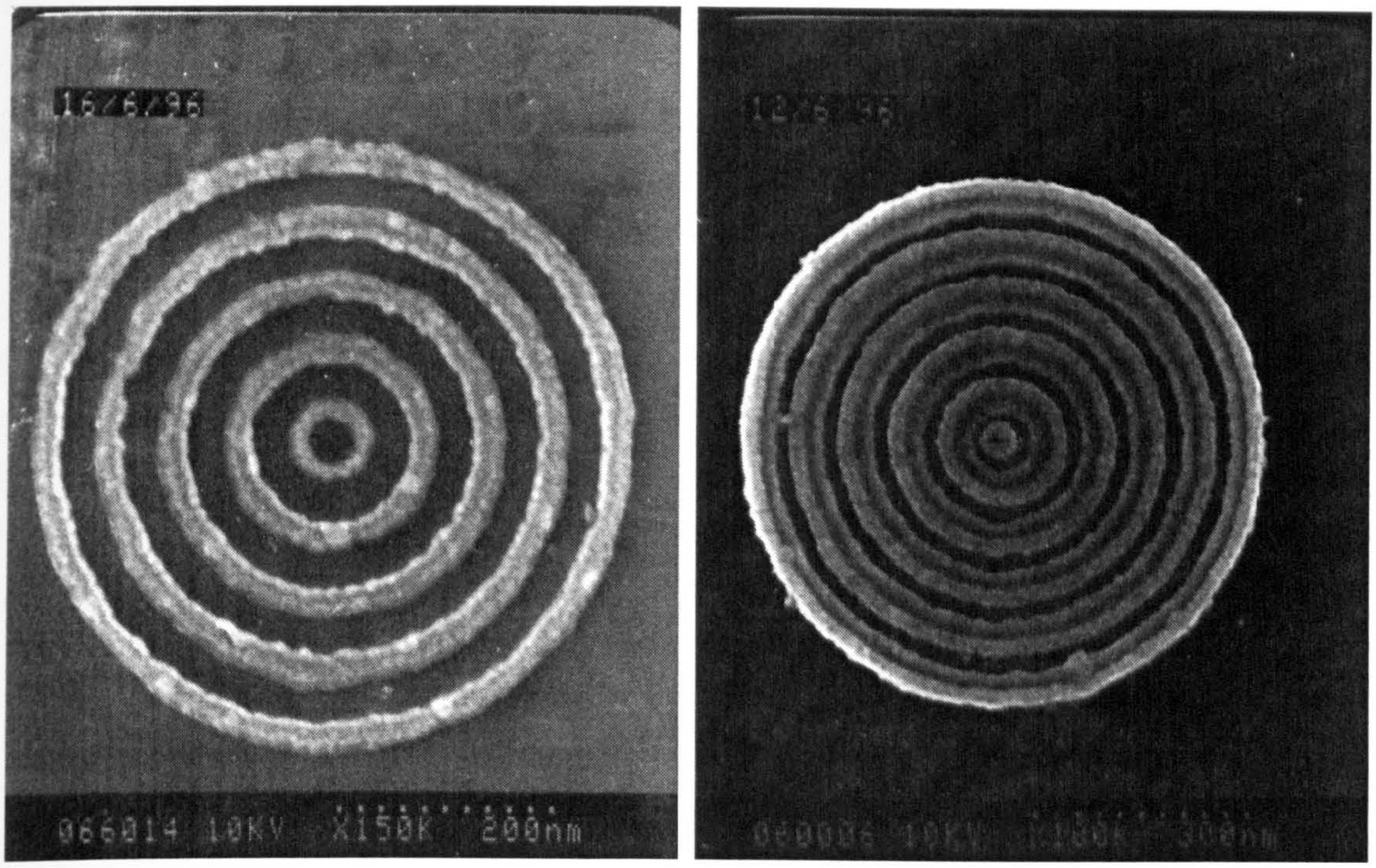


Figure 6.2.40: Micrographs illustrating an attempt to fabricate circular gratings in nichrome using lift-off from cross-linked PMMA ('ama0035z'). The samples were not coated with a charge conduction layer. (a) Left Image: a successful lift-off; (b) Right Image: a failed lift-off.

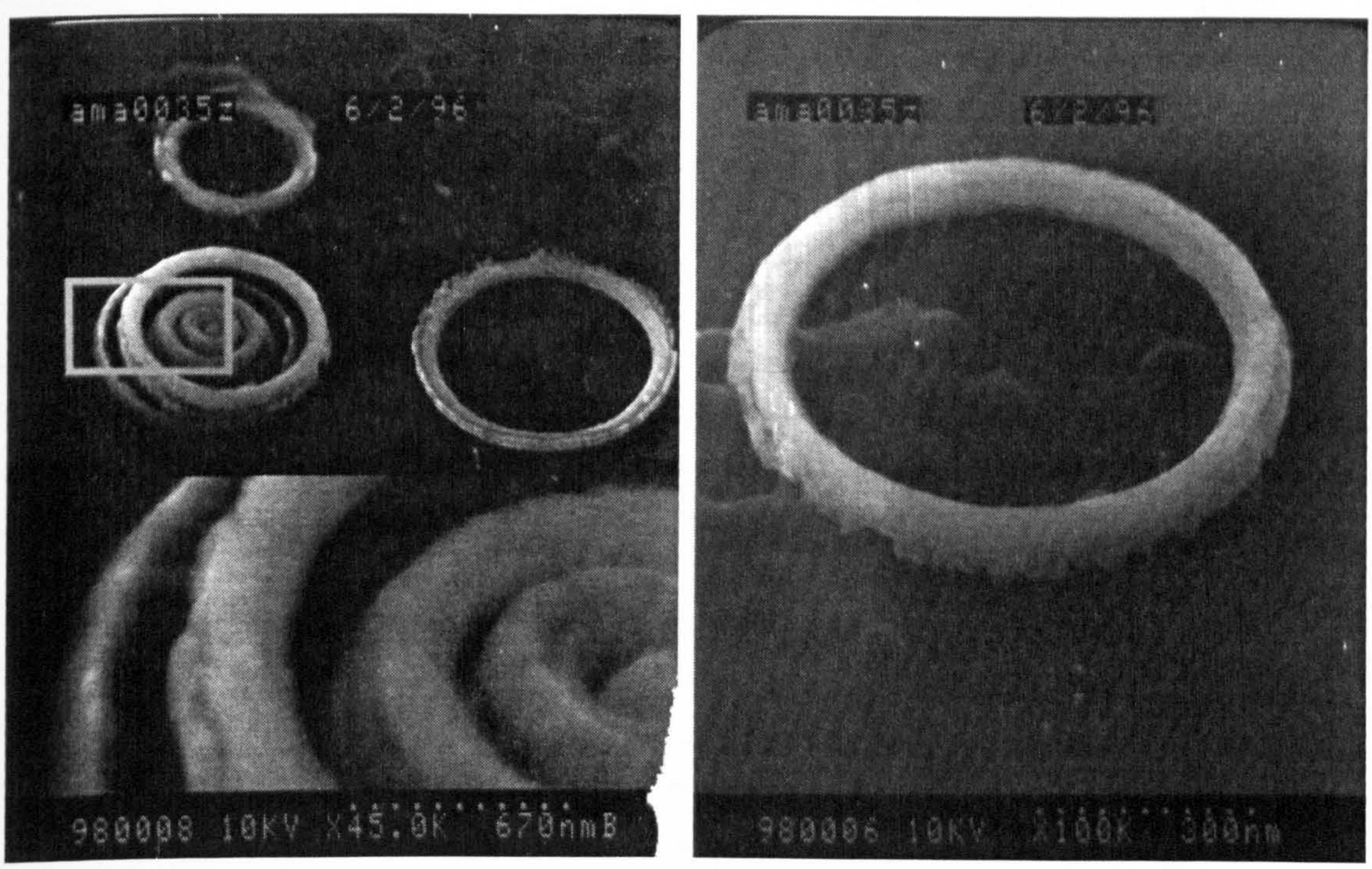


Figure 6.2.41: Micrographs illustrating an attempt to fabricate circular gratings in nichrome using lift-off from cross-linked PMMA ('ama0035z'). The samples were not coated with a charge conduction layer. (a) Left Image: a partial lift-off. Some rings of nichrome appear to have lifted-off and resettled on the substrate surface. The bottom half is a 5x magnification over the displayed scale; (b) Right Image: a close-up of a resettled nichrome ring. This one appears to have settled on the surface right-side up.

[As an aside, one test which was carried out to determine some of the lithographic possibilities with this new technique, was the exposure and nanostrip lift-off of concentric rings (*i.e.* a circular grating). Figure 6.2.40 shows both a successfully and unsuccessful lifted off grating. However, one unusual aspect of these patterns was the situation when only partial lift-off was achieved. This is illustrated in Figure 6.2.41 where some of the lifted-off rings have settled onto the sample surface, presumably stuck to it by capillary action during the drying of the sample. Figure 6.2.41b shows a close-up of such a ring which has also managed to remain upright, unlike those of Figure 6.2.41a.]

The way forward now seemed more clear. It was thought that a determination of the threshold dose required to form a cross-linked PMMA dot would indicate the resolution limits imposed by this technique. There was also some debate as to whether charging effects played any role in the determination of feature size and it was also speculated that a single layer of PMMA resist may develop the necessary overhanging profile for lift-off to succeed without the need for an additional, higher molecular weight PMMA layer. To answer these questions a test was devised based on five samples:

(1) A bi-layer of PMMA consisting of 4% BDH and 4% Elvacite and with a 30nm layer of nichrome deposited by the Plassys MEB 450 on top as a charge conduction layer.
(2) A bi-layer of PMMA consisting of 4% BDH and 2.5% Elvacite and with a 30nm layer of nichrome deposited by the Plassys MEB 450 on top as a charge conduction layer.
(3) A bi-layer of PMMA consisting of 4% BDH and 4% Elvacite and with no charge conduction layer.
(4) A bi-layer of PMMA consisting of 4% BDH and 2.5% Elvacite and with no charge conduction layer.
(5) A single layer of 4% BDH with no charge conduction layer.

In each case the resist was spun at 5000 rpm for 60 seconds, the samples were baked overnight at 180°C, and all bi-layers also had a 180°C, one hour bake between each layer. A pattern was then devised as follows:

(a) a basic unit of four 2μm diameter circles (positively exposed) with a single 5nm pixel at the centre (negatively exposed).
(b) the basic unit of (a) was used to form a five by five array, in which the cell at the bottom left was exposed once (no overlaps), while that at the top right was exposed 25 times in succession (24 overlaps)
(c) the pattern of (b) was then exposed as an exposure test in an 8 by 8 arrangement (75μm step and repeat distance) with a multiplicative increment of 1.052 (*i.e.* a geometric progression) to give a range of 64 doses from 17,500μC/cm^2 to 426,609μC/cm^2 per overlap (since the top end of 25 times nominal exposure would correspond to a dose of 437,000μC/cm^2 at

the lowest exposure dose setting, this pattern effectively covered a range of doses from 17,500μC/cm² to 10,665,216μC/cm² to an accuracy of 5.2%)

(d) the pattern of (c) was then repeated in a 2 by 2 array (1mm pitch) to allow verification to be obtained of any threshold dose measured.

The samples were then exposed at 100kV. The (positive) circles were exposed with a 112nm spot, a dose of 800μC/cm² and a resolution of 50nm. The negative dots were exposed with a 12nm spot and a resolution of 5nm. There were 25,600 dots per sample, but despite this the total job time was only 0.55 hours per sample. This illustrates the dramatic effect of a reduced exposure area despite the low sensitivity of this technique.

The 30nm nichrome charge conduction layer was then removed using chrome etch [§3.7.3] on the appropriate samples. The samples were all then developed in IPA:MIBK 2.5:1 for 60 seconds, before 50nm of nichrome was evaporated on them using the Plassys MEB 450 evaporator. Nanostrip lift-off was then performed and the samples were examined in the Hitachi S900 SEM. Some of the results were surprising and some were as expected. The not unexpected result was that lift-off universally failed on all the patterns on the sample with the single PMMA resist layer. The surprising result was that all the samples which had the charge conduction layer appeared to be underexposed. All that was left was a slight bump at the centre of the positively exposed circles (up to a dose of 10,665,216μC/cm²). This suggested that the charge conduction layer had somehow hindered the cross-linking of PMMA. One explanation for this might be that heating by the electron beam is an important factor in the cross-linking process [Yasuda, *et al.*[50] (1994)] and that the 30nm metal layer had acted as a heat sink, by dissipating any heat generated over a much wider area. Another explanation that the charge conduction layer had somehow become electrically charged by the electron beam and then deflected the beam away seemed less likely as the samples are clamped into earthed holders in the beamwriter with the clamps firmly on the surface (*i.e.* making electrical contact with the charge conduction layer). In the absence of further tests it is difficult to come to any definite conclusion, however, one test would be to compare these results with that on a sample having the 30nm nichrome layer beneath the PMMA resist layer.

The remaining two samples both yielded some usable data. In the case of the sample coated with 4% BDH and 2.5% Elvacite, it was found that in one of the patterns, the first quartet where all four dots (out of four) worked corresponded to the 51st exposure dose setting with 15 overlaps. This is a dose of $17{,}500 \times 1.052^{50} \times 15 = 3{,}310{,}673\mu C/cm^2$. Thus a safe dose would be 3,600,000μC/cm² corresponding to 9 overlaps with the maximum 400,000μC/cm² dose setting. With the sample coated with 4% BDH and 4% Elvacite it was found that the first quartet where all four dots worked corresponded to the 64th exposure dose with 13 overlaps. This is a dose of $17{,}500 \times 1.052^{63} \times 13 = 5{,}545{,}912\mu C/cm^2$. Thus a safe dose would be 6,000,000μC/cm² corresponding to 15 overlaps with the maximum 400,000μC/cm² dose setting. This also verified the observation that the threshold dose scaled with the thickness of the resist. That this

relationship is direct proportion cannot be verified on the basis of two points. [The author notes that, from Table 3.5.1, a bi-layer of 4% BDH, 2.5% Elvacite has a nominal total thickness of 109nm, while that of 4% BDH, 4% Elvacite is 169nm. Based on the ratio of these two, if the dose required for the thinner resist coating is as above, the expected dose for the thicker resist coating would be 5,133,062μC/cm^2, which is within 10% of the measured value.]

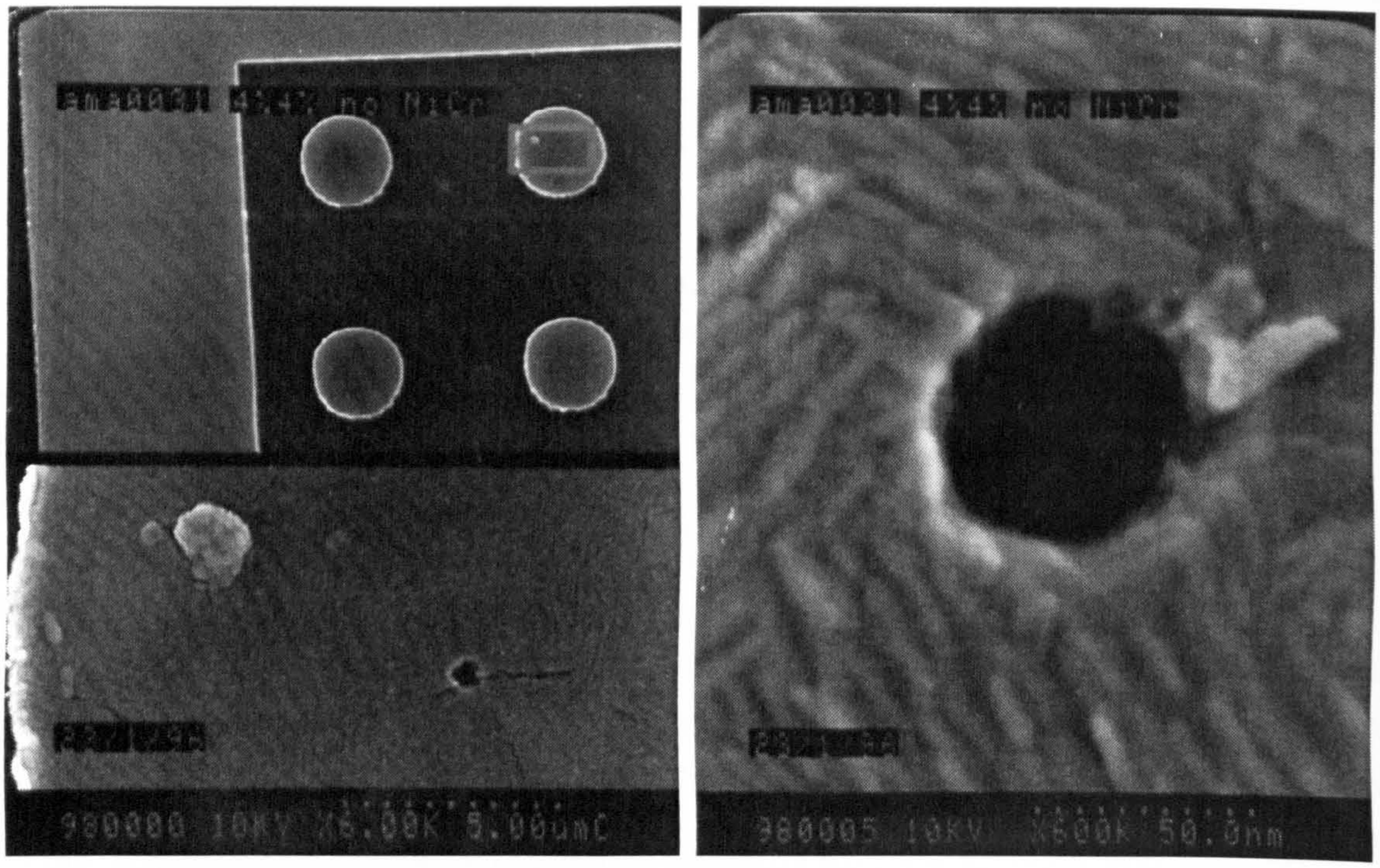


Figure 6.2.42: Micrographs showing some results from a test investigating lift-off using cross-linked PMMA as a function of dose ('ama0031'). This sample used a 4% BDH, 4% Elvacite bi-layer resist coating. (a) Left Image: highest dose, 25 overlaps. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: highest dose, 13 overlaps. See text for details.

Figure 6.2.42a shows the quartet on the sample with 4% BDH, 4% Elvacite corresponding to the highest dose setting with 25 overlaps. Figure 6.2.42b shows a close-up of one of the holes on the same sample corresponding to the highest dose setting with 13 overlaps. This hole is about 55nm in diameter. Figure 6.2.43a shows a close-up of another hole in the same quartet – it is virtually identical to that of Figure 6.2.42b. Finally, Figure 6.2.43b shows a 40° tilted close-up view of a hole corresponding to the highest dose setting with 14 overlaps in a different pattern on the same sample. This hole appears to be a little bigger. The minimum hole diameter was found to be almost the same on both samples, despite using different thicknesses of resist.

The final set of tests before attempting to use this technique in fabricating a functional SNOM device was to determine which resist should be used with the float coating technique [§5.2]. A sample was prepared which would also allow statistics to be obtained regarding aperture variability, as had been done with the two previous fabrication methods [§6.2.1 and §6.2.2]. This consisted of defining the apertures on silicon nitride beams as had been done previously [see §6.2.1].

The pattern consisted of:

(1) 32 overlaps of a single 5nm pixel repeated 3 x 10 times

(2) 10 overlaps of a 10nm square repeated 3 x 10 times

(3) 5 overlaps of a 15nm circle repeated 3 x 10 times

(4) 3 overlaps of a 20nm circle repeated 3 x 10 times

(5) 2 overlaps of 25nm, 30nm, 35nm, 40nm and 45nm circles each repeated 3 x 10 times

(6) 1 'overlap' of a 50nm circle repeated 3 x 10 times

(7) 2 overlaps of 60nm, 70nm, 80nm, 90nm, 100nm, 110nm, 120nm, 130nm, 140nm, 150nm, 160nm, 170nm, 180nm, 190nm, 200nm, 210nm, 220nm, 230nm, 240nm and 250nm circles each repeated 3 x 10 times.

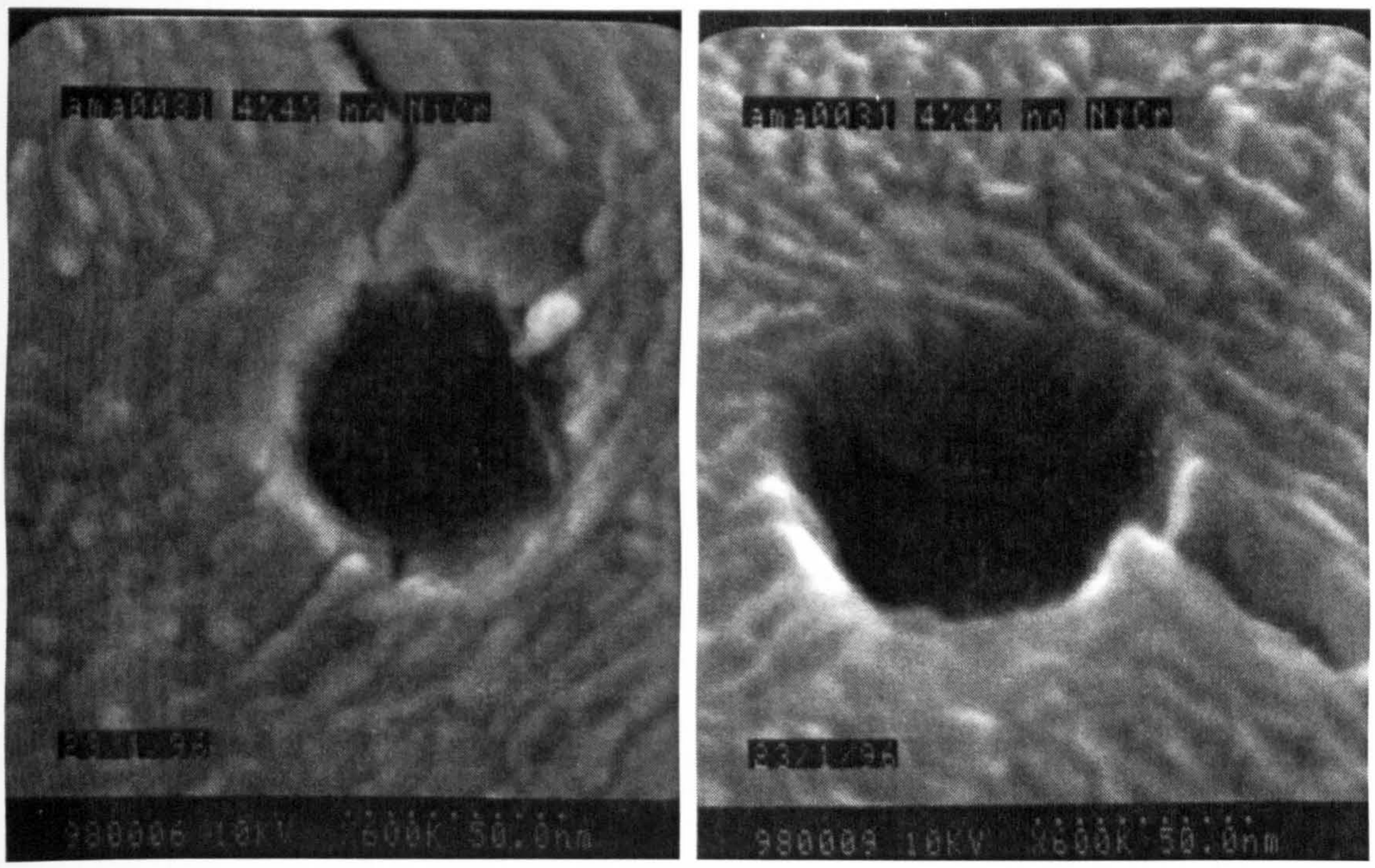


Figure 6.2.43: Micrographs showing some results from a test investigating lift-off using cross-linked PMMA as a function of dose ('ama0031'). (a) Left Image: another aperture from the same quartet as Figure 6.2.42b; (b) Right Image: highest dose, 14 overlaps. 40° tilted view. See text for details.

The number of overlaps used for each structure was approximately reduced with the pattern area (square of pattern diameter) with a minimum number of overlaps of two (the 50nm circle pattern was mistakenly left as a 'single overlap'). This was done so that the total exposing charge was independent of pattern – this was a crude way of taking into account the additional exposure by scattered electrons expected with larger structures. The sample used had a 100nm thick LPCVD silicon nitride layer and was processed as follows:

(a) 2.5% equivalent thickness Aldrich was float coated onto the substrate [§5.2].

(b) The resist was dried by baking for 30 minutes at 80°C.

(c) 1.5% Elvacite was float coated onto the substrate [§5.2].

(d) The resist was dried by baking for 30 minutes at 80°C.

(e) The resist was exposed using a resolution of 5nm and a spot of 12nm at 100kV, with global alignment. The pattern was exposed with a dose of 400,000μC/cm^2 per overlap.

(f) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(g) 50nm of nichrome was evaporated onto the resist using the Plassys MEB450.

(h) Nanostrip lift-off (two applications) was carried out.

(i) The sample was dry etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 10 minutes in the Plasma Technology BP80 (only 100nm of LPCVD silicon nitride to etch through, with an additional 300nm of silicon to allow shielding metal to be deposited).

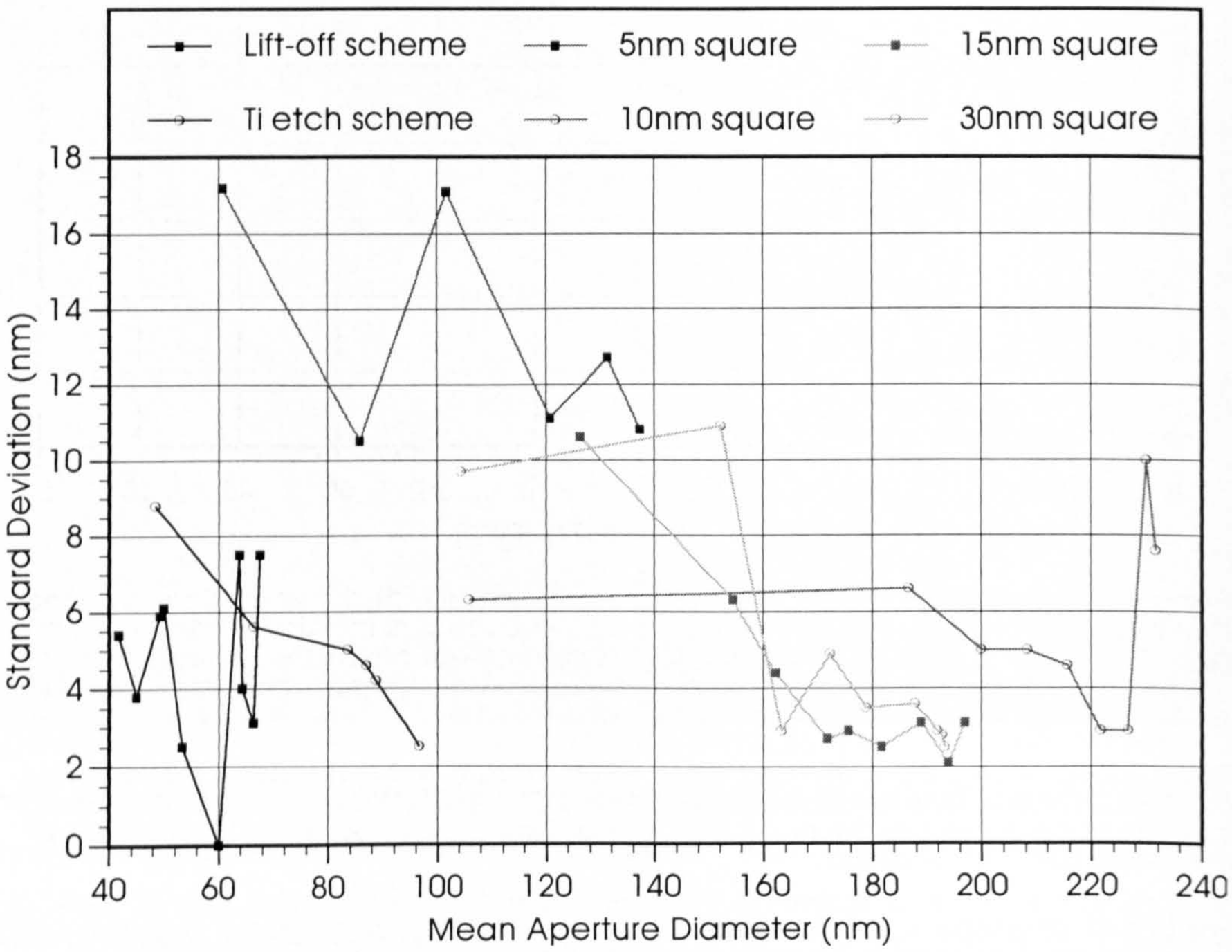


Figure 6.2.44: Comparison of the aperture diameter variability with mean aperture diameter between the nichrome lift-off scheme, the titanium etching scheme with a single pixel pattern and the direct etching scheme with the smaller square patterns [see Figures 6.2.17 and 6.2.27]. The points associated with each pattern style and size have been connected together by lines for clarity.

After etching, the nichrome masking layer was stripped off in chrome etch and, as the whole pattern had been repeated several times across the sample, the sample was split up and one part was wet etched in potassium hydroxide (as in §6.2.1). A 10nm thick layer of nichrome was then evaporated onto it. This piece was then examined in the Hitachi S900 SEM. It became clear that the yield was extremely low for most of the large aperture sizes, with many being occluded or partially occluded. [This may have been

because the dose had been insufficient, perhaps because the resist was substantially thicker than that used in the previous tests or because the crude manner in which the number of overlaps had been scaled down with pattern area underestimated the dose required for larger patterns.] However, at small aperture sizes the results were much better. Figure 6.2.44 shows the aperture diameter variability for the smaller apertures as compared to the titanium etching and direct PMMA definition schemes. It can be seen that for small aperture sizes, the (nanostrip) lift-off scheme is the best with an aperture variability of less than 6nm at a diameter of 42nm.

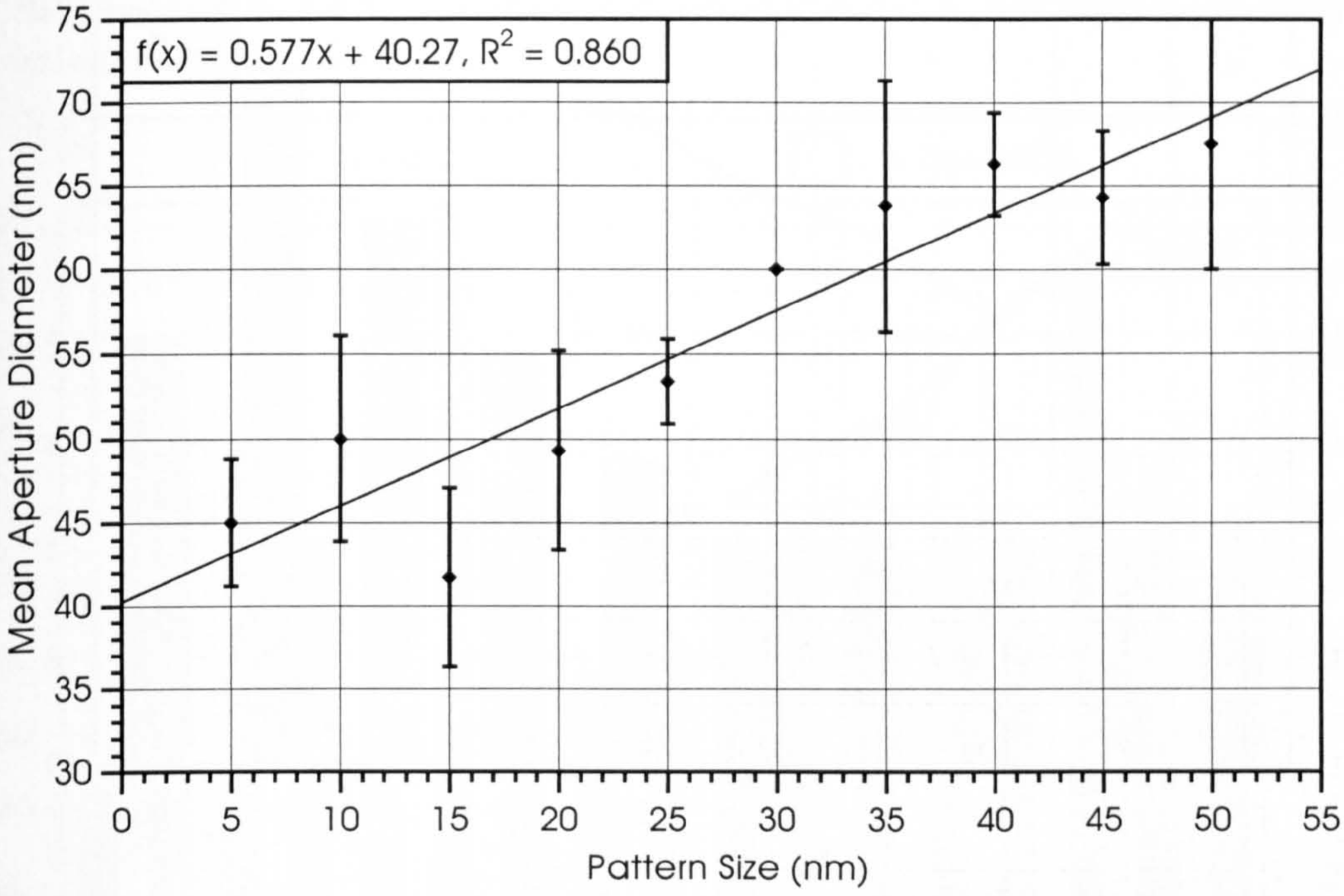


Figure 6.2.45: Variation of the mean aperture diameter with drawn pattern size. The variability in aperture diameter is represented by error bars marking a range of one standard deviation above and below the mean. A line has been fitted by least squares giving the mean aperture diameter, f, as a function of the pattern size, x, to aid in extrapolation to larger diameters.

Figure 6.2.45 shows how the variability of aperture diameter using the new scheme varies with the aperture diameter and the pattern used. It is interesting to note that the aperture size achieved using a drawn circle with a diameter of 15nm produced a smaller aperture than that with a single 5nm pixel. The smallest aperture diameter on this sample using the single 5nm pixel pattern was 37.5nm, while that with both the 10nm square and 20nm circular patterns was 45nm, and 30nm was achieved with the 15nm circular pattern (which turned out to be the smallest aperture examined on the sample). For patterns larger than this the smallest aperture with each pattern then increased more uniformly with drawn diameter. This suggests that further investigations need to be carried out to find out how the threshold dose varies with pattern size: as mentioned earlier the dose assignments (number of overlaps) for each pattern were done in a crude and ad hoc manner. Figure 6.2.46 shows the total exposing charge for each pattern size and the mean size of the resulting aperture. The author speculates that for pattern sizes

much smaller than the cross-linked dot size, the threshold dose required corresponds to a fixed total deposited charge independent of pattern size (this was part of the motivation behind the crude dose assignment scheme used). The only evidence that this is the case comes from the correlation between the mean aperture diameter and exposing charge in Figure 6.2.45 (excluding the result for a drawn pattern size of 50nm). For larger sizes this relationship may break down. However, as the doses were fixed for each pattern in this experiment it is clear that further tests are required before a theory can be advanced with any degree of confidence.

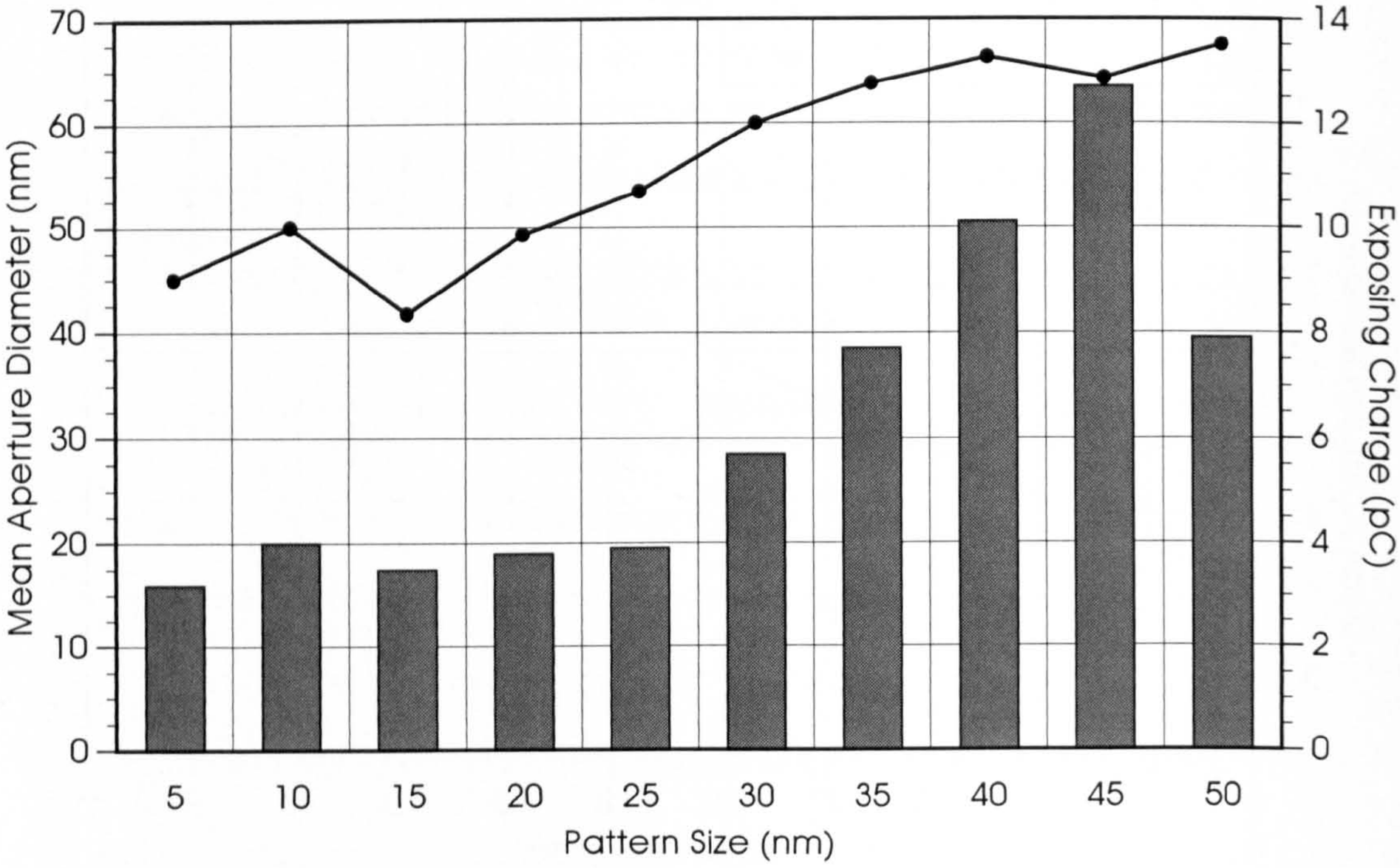


Figure 6.2.46: Variation of the mean aperture diameter (line) with drawn pattern size and exposing charge (bars).

At this point it was decided, that although the yield in the larger apertures was low, it would be worthwhile finding out how much the apertures closed up by having a metal shielding layer deposited on top. This would give an indication as to the range of aperture sizes required in the silicon nitride layer to give the desired range in the metal shielding layer. Therefore, with one of the remaining pieces, a metal shielding layer consisting of a 10nm nichrome sticky layer followed by a 200nm palladium shielding layer was evaporated using the Plassys MEB 450. Palladium was chosen because it meant that the aperture shape would be better defined than would be the case with aluminium, as already discussed in §6.2.1.

Figure 6.2.47 shows the mean aperture diameter after the shielding layer is deposited against the drawn circle diameter. Also shown are the yield and variability (as a fraction of the mean aperture diameter). It should be noted that some of the smaller apertures were achieved because the aperture was partially occluded by a partial lift-off of the cross-linked PMMA layer prior to dry etching. The diameter of such non-circular apertures was estimated to be the average of the largest and smallest diameter. Apertures

which failed did so as a result of unsuccessful lift-off. A high proportion of the apertures were above 50% yield, but it was felt that given the robustness demonstrated by the technique so far it ought to be possible to get this yield much closer to 100%. It was noted that the variability in aperture diameter for physical diameters above 100nm was largely below 10%, which was acceptable. The increase in the variability of aperture diameter for sub-100nm (physical) diameter apertures was due to the large number of occluded apertures and the method used to estimate their diameter. The variability in aperture diameter could be reduced by the same measures which improve the yield of these apertures, if it is assumed that these measures would also reduce the number of occluded apertures.

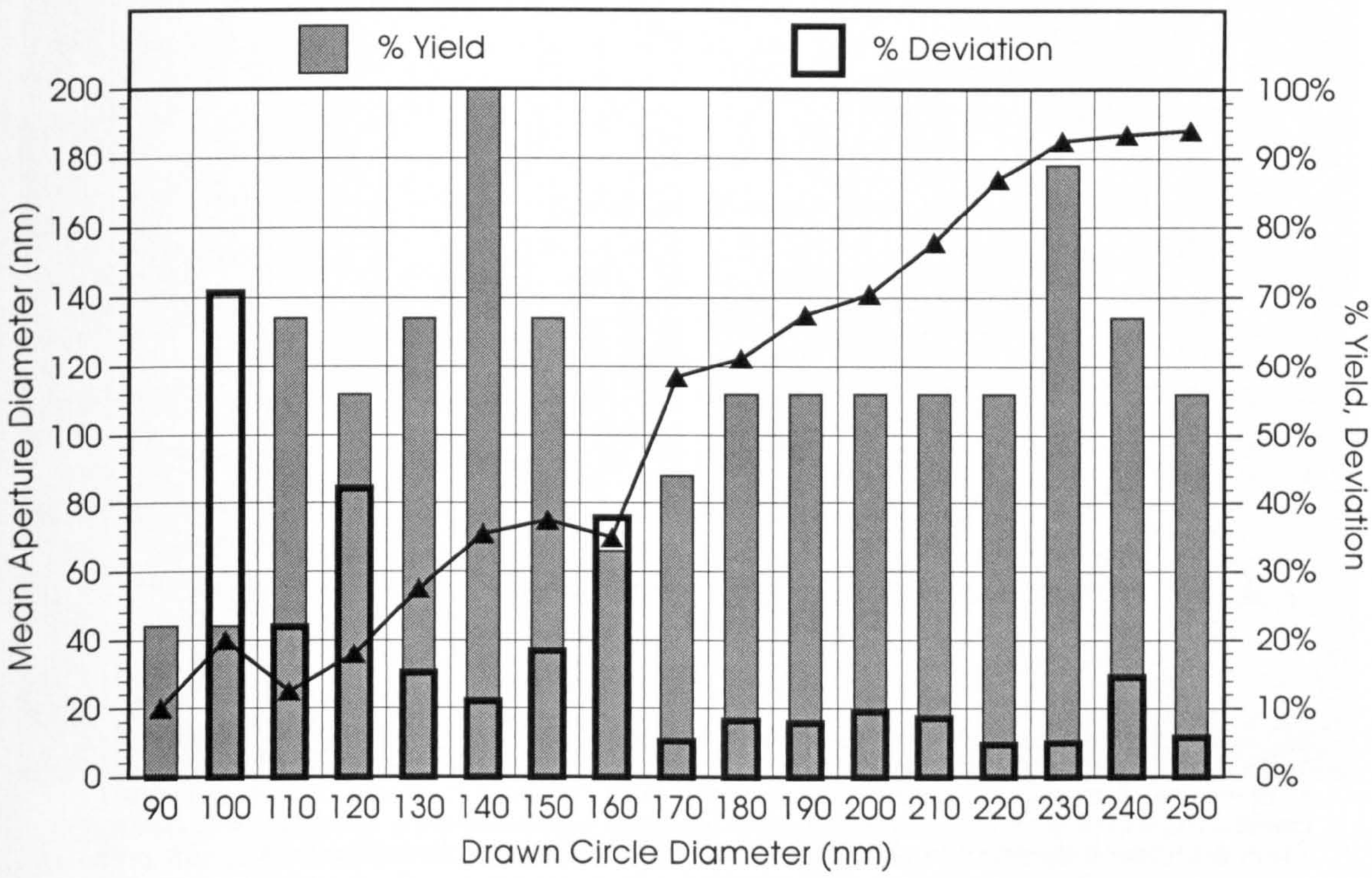


Figure 6.2.47: The mean aperture diameter (line), percentage yield (grey bars) and percentage deviation (clear bars) achieved with a 200nm palladium shielding layer as a function of the drawn circle diameter.

Figure 6.2.48 shows the aperture diameter variability (in absolute terms) and mean aperture diameter as a function of the drawn aperture diameter. A straight line seems to reasonably describe the variation in mean aperture diameter with drawn circle diameter. [This fit is clearly much better than that in Figure 6.2.45.] This line can be used to give a best guess at the decrease in aperture diameter caused by the deposition of the shielding layer. If the line is extrapolated to a zero drawn circle diameter, the mean aperture diameter corresponds to a drawn circle diameter of -94.57nm (to two decimal places). If the line is extrapolated to the point where the mean aperture diameter is zero, the corresponding drawn circle diameter is 80.27nm (to two decimal places). Thus, on average, the physical aperture diameter is reduced by 87.5nm (to the nearest 0.5nm) from the drawn circle diameter, for a deposition layer 210nm thick. From Figure 6.2.45, it can be seen that for the largest pattern shown, the physical aperture diameter is

17.5nm larger than the drawn circle diameter of 50nm. Using this as an estimate of the minimum amount by which the physical aperture diameter in the silicon nitride exceeds the drawn diameter of the pattern, the change in the physical aperture diameter (from that obtained in the silicon nitride layer) as a result of the deposition of the metal shielding layer is estimated to be 70nm. This equates to a tapering angle (angle between the sides of the aperture and the perpendicular to the plane of the aperture) of about 9.5° per side, or a full cone angle (as defined in §6.5) of 19° (to the nearest degree).

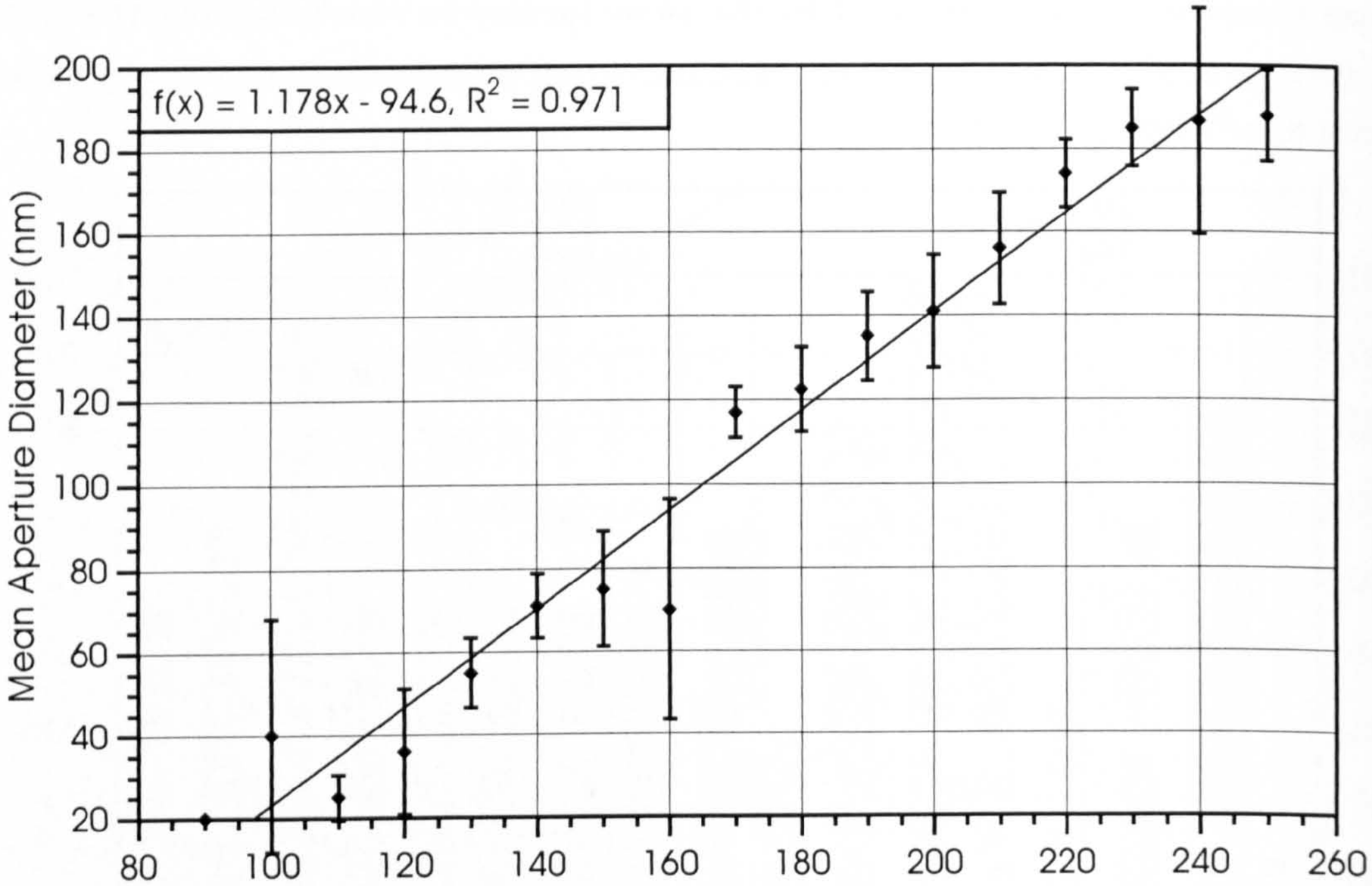


Figure 6.2.48: The variability in aperture diameter and mean aperture diameter as a function of drawn circle diameter with a 200nm palladium shielding layer. The variability in aperture diameter is represented by error bars marking a range of one standard deviation above and below the mean. A line has been fitted by least squares giving the mean aperture diameter, f, as a function of the drawn circle diameter, x.

From this estimate, the physical aperture diameter required in the silicon nitride to achieve a physical aperture diameter of 50nm in the shielding layer is 83nm if 100nm of aluminium is used and 120nm if 10nm of nichrome and 200nm of palladium is used as a shielding layer. From the preceding discussion, this would then correspond to drawn circle diameters of 66nm and 103nm (to the nearest nm) respectively.

Figure 6.2.49 shows two micrographs of apertures on this sample. The upper half of Figure 6.2.49a shows a partially occluded aperture while the lower half shows an aperture around 50nm in diameter. Again it is noticeable that the edge roughness is dictated by the grain size of the metal. Figure 6.2.49b shows one of the smallest better shaped apertures. This is around 35nm in diameter.

To improve the yield and variability of larger apertures, another test was carried out. It was thought that using a thicker resist would also aid nanostrip lift-off by improving chemical access to the resist. The pattern consisted of:

(1) 64 overlaps of a single 5nm pixel repeated 3 x 10 times
(2) 20 overlaps of a 10nm square repeated 3 x 10 times
(3) 10 overlaps of a 15nm circle repeated 3 x 10 times
(4) 6 overlaps of a 20nm circle repeated 3 x 10 times
(5) 4 overlaps of 25nm, 30nm, 35nm, 40nm and 45nm circles each repeated 3 x 10 times
(6) 2 overlaps of a 50nm circle repeated 3 x 10 times
(7) 4 overlaps of circles with diameters from 60nm to 350nm in increments of 10nm each repeated 3 x 10 times.

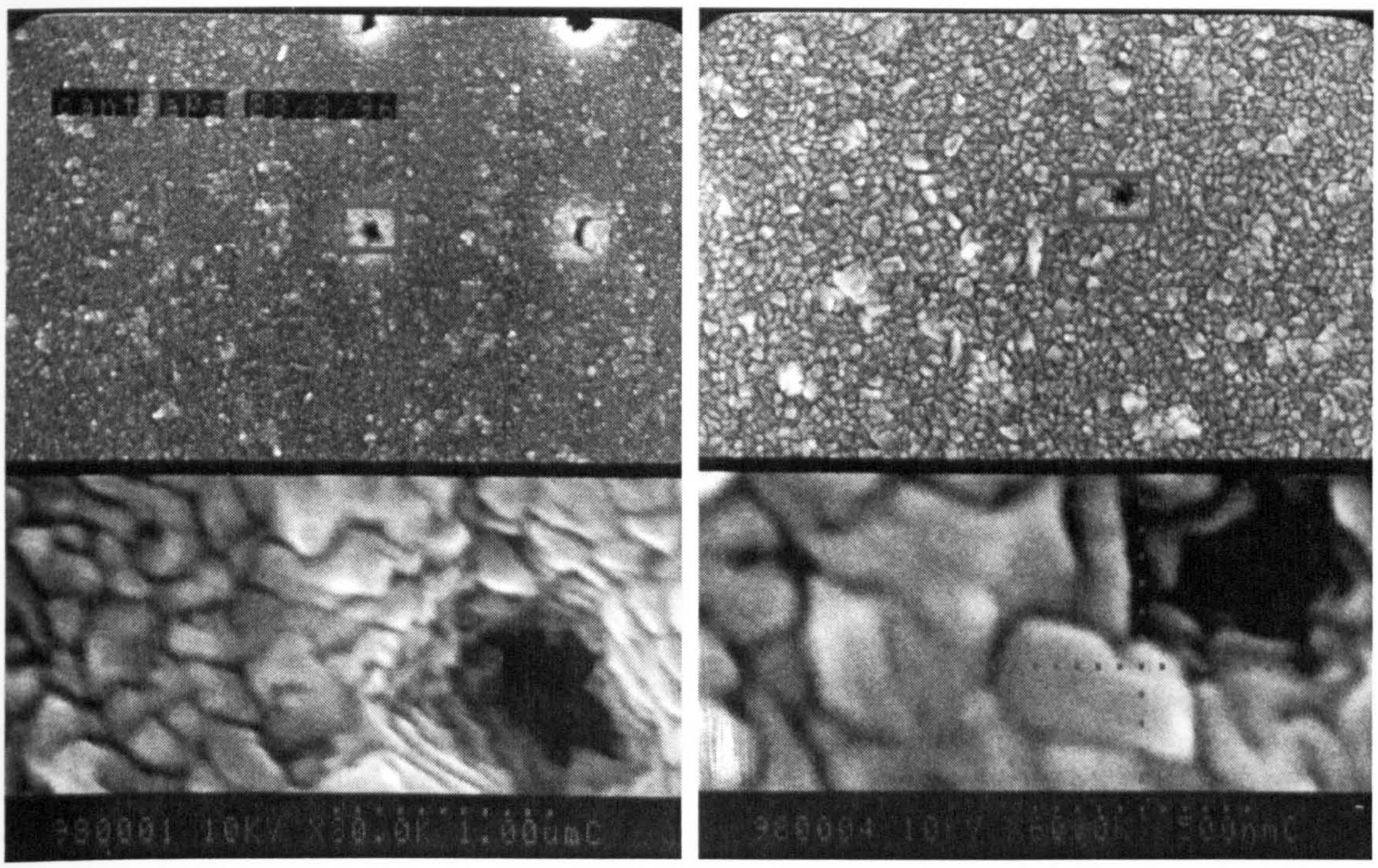


Figure 6.2.49: Micrographs showing some apertures in 200nm thick palladium made using the cross-linked PMMA scheme. (a) Left Image: four apertures from the same pattern exposed with the same dose. The bottom right aperture is partially occluded; (b) Right Image: one of the smallest apertures achieved on this sample. The bottom images are a 10x magnification over the displayed scales which apply to the top images. See text for details.

The attempt was made with a bi-layer of PMMA consisting of 4% equivalent thickness Aldrich and 2.5% Elvacite float coated onto a sample with 100nm of LPCVD silicon nitride. The sample was then exposed, developed and evaporated upon (50nm nichrome) as before. Nanostrip lift-off was then carried out (this time with three applications in place of two) before the sample was examined in the Hitachi S900 SEM. As can be seen from Figure 6.2.50 the larger apertures with 4 overlaps seemed to have lifted off with a good yield, however, the smallest of these larger apertures, with a drawn diameter of 60nm, has resulted in a physical diameter of about 150nm, as seen from Figure 6.2.50b. This would be a reasonable size if a 200nm thick palladium shielding layer was used. The large growth suggests that the dose may have been on the high side, and that the actual reason the yield was low in the previous test was, not so much insufficient exposure, as an insufficient number of applications of nanostrip during lift-off.

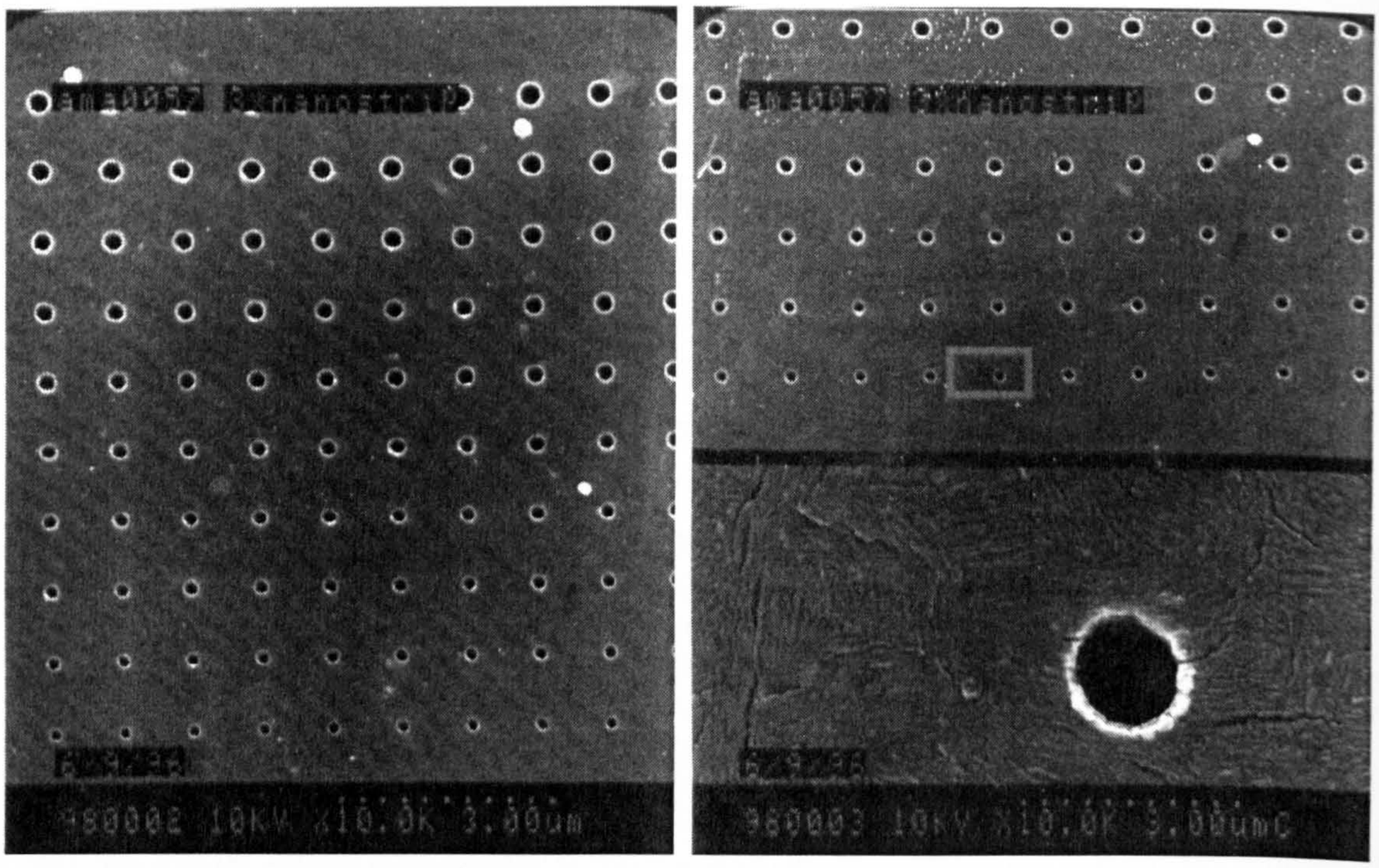


Figure 6.2.50: Micrographs showing some test apertures in 50nm thick nichrome made using the cross-linked PMMA scheme ('ama0057'). (a) Left Image: a much improved yield at larger aperture diameters; (b) Right Image: a close-up of one of the smallest apertures on this sample. The bottom half is a 10x magnification over the displayed scale. See text for details.

The result with the smaller apertures was quite different on this sample. It appeared that many of them had been underexposed. This was probably as a consequence of the thicker resist layer. From Table 5.2.1, the resist thickness is expected to be in excess of 600nm. Thus it was decided that a thinner resist layer be used and a further test was then carried out.

This attempt was made with a bi-layer of PMMA consisting of 3.25% equivalent thickness Aldrich and 2% Elvacite float coated onto a sample with 100nm of LPCVD silicon nitride. After each layer the sample was baked at 80°C for 30 minutes. The pattern consisted of:

(1) 64, 128 and 192 overlaps of a single 5nm pixel each repeated 10 times
(2) 20, 40 and 60 overlaps of a 10nm square each repeated 10 times
(3) 10, 20 and 30 overlaps of a 15nm circle each repeated 10 times
(4) 6, 12 and 18 overlaps of a 20nm circle each repeated 10 times
(5) 4, 8 and 12 overlaps of each of 25nm, 30nm, 35nm, 40nm and 45nm circles each repeated 10 times
(6) 2,4 and 6 overlaps of a 50nm circle each repeated 10 times
(7) 4, 8 and 12 overlaps of circles with diameters from 60nm to 350nm in increments of 10nm each repeated 10 times.

This sample was then exposed and processed as before. As seen from Figure 6.2.51 the smaller apertures with the least number of overlaps (as given above) seemed to have lifted off with a good yield. Figure 6.2.51b shows one of the smallest apertures, corresponding to a 15nm diameter drawn circle (with 10 overlaps), which resulted in an aperture with a

physical diameter of 65nm. However, as can be seen from Figure 6.2.51a, the aperture corresponding to a drawn pattern diameter of 30nm (with 4 overlaps) resulted in an aperture with a physical diameter just under 75nm, which is almost ideal for a shielding layer of 100nm thick aluminium. These results suggest that for the smallest apertures, the best choice is a float coated bi-layer of 3.25% equivalent thickness Aldrich and 2% Elvacite with drawn circular pattern diameters from 30nm upwards and a total exposure dose of 1,600,000μC/cm^2.

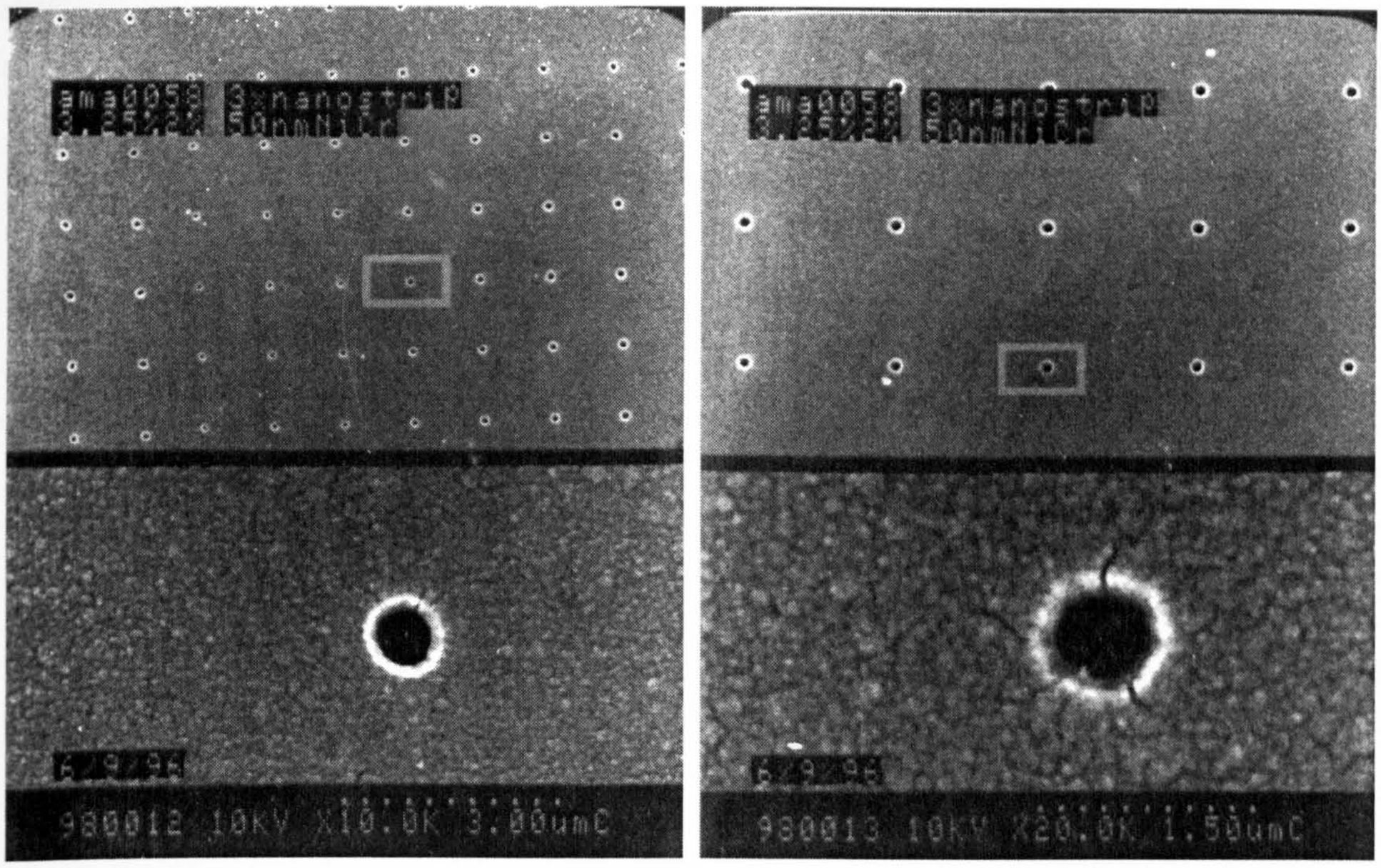


Figure 6.2.51: Micrographs showing some test apertures in 50nm thick nichrome made using the cross-linked PMMA scheme ('ama0058'). (a) Left Image: a close-up of an aperture corresponding to a drawn diameter of 30nm; (b) Right Image: a close-up of one of the smallest apertures on this sample. See text for details. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

A further observation made from the higher dose, smaller apertures showed that, although they had lifted off, they were more misshapen than those with a lower dose. This also seemed to be true of the lowest dose apertures derived from 5nm and 10nm square patterns. All this seemed to suggest that with extremely high doses, either thermal drift started to play a noticeable role in the shape of the apertures or that charging effects would result in the electron beam being deflected [Liu, *et al.*[30] (1995)]. The two micrographs in Figure 6.2.52, where the both sets of apertures were generated from identical patterns on the same sample and exposed in the same run (smaller aperture pattern) but with Figure 6.2.52b being at double the dose of Figure 6.2.52a can be compared. It is quite noticeable from Figure 6.2.52b, that other than the first two columns of apertures, all the apertures appear to be ellipsoidal in the same direction. At first this might be thought to rule out charging effects. However, the first two columns *also* show some misshapen apertures, except that here the directions and shape appear to be more random. This would seem to rule out thermal drift as the cause. These first

two columns correspond to patterns of a single 5nm square and a single 10nm square while the other columns are all circular patterns from 15nm in diameter to 50nm in diameter (in increments of 5nm). It is clear from Figure 6.2.52a, that other than the first two columns, all the apertures with half the dose have come out as very well-shaped circular apertures. The first two columns of apertures here also seem to have some degree of ellipticity, but not as pronounced, although just as random, as in Figure 6.2.52b. It can be speculated that these first two columns are overexposed in both cases, but that the other apertures are only overexposed in the second micrograph. This would help explain why the 15nm drawn circular patterns have produced smaller apertures, in both cases, than the 5nm and 10nm drawn squares.

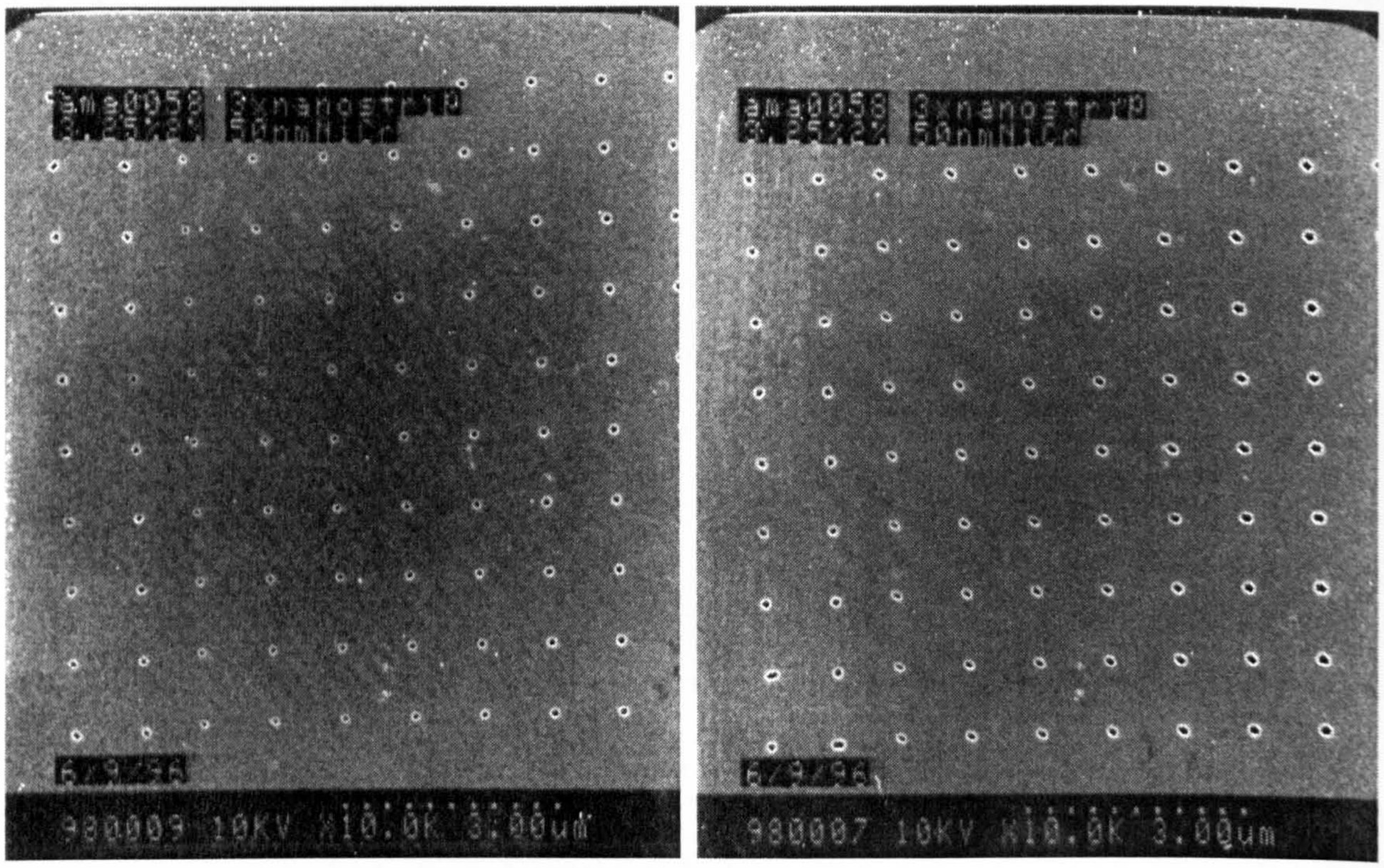


Figure 6.2.52: Micrographs showing some test apertures in 50nm thick nichrome made using the cross-linked PMMA scheme ('ama0058'). Investigation of the variability in aperture shape with dose. (a) Left Image: these apertures are less elliptical than in (b); (b) Right Image: these apertures were exposed with twice the dose of (a). See text for details.

[The author speculates that charging effects could be causing this problem, but that the direction of the deflection, which would normally be random, is dictated somewhat by the order in which the pixels making up the more composite patterns (circles) are exposed. If the beamwriter were to expose the circles from right to left and bottom to top in the case of the samples in question, it would be expected that charge would gradually build up towards the bottom and right of the pattern and cause the beam to deflect towards the top left, creating the effect as seen. In the case of the single pixel and four pixel square patterns, particularly the former, such a direction would not be as well defined and result in a more random deflection. Unfortunately, without more work, this cannot be verified.]

Many attempts were also made to define apertures using this technique on 100nm LPCVD silicon nitride membranes with through-wafer access so that light intensity

measurements could be made. Each time, if the apertures had lifted-off, the membranes would shatter spectacularly during dry etching. Several possibilities, such as air being trapped behind the membranes when the chamber was pumped out, were put forward. However, the most likely explanation was only found afterwards through scouring the literature. The previously successful attempts (using the direct PMMA technique) had been carried out on 50nm thick LPCVD silicon nitride membranes.

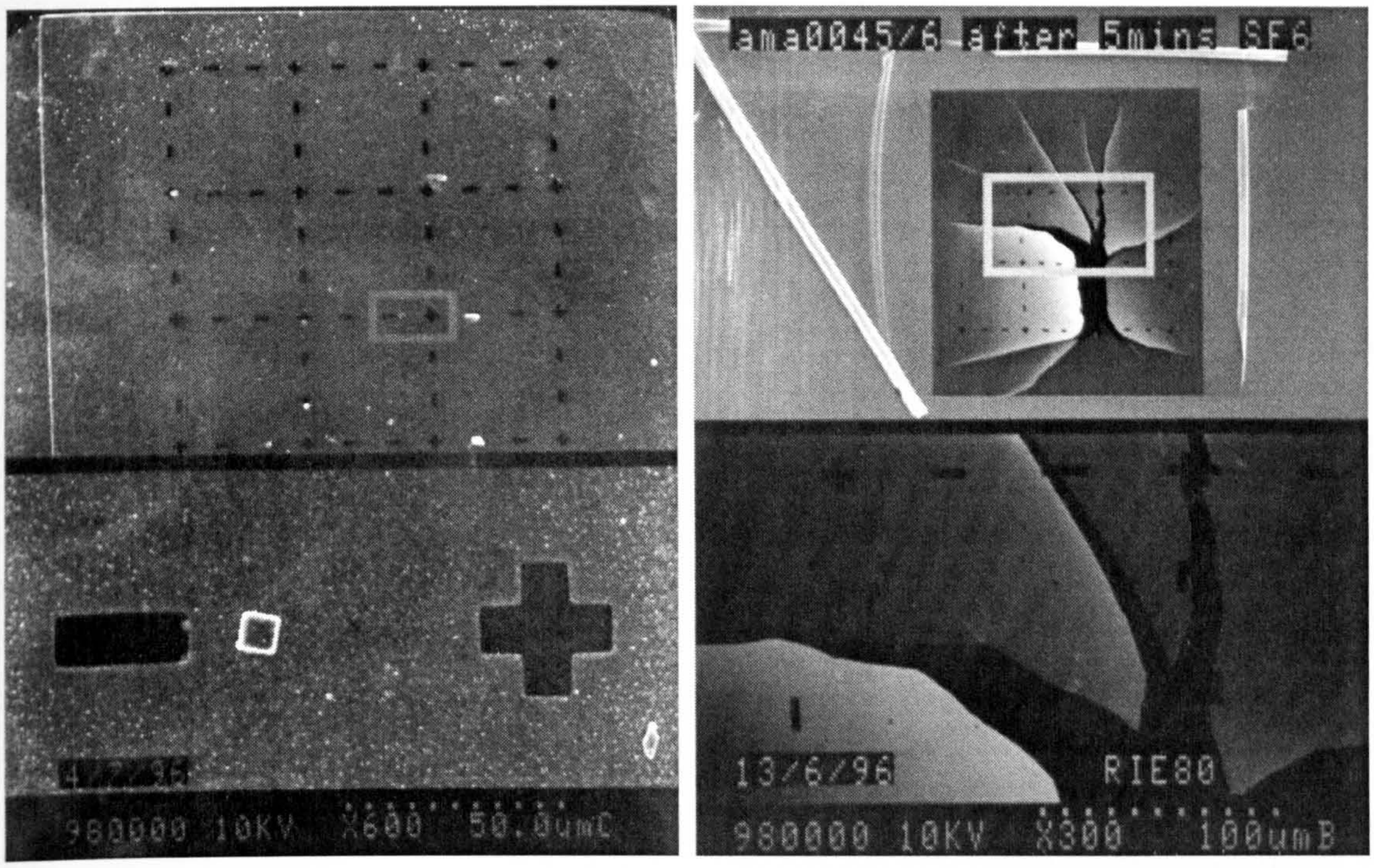


Figure 6.2.53: Micrographs showing the effects of stress in fracturing silicon nitride membranes in which apertures have been defined ('ama0045'). (a) Left Image: this membrane has miraculously survived despite the large location apertures having been etched through. Unfortunately none of the measurement apertures have been etched through. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: The measurement apertures have been etched through here but unfortunately the membranes have shattered. The bottom half is a 5x magnification over the displayed scale.

The probable cause of the shattering was that the residual stress within the 100nm thick silicon nitride was sufficiently high that when the apertures etched through, the membranes fractured. This phenomenon has been observed before with cracks tending to propagate across the membrane from stress centres formed during dry etching [Keith, Soave and Rooks[54] (1991)]. The pattern consisted of many rectangular and 'cross-shaped' (sharp cornered) apertures for location purposes which were observed to act as stress centres for the fracturing process. The remedy would have been to either stress relieve the membranes, *e.g.* by ion implantation [Wong and Ingram[55] (1992)], to remove the LPCVD silicon nitride layers and substitute them with 500nm thick, low stress, PECVD silicon nitride, or to design the pattern so as to minimise the stress concentration (*e.g.* by having fewer location apertures, and with more rounded corners). However, these solutions were found too late for any further tests to be carried out. Figure 6.2.53 shows two membranes, with one of them intact (the apertures had not lifted off, even though the larger crosses and dashes, used for location purposes, have been largely successful) and

the other one broken (some etched through apertures can just be made out upon close examination).

6.3 FABRICATION PROCESS INTEGRATION

The main issues in the integration of the processes developed in §6.2 were to do with arranging for the dry etching of the aperture and SNOM shield evaporation within the existing fabrication process. Due to the parallel development strategy being followed, many of the problems discussed in Chapters 4 and 5 were being addressed simultaneously with the developments of the last section. In particular, the aperture fabrication processes of the last section were being integrated with the cantilever fabrication process even while both were in the midst of being developed as separate processes.

One of the first attempts to fabricate an aperture on a pyramidal tip, (albeit one that had been defined on a substrate not already micromachined using the photolithographic processes of §4.2, and also missing a silicon nitride layer) was based on the titanium etching process of §6.2.2. At this time, as discussed in §5.1, there were problems with the method by which defocusing of the beamwriter could be achieved (misalignment caused by using defocus), so that the aperture was defined without the use of any defocus. The sample was processed as follows:

(a) 30nm of titanium was evaporated onto the sample in the Plassys MEB 450.
(b) 15% BDH was spin coated at 1000 rpm for 60 seconds.
(c) The resist was dried by baking at 180°C for one hour.
(d) 8% BDH was float coated on top of the previous layer [§5.2].
(e) The resist was dried by baking at 80°C for 30 minutes.
(f) The resist was exposed using a resolution of 5nm and a spot size of 12nm at 100kV, with local alignment to each of an array of pyramidal tips, and without defocus. The pattern consisted of (positively) exposing a single 5nm pixel on each tip with a range of doses from 15,000 to 30,491$\mu C/cm^2$ (exposure test).

Unfortunately, the apices of most of these pyramids had been damaged by mishandling, and so although a range of doses was used, only a few of the pyramids were useful for the purposes of determining whether the subsequent titanium etching had been successful. On the tip with the highest dose, at the centre of the apex, what looked like an aperture with a diameter of between 100 and 150nm could be seen, but not clearly, as shown in Figure 6.3.1a. There was some debate about whether the etching had actually penetrated through the titanium film. As was found subsequently, even with a spin-coated bi-layer of 8% BDH and 4% Elvacite (5000 rpm, 60 seconds, 180°C bake of one hour between each layer and overnight after the final layer), a dose of at least 20,000$\mu C/cm^2$ was

required to (positively) expose a single pixel [Figure 6.2.15]. With a thicker resist layer, the dose would be higher.

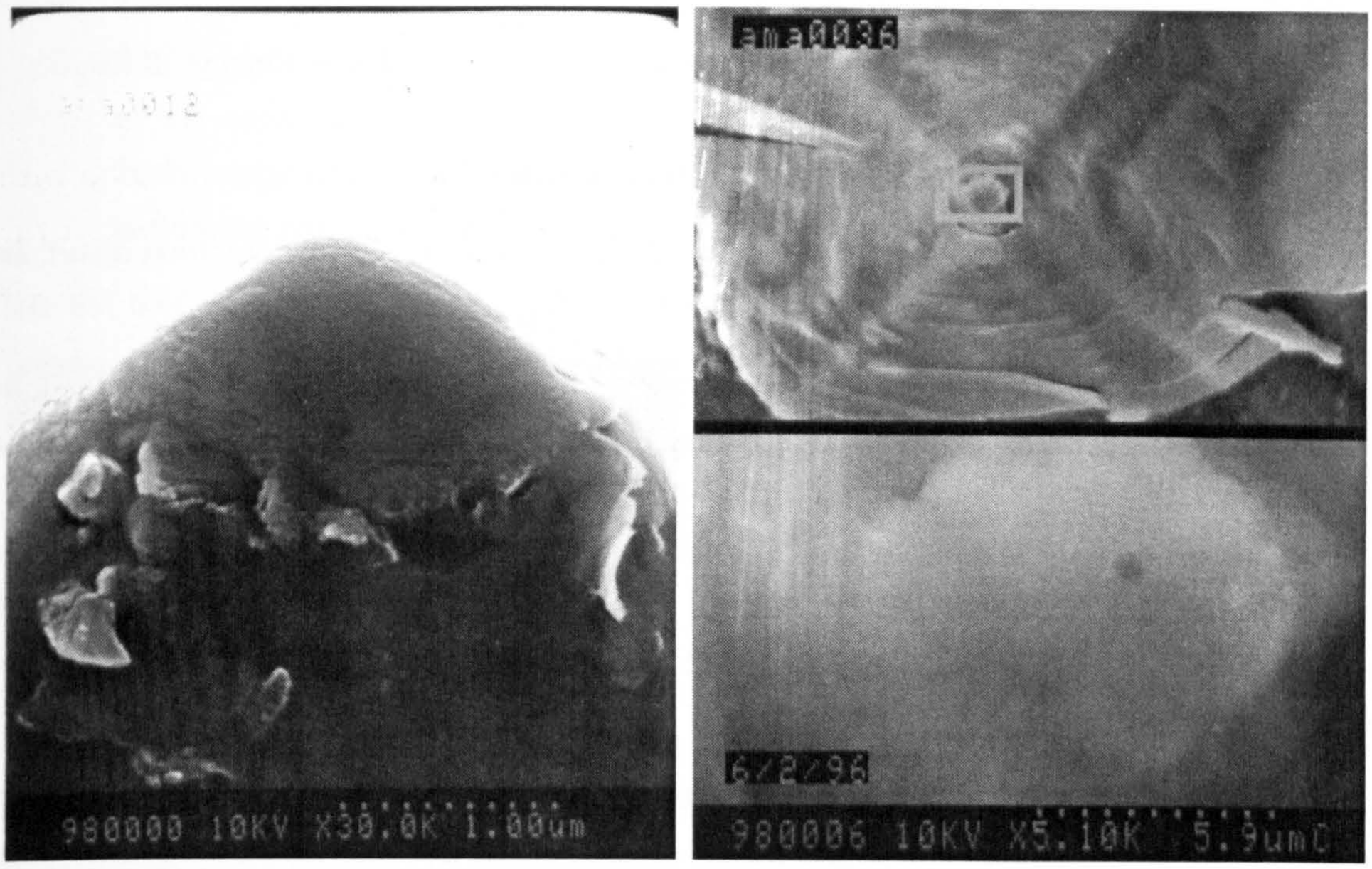


Figure 6.3.1: Micrographs showing early attempts at defining apertures on pyramidal tips. (a) Left Image: a tip upon which a 100nm to 150nm diameter aperture appears to have been defined using the titanium etching scheme ('ama0012'); (b) Right Image: a tip upon which an 80nm diameter aperture appears to have been defined using the cross-linked PMMA scheme ('ama0036'). The bottom half is a 10x magnification over the displayed scale.

The first attempt to define an aperture on top of a pyramidal tip using the cross-linked PMMA scheme was made after the defocus problem had been resolved (coincidentally), but prior to many of the exposure tests described in §6.2.3. A test was first carried out with a bad sample, suffering from many cracks [see §4.2.3 for an explanation], to determine the exposure dose required to produce an aperture on a pyramidal tip. The sample, (which was based on the 'amm0007' mask set as described in Appendix F) consisted of a 7 x 3 array of paired devices, and did not have a silicon nitride layer deposited onto it. The sample was processed as follows:

(a) 15% BDH was spin coated at 2500 rpm for 60 seconds.

(b) The resist was dried by baking at 180°C for one hour.

(c) 2.5% BDH was float coated on top of the previous layer [§5.2].

(d) The resist was dried by baking at 80°C for 30 minutes.

(e) 1.5% Elvacite was float coated on top of the previous bi-layer [§5.2].

(f) The resist was dried by baking at 80°C for 30 minutes.

(g) The resist was exposed using a resolution of 5nm and a spot size of 12nm at 100kV, with local alignment to each of an array of pyramidal tips, and with 20μm of defocus. The pattern consisted of (negatively) exposing a single 5nm pixel on each tip with a range of doses from 100,000 to

320,713μC/cm² per overlap (exposure test). 23 overlaps, corresponding to 24x exposure was used.

(h) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(i) 50nm of nichrome was evaporated onto the resist using the Plassys MEB 450.

(j) Nanostrip lift-off using two applications was carried out.

This resulted in what appeared to be an 80nm diameter aperture (at the highest dose), as shown in Figure 6.3.1b.

The results of this test suggested that an aperture defined in this way would be too small, (due to closure during evaporation of the shielding layer) as at this time 200nm of palladium was the preferred metal shielding layer. Therefore the pattern used to expose an aperture on the better sample was changed to a 'single overlap' of a 150nm diameter drawn circle. The resist structure was as with the test sample and the exposure conditions were identical except that the dose was fixed at 250,000μC/cm² and the defocus used was 18μm.

At the time when PMMA resist was being used to directly mask the cantilever definition dry etch (as opposed to the nichrome and chromium masking layers in subsequent evolutions of this lithographic level – see Appendix L (§L.1)), the idea was to use the cantilever definition dry etch to simultaneously define the aperture and cantilevers as it did not matter if the cantilever definition etch was extended in duration to accommodate the aperture definition etch (PMMA etches at about 50nm per minute and a 15% BDH layer should therefore be capable of surviving for 25 minutes when spun at 5000 rpm – see Table 3.5.1). With the cross-linked PMMA process and the use of a nichrome masking layer for the cantilever definition etch, it was decided that the same lithographic level as used to define the aperture (negatively) also be used to define the cantilever shape by exposing this region (positively), *i.e.* the nichrome lift-off step would leave a triangular cantilever-shaped nichrome layer, encompassing the pyramidal tip and with an aperture at the apex of the tip. This cantilever level was exposed as a separate matrix [§3.5] with a spot size of 300nm and a dose of 1200μC/cm² (high dose because the resist is extremely thick, exceeding 2μm, and the beam energy is 100kV).

After exposure, the sample was developed and lift-off carried out as with the test sample. Before the aperture and cantilever could be dry etched, the remaining areas of the substrate also had to be protected from the dry etch. Thus another level of lithography was carried out to protect these areas. This was carried out as follows:

(a) 15% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(b) The resist was dried by baking it overnight at 180°C.

(c) The resist was exposed using a resolution of 200nm and a spot of 400nm at 50kV, with global alignment, and without defocus. The pattern consisted of exposing large rectangular strips placed over the rows of

cantilevers (as the nichrome covering the cantilevers would mask it from the etch) with a dose of 800μC/cm^2.

(d) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(e) The sample was dry etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 10 minutes in the Plasma Technology BP80 (only 200nm of PECVD silicon nitride was being used to define the cantilever at this time).

[Note the fact that the pyramids would not be coated with resist was an advantage here.]

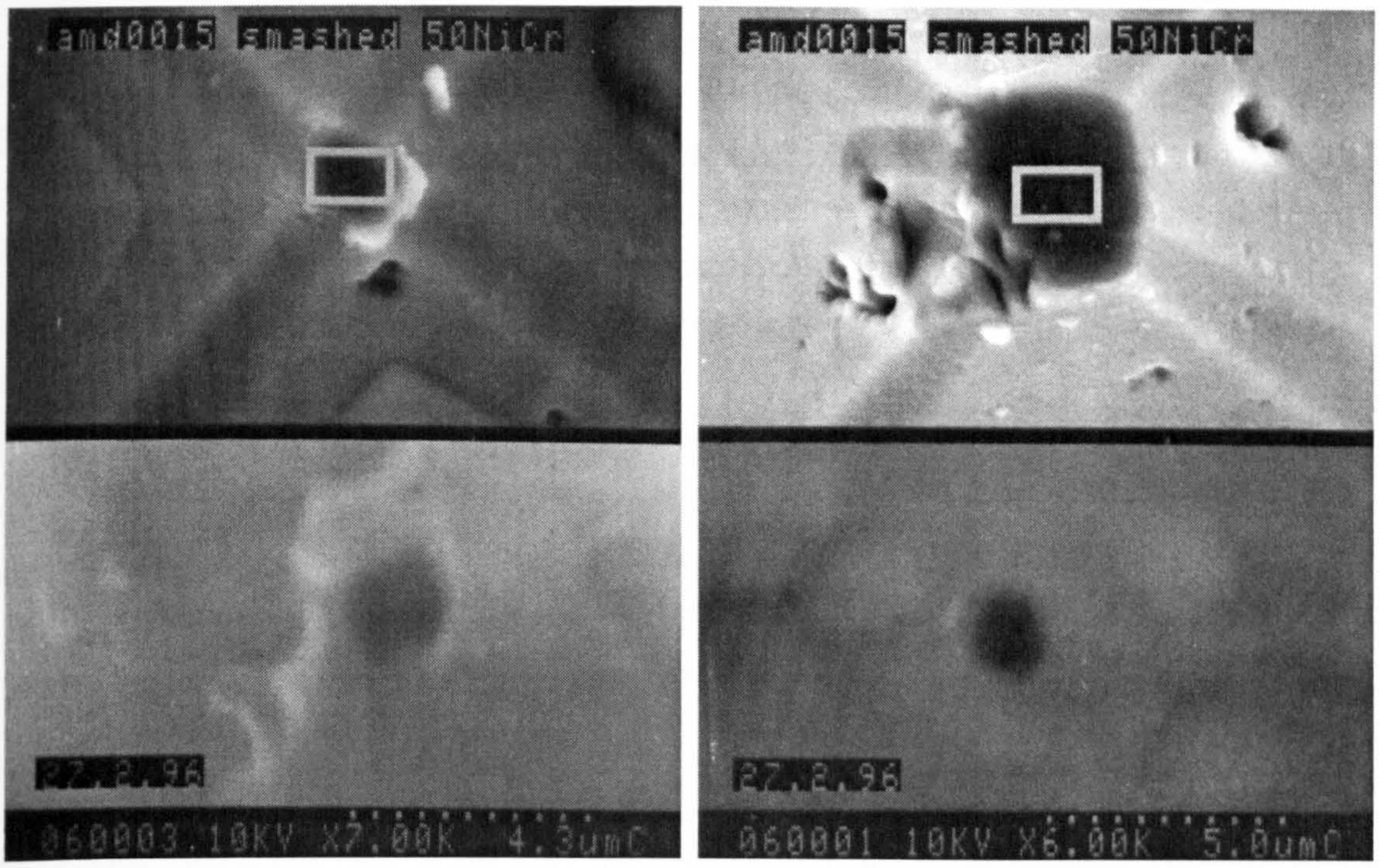


Figure 6.3.2: Micrographs showing early attempts at defining apertures on pyramidal tips ('amd0015'). Two apertures defined in 200nm thick PECVD silicon nitride on pyramidal tips using the cross-linked PMMA scheme. Both apertures appear to be less than 200nm in diameter. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

Unfortunately, at this time severe problems were being experienced with the substrates cracking [see §4.2.3 and §4.3.2.1], particularly through the cantilever membrane areas and after the pyramidal tips had been defined. After dry etching, it was found that the sample had fallen apart, with at most one half remaining as the largest piece. The other fragments were then examined in the Hitachi S800 SEM (the pieces were too big for the higher resolution Hitachi S900). Figure 6.3.2 shows two of the apertures which appear to be between 150 and 175nm in diameter. This would have been perfectly acceptable for the 200nm palladium shielding scheme as discussed in §6.2.3.

The final lithography level before the release etch was then carried out on the largest piece of this sample. Such a level was needed because cantilever release tests [§4.4.2] had shown that if the cantilever was covered in a layer of palladium 200nm thick, the stress within the metal layer would cause the cantilevers to bend upwards even further

(at this point, low stress silicon nitride deposition was not available). Thus the region of metallisation had to be limited to the area surrounding the pyramidal tip by an additional lift-off level. [As noted in §6.2.1, this problem would be less severe with an aluminium film because the 100nm layer used seems to have much lower residual stress.] The process used was as follows:

(a) The nichrome masking layer was stripped off using chrome etch [§3.7.3].
(b) 12% BDH was spin coated at 2500 rpm for 60 seconds.
(c) The resist was dried by baking at 180°C for one hour.
(d) 4% Elvacite was spin coated at 2500 rpm for 60 seconds.
(e) The resist was dried by baking it overnight at 180°C.
(f) The resist was exposed using a resolution of 200nm and a spot size of 400nm at 50kV, with global alignment, and without defocus. The pattern consisted of exposing the final 100μm of the cantilevers (part of the flat cantilever area had to act as a mirror for the interferometric detection of the cantilever deflection required for contact mode AFM – see §1.5) with a dose of 800μC/cm^2.
(g) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.
(h) 10nm of nichrome and 150nm of palladium were evaporated onto the resist using the Plassys MEB 450.
(i) Lift-off was carried out using acetone.

Unfortunately, the whole sample fell apart at this stage and nothing further could be done with it.

The next attempt was carried out on samples produced from a succeeding mask set ('amm0008' and 'amm0009' – see Appendix F). In this mask set, measures had been taken to reduce the likelihood of the sample developing cracks by the introduction of plateaux on the cantilever membranes (which were written during the tip definition level – see §4.3.2.1). However, at this time there was an awareness of the problem with stress in the PECVD silicon nitride layer although no solution had yet been found, and the problem with the tips being misplaced with respect to the photolithography levels had not been discovered [Appendix F]. At this time it was also clear that the thickness of the PECVD silicon nitride layer used for the cantilever would have to be increased from 200nm to 500nm thick in order to obtain a cantilever with an adequate spring constant [see §4.4.2].

The scheme used for these samples was much the same as that used for the previous sample. The sample now consisted of a 8 x 3 array of pairs of devices. The sample was processed as follows:

(a) 500nm of silicon nitride was deposited onto the sample by PECVD.
(b) 8% BDH was spin coated at 2500 rpm for 60 seconds.
(c) The resist was dried by baking at 180°C for one hour.
(d) 2.5% BDH was float coated on top of the previous layer [§5.2].

(e) The resist was dried by baking at 80°C for 30 minutes.
(f) 1.5% Elvacite was float coated on top of the previous bi-layer [§5.2].
(g) The resist was dried by baking at 80°C for 30 minutes.
(h) The resist was exposed using a resolution of 5nm and a spot of 12nm at 100kV, with local alignment to each of an array of pyramidal tips, and with 20μm of defocus. The pattern consisted of (negatively) exposing a 'single overlap' of a 50nm diameter circle on each tip (this was prior to most of the characterisation tests discussed in §6.2.3) with a dose of 400,000μC/cm^2.
(i) The resist was also exposed using a resolution of 200nm and a spot size of 300nm at 100kV, with global alignment, and without defocus. The pattern consisted of (positively) exposing the cantilever shape required (the cantilever pattern had been changed to a later design [§4.4.2]) with a dose of 1000μC/cm^2.
(j) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.
(k) 50nm of nichrome was evaporated onto the resist using the Plassys MEB 450.
(l) Nanostrip lift-off using two applications was carried out.

The next level, which protected the remaining areas of the substrate from the cantilever definition dry etch, was then carried out as follows:

(a) 15% BDH was spin coated at 2500 rpm for 60 seconds.
(b) The resist was dried by baking it overnight at 180°C.
(c) The resist was exposed using a resolution of 200nm and a spot size of 400nm at 50kV, with global alignment, and without defocus. The pattern consisted of exposing large rectangular strips placed over the rows of cantilevers (as the nichrome covering the cantilevers would mask it from the etch) with a dose of 600μC/cm^2.
(d) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.
(e) The sample was dry etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 20 minutes in the Plasma Technology BP80.

Then the pattern defining the region of the SNOM shielding layer (by lift-off) had to be exposed. This was done as follows:

(a) The nichrome masking layer was stripped off using chrome etch [§3.7.3].
(b) 15% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].
(c) The resist was dried by baking at 180°C for one hour.
(d) 2.5% Elvacite was spin coated at 2500 rpm for 60 seconds.
(e) The resist was dried by baking it overnight at 180°C.
(f) The resist was exposed using a resolution of 200nm and a spot size of 300nm at 50kV, with global alignment, and without defocus. The pattern

consisted of exposing the final 100μm of the cantilevers (as before) with a dose of 350μC/cm^2.

(g) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(h) 10nm of nichrome and 150nm of palladium were evaporated onto the resist using the Plassys MEB 450.

(i) Lift-off was carried out using acetone.

The sample was then ready for release etching. However, by this time, it had been discovered that the resist layer being used as protection against the cantilever definition etch was not working optimally. The substrate roughness (upon which 500nm of PECVD silicon nitride had been deposited) manifested itself in the form of myriads of small pyramids. [The cause of this roughness was discussed in Appendix J (§J.2).] The spin coated resist layer was not protecting the apices of these pyramids from the cantilever definition dry etch, and so this step was resulting in large numbers of pinholes being formed in the silicon nitride. These holes would then be undercut by the silicon wet etch (release) step.

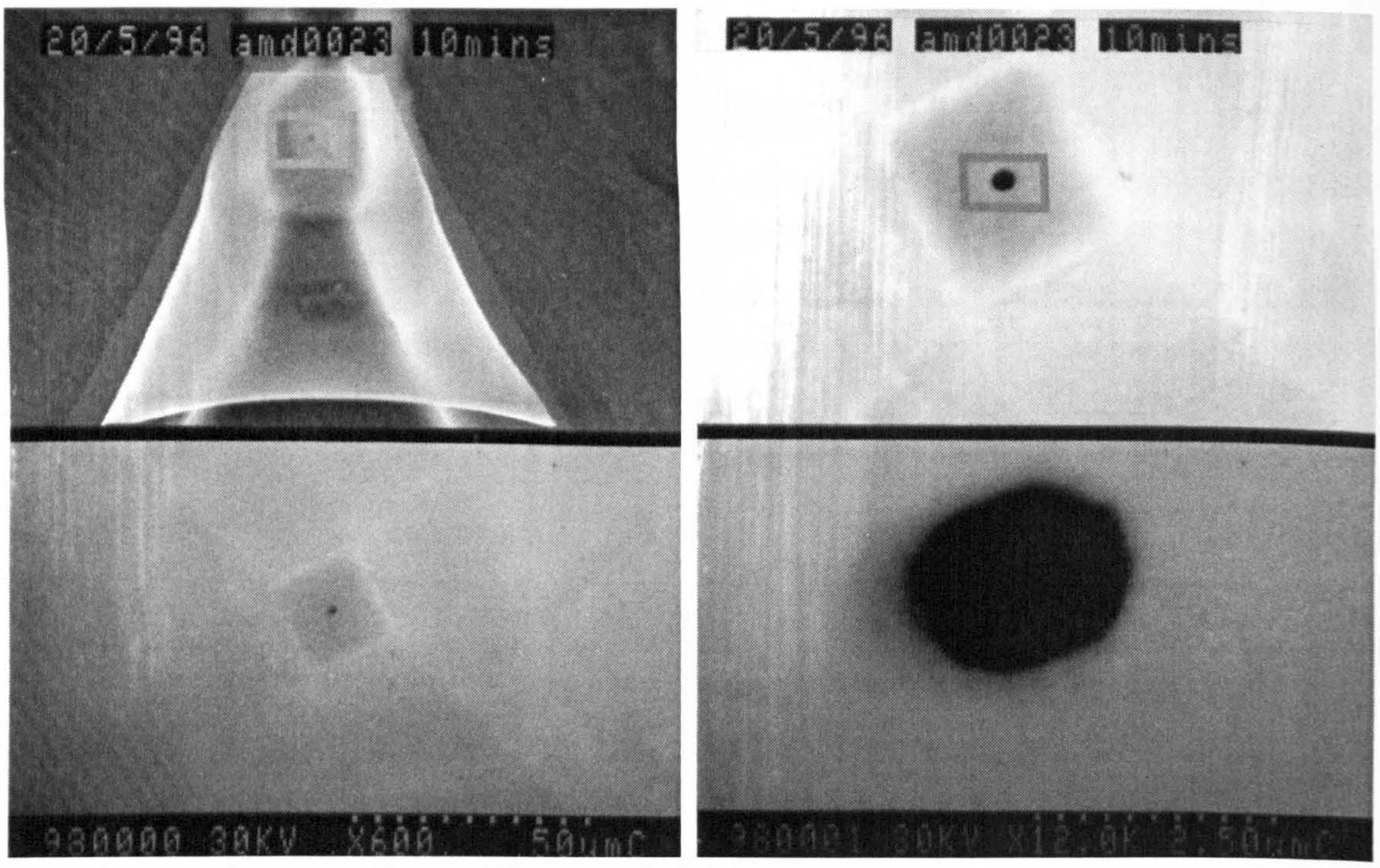


Figure 6.3.3: Micrographs showing attempts to define apertures on functional AFM device substrates ('amd0023'). An aperture defined in 150nm thick palladium on 500nm thick PECVD silicon nitride cantilevers using the cross-linked PMMA scheme. The sample has been partially released. The aperture appears to be around 225nm in diameter, but the metallic shielding appears to have lost its adhesion to the cantilever. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

It was therefore decided that an attempt be made to determine the behaviour of the SNOM shielding layer in the release etch by reducing the etch time to 10 minutes in 7 molar (+5%) potassium hydroxide solution at 65°C. [Note that prior to the release etch, no deoxidation in hydrofluoric acid takes place as this would attack the PECVD silicon nitride cantilever.] Figures 6.3.3 and 6.3.4 show the results of this. From Figure 6.3.3 it

is clear that an aperture has been formed with a diameter of around 225nm. However, by examining Figure 6.3.4 which is taken at an angle, it seems as though the aperture consists of a central core surrounded by a more translucent border. This could simply be an artefact (the aperture acting as a faraday cup), or it could be contamination from the SEM or alternatively it could indicate that the metal profile of the aperture is tapered (as desired). Another observation was that the 50nm thick nichrome layer used to protect and define the cantilever during the cantilever definition dry etch was not adequately protecting the rougher areas of the surface (not conformally deposited) resulting in the introduction of many pinholes in the cantilever. Also noted was the fact that the SNOM shielding layer appeared to have lost its adhesion to the silicon nitride of the cantilever. This was probably due to residual stress in the metal layer causing the edges of the shield to curl up off the cantilever surface.

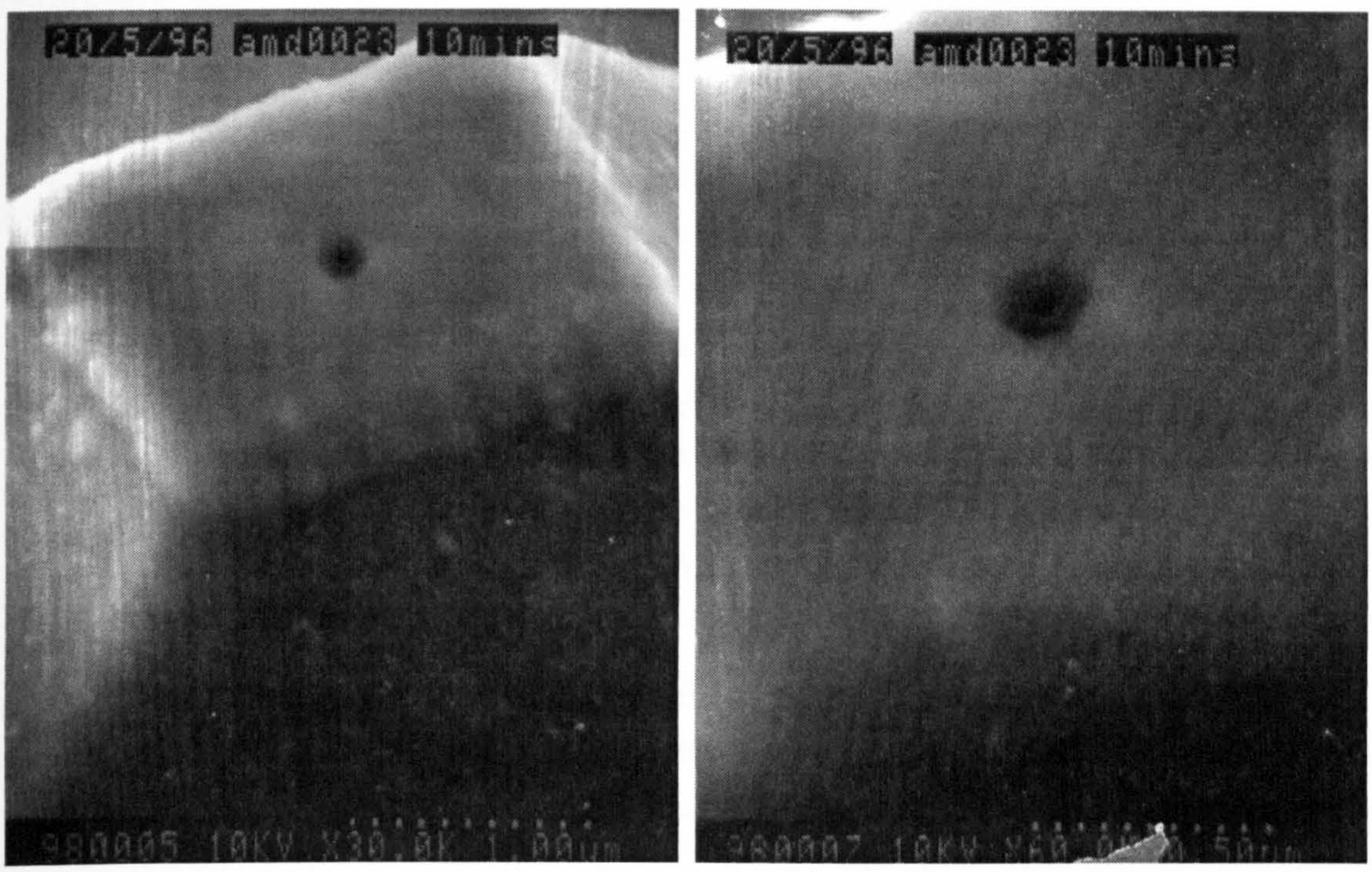


Figure 6.3.4: Micrographs showing attempts to define apertures on functional AFM device substrates ('amd0023'). An aperture defined in 150nm thick palladium on 500nm thick PECVD silicon nitride cantilevers using the cross-linked PMMA scheme. Tilted view of the aperture of Figure 6.3.3. A translucent border can be seen around the innermost core of the aperture.

The next attempt was made after yet another mask had been produced ('amm0011' – see Appendix F). At this point, low stress PECVD silicon nitride had become available, and was now being used exclusively. To tackle some of the other problems of the previous sample, it was decided that the mask for the cantilever definition dry etch be made entirely of nichrome. The whole substrate could be coated with nichrome after deposition of the PECVD silicon nitride layer and the areas to be dry etched could be defined by wet etching the nichrome with chrome etch. However, care had to be taken in using this scheme as if resist was simply spin coated then the apices of spurious pyramids on the substrate would not be coated and any chrome etching would result in holes in the

nichrome film, which would then be transferred to the silicon nitride and silicon during subsequent processing steps. For the SNOM process it would also be necessary to remove the nichrome at the apex of each tip so that an aperture could be defined at the next level using the cross-linked PMMA lift-off method. For the first attempt the process used was as follows:

(a) 500nm of silicon nitride was deposited onto the sample by PECVD.
(b) 100nm of nichrome was evaporated onto the resist using the Plassys MEB 450.
(c) 8% equivalent thickness Aldrich was spin coated at 2500 rpm for 60 seconds (the switch to Aldrich had to be made at this point in order to develop a future-proof process – see §3.5.3).
(d) The resist was dried by baking at 180°C for one hour.
(e) 4% Elvacite was float coated on top of the previous layer [§5.2].
(f) The resist was dried by baking at 80°C for 30 minutes.
(g) The resist was exposed using a resolution of 200nm and a spot size of 300nm at 50kV, with global alignment, and without defocus. The pattern consisted of (positively) exposing the cantilever shape required with a dose of 350μC/cm^2.
(h) The resist was also exposed using a resolution of 50nm and a spot size of 112nm at 50kV, with local alignment to each of an array of pyramidal tips, and with 20μm of defocus. The pattern consisted of (positively) exposing a 5μm square ('window') at the apex of each tip with a dose of 350μC/cm^2.
(i) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.
(j) The nichrome was then wet etched in chrome etch [§3.7.3] for 5 minutes.

The sample was then examined with the optical microscope. It appeared that the 4% float coated Elvacite layer had been too thick so that many of the large areas had not cleared out during development.

This procedure was then repeated, but with the float coated layer consisting of 4% equivalent thickness Aldrich instead of 4% Elvacite. This proved successful. Without removing the resist after the chrome etch, the processing continued as follows:

(a) 2.5% equivalent thickness Aldrich was float coated on top of the previous layer [§5.2].
(b) The resist was dried by baking at 80°C for 30 minutes.
(c) 1.5% Elvacite was float coated on top of the previous layer [§5.2].
(d) The resist was dried by baking at 80°C for 30 minutes.
(e) The resist was exposed using a resolution of 200nm and a spot size of 300nm at 100kV, with local alignment to each of an array of pyramidal tips, and with 20μm of defocus. The pattern consisted of (positively) exposing a 15μm square cap over each tip (to cover the 5μm square 'window') with a dose of 1000μC/cm^2.

(f) The resist was also exposed using a resolution of 5nm and a spot size of 12nm at 100kV, with local alignment to each of an array of pyramidal tips, and with 20μm of defocus. The pattern consisted of (negatively) exposing a 'single overlap' of a 50nm diameter circle on each tip with a dose of 400,000μC/cm^2.

(g) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(h) 50nm of nichrome was evaporated onto the resist using the Plassys MEB 450.

(i) Nanostrip lift-off using two applications was carried out.

(j) The sample was dry etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 30 minutes in the Plasma Technology BP80.

(k) The nichrome masking layer was stripped off using chrome etch [§3.7.3].

(l) 15% BDH was spin coated at 2500 rpm for 60 seconds.

(m) The resist was dried by baking at 180°C for one hour.

(n) 2.5% Elvacite was spin coated at 2500 rpm for 60 seconds.

(o) The resist was dried by baking it overnight at 180°C.

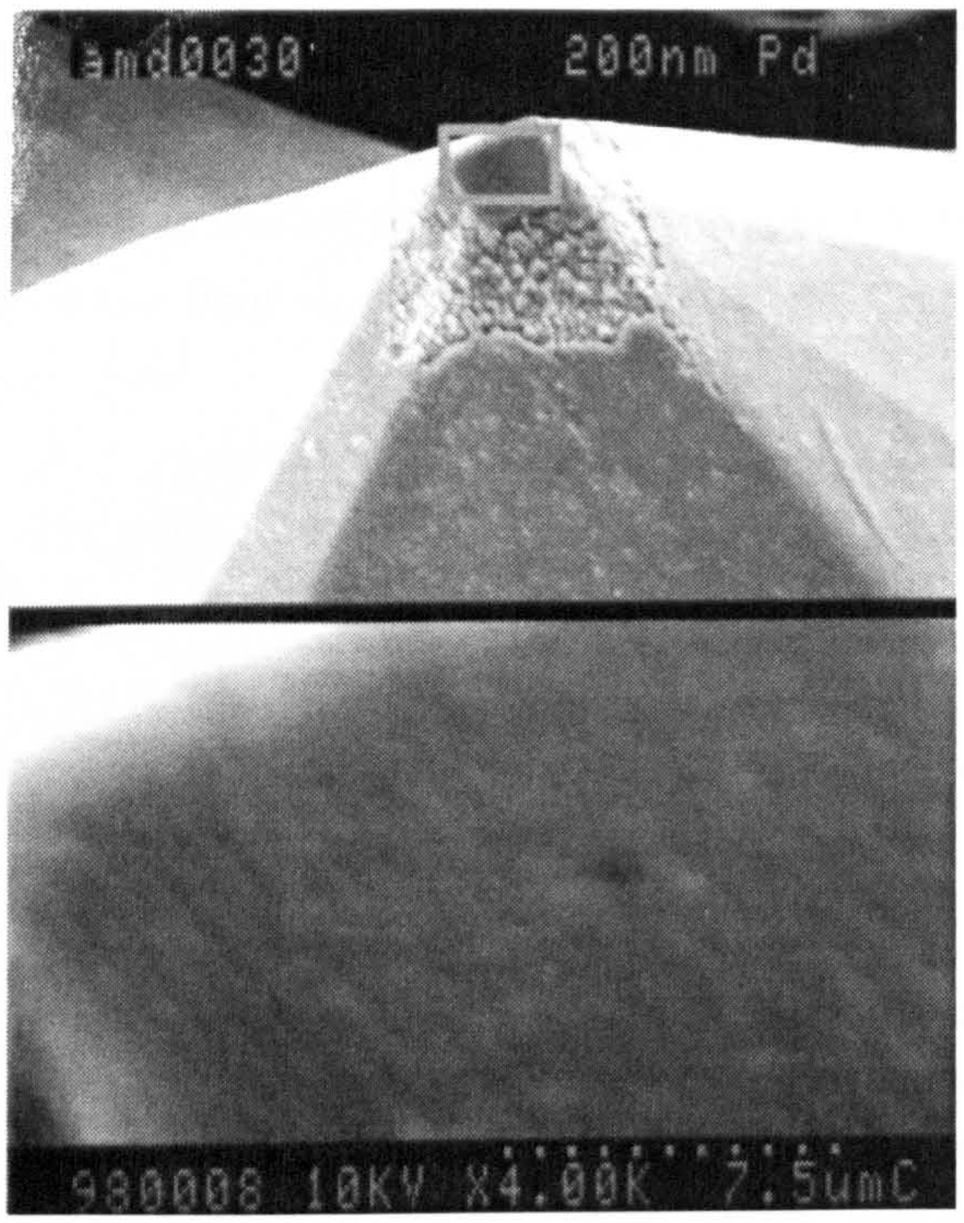
(p) The resist was exposed using a resolution of 200nm and a spot size of 300nm at 50kV, with global alignment, and without defocus. The pattern consisted of exposing the final 100μm of the cantilevers (as before) with a dose of 500μC/cm^2.

(q) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(r) 10nm of nichrome and 200nm of palladium were evaporated onto the resist using the Plassys MEB 450.

(s) Lift-off was carried out using acetone.

This sample was then examined in the Hitachi S800 SEM before any further steps were carried out. As shown in Figure 6.3.5a, an aperture of around 75 to 100nm in diameter appears to have been defined. The sample was then release etched for 75 minutes in 7 molar (+5%) potassium hydroxide solution at 65°C. The release etching proved successful and the SNOM shield did not lose its adhesion to the cantilever this time. Figure 6.3.5b shows both secondary and backscatter images of what appears to be a sub-100nm diameter aperture. However, although at first it seemed as if success had finally been achieved, an examination with the Jeol 100 CXII STEM in transmission electron mode seemed to show that the aperture had not cleared out. It was possible that the hole in the silicon nitride had been so small that deposition of the shielding layer had closed up the aperture and what was being seen in the SEM images was a difference in the metal thickness rather than an absence of metal. It is also possible that the metal deposited within the aperture itself, which should have been attached to silicon, had not been removed during the release etch as expected [Figure 6.2.2].



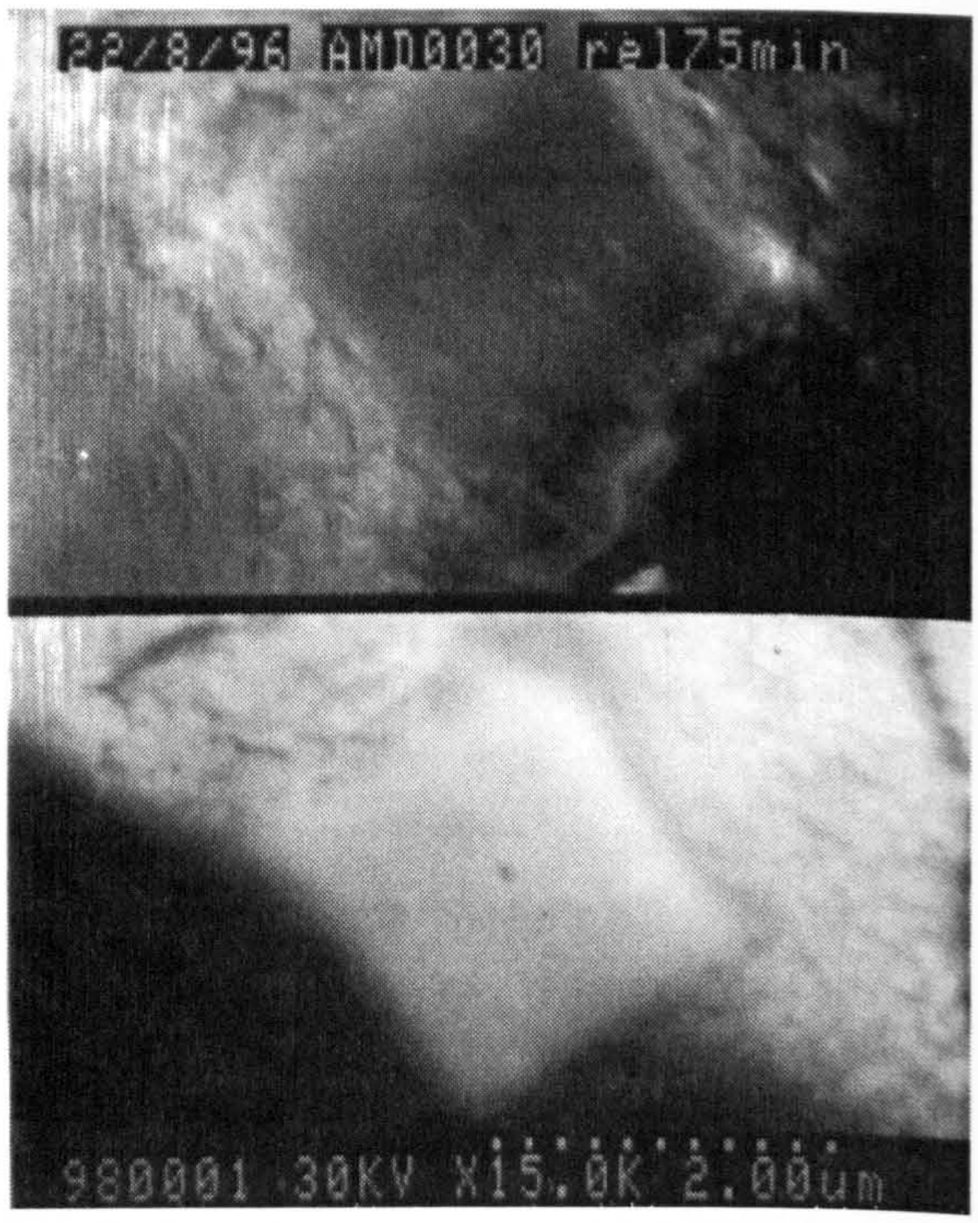


Figure 6.3.5: Micrographs showing attempts to define apertures on functional AFM device substrates ('amd0030'). An aperture defined in 200nm thick palladium on 500nm thick PECVD silicon nitride cantilevers using the cross-linked PMMA scheme. The aperture appears to be between 75nm and 100nm in diameter. (a) Left Image: angled view of the tip and aperture prior to release. The surface of the tip where the nichrome 'window' was defined prior to exposing the aperture appears to be particularly rough. The bottom half is a 10x magnification over the displayed scale; (b) Right Image: a backscatter (top half) and secondary (bottom half) image of the aperture in (a) after release.

Also noticeable was the patchy nature of the metal shielding in the region where the nichrome 'window' had been defined. This was attributed to a residue being left behind by the chrome etch, which affected the subsequent nichrome evaporation and was then transferred to the silicon nitride during the cantilever and aperture definition dry etch. This could be avoided by thoroughly cleaning the substrate (using both inorganic and organic cleans – see Appendix D) prior to the aperture lithography level, although this would mean that the resist used to define the wet etching areas could not be reused as part of the resist layer for the aperture definition level.

An alternative explanation [also see §3.7.3] based on the fact that both nickel and chromium intermix with silicon even at room temperature [Franciosi, Peterman and Weaver[56] (1981); Grunthaner, Grunthaner, Madhuker and Mayer[57] (1981)] is that some atoms of nickel, chromium or both had diffused into the silicon nitride layer and were not all removed by the chrome etch. A sufficient number of these atoms may have remained to affect grain formation during the subsequent deposition of nichrome.

[A peculiar effect had been seen once before when lifted-off nichrome was being used as a masking layer for the cantilever dry etch. On a few occasions where the nichrome masking layer had been subsequently stripped off using chrome etch, a 'ghost' image of the area where the nichrome had made contact with the silicon nitride could be seen with the optical microscope. Although this could again be attributed to a residue left by the

chrome etch, on one occasion the 200nm thick silicon nitride layer beneath had also been stripped off, using hydrofluoric acid, and a new silicon nitride layer deposited by PECVD. However, the silicon nitride deposited in the area where the nichrome had been was a slightly different colour from the rest of the layer. Colour variations are normally attributed to a difference in thickness. As it is known that PECVD silicon nitride is an inferior diffusion barrier compared with stoichiometric LPCVD silicon nitride [Sze[58] (1985)], it is not inconceivable that nickel or chromium atoms could have diffused through the PECVD silicon nitride and reached the silicon surface. Then in the subsequent deposition step, any remaining atoms may have affected the rate at which silicon nitride growth started. However, this is highly speculative and without further tests no firm conclusions can be drawn.]

Unfortunately, at this point time ran out and no further fabrication attempts could be made.

One additional problem which should be mentioned here concerned local alignment to individual devices [see §3.5.2] and was related to the algorithms used by the beamwriter alignment software. It was found that the beamwriter software always assumed that there were no alignment markers present at the extreme right and at the top of the sample, unless the number of repeats was set to 1 x 1 (*i.e.* global alignment). Thus all devices at the top and right edges were aligned by extrapolating alignment correction parameters from other cells. To avoid this problem, and still use local alignment to each (pair of) device(s), the software could be told that the number of repeats exceeded the actual number in each direction by 1, *i.e.* to align to an array of 8 x 3 devices (x, y), the beamwriter would be told that the number of repeats was 9 x 4. Then the matrix would be set up as a complex matrix, so that those virtual devices at the top and right edges of the sample could be 'dropped out' [Thoms[59] (1993-1996)]. Using this scheme successfully avoids the problem, at the expense of some additional job execution time due to the use of a complex matrix rather than a simple one. The BWL software supports all these features, so for the example above the matrix would simply be (if typed directly into BWL, this has to be vertically inverted, or alternatively a text file should be used):

```
- - - - - - - - - -
1 1 1 1 1 1 1 1 -
1 1 1 1 1 1 1 1 -
```

6.4 SNOM PROBE FABRICATION SUMMARY

The results of the tests carried out in section §6.2.3 and §6.3 suggested that the way forward was to use a float coated 3.25% equivalent thickness Aldrich, 2% Elvacite bi-layer with 4 overlaps of drawn circular patterns ranging from 30nm to 100nm in diameter, all exposed at 400,000$\mu C/cm^2$ per overlap.

In addition, the masking layer for the cantilever definition dry etch had been changed from 100nm thick nichrome to 100nm thick chromium, which was deposited by sputtering after silicon nitride deposition. This was in response to the stress and cracking problems being experienced with thicker nichrome layers, as discussed in Appendix L (§L.1).

The final process to define SNOM probes, from the second EBL level onwards, although not demonstrated to completion, is envisaged to be as follows:

(a) Between 250nm and 750nm of stress free PECVD silicon nitride (depending on the required cantilever compliance) is deposited onto the (pyramidal tip side of the) substrate.
(b) A 100nm thick layer of chromium is sputtered onto the substrate [§4.4.2].
(c) 8% equivalent thickness Aldrich is spun onto the substrate at 2500 rpm for 60 seconds [Table 3.5.1].
(d) This resist layer is dried by prebaking at 180°C for at least 6 hours (or overnight).
(d) 4% equivalent thickness Aldrich is float coated onto the substrate [Table 3.5.1].
(e) This resist layer is dried by baking at 80°C for 30 minutes.
(f) Two patterns are exposed at 50kV using one of the global EBL markers, one pattern with a spot size of 400nm and a dose of 350μC/cm^2, the other with a spot size of 112nm and a dose of 350μC/cm^2. The first pattern comprises two layers: the layer (layer 2) containing the cantilever pattern (including the shoulder and tapered region beneath – §4.4.2) is subtracted from a layer (layer 9) containing a pattern consisting of long rectangular strips enclosing the cantilever patterns. The second pattern consists of a 5μm square centred at the apex of each tip ('aperture window').
(g) The pattern is developed in MIBK:IPA 1:1 at 23°C for 35 seconds.
(h) The chromium layer is wet etched using chrome etch [§3.7.3] for 5 to 10 minutes or until it is clear.
(i) The resist is stripped off in acetone and the substrate is cleaned thoroughly with a twenty minute dip in nanostrip diluted with water in the ratio 1:1 (this solution should heat up to 120°C), followed by two five minute dips in ultra-pure water and a five minute dip in IPA before being blow dried and baked for five minutes at 180°C [App. D].

[At this point, if the silicon nitride being used is too thick to allow reliable apertures to be formed, the silicon nitride can be partially thinned by dry etching. However, care should be taken to use a low bias and power dry etch so as to end up with an overcut rather than undercut edge profile: when the shielding metal is deposited over the apices of the tips, an undercut profile could result in a break in the shielding layer over the edge defined by the transition in silicon nitride thickness. Alternatively, the 5μm square

exposed in step (f) could be enlarged so that if the shielding layer has a small break in it, it would be sufficiently far away from the aperture not to pose any imaging problems. Another idea which makes the thinning of silicon nitride, or even complete removal and re-deposition, attractive would be the tuning of the silicon nitride thickness to give maximum reflectivity at the operating wavelength of the probe by thin film interference.]

(j) 12% equivalent thickness Aldrich is spin coated onto the substrate at 2500 rpm for 60 seconds [Table 3.5.1].

(k) The resist layer is prebaked for at least 6 hours (or overnight) at 180°C.

(l) 3.25% equivalent thickness Aldrich is float coated onto the substrate [§5.2].

(m) This resist layer is dried by baking for 30 minutes at 80°C.

(n) 2% Elvacite is float coated onto the substrate [§5.2].

(o) The resist is dried by baking for 30 minutes at 80°C.

(p) Two patterns, one consisting of 15μm squares placed at the apices of each pyramidal tip, and another of circles of a diameter ranging from 30 to 100nm (depending on the diameter of aperture required and thickness of shielding metal chosen) and using a resolution of 5nm placed at the centre of each pyramidal tip, are exposed at 100kV. The first pattern uses one set of global EBL markers for alignment and is exposed with a spot size of 300nm, a dose of 1000μC/cm^2 and defocus of the pyramidal tip height [§5.1]. The second pattern uses one set of local EBL markers for alignment, using a complex matrix with drop outs as explained in §6.3, and is exposed with a spot size of 12nm and a total dose of 1,600,000μC/cm^2 (generated using 4 overlaps and a dose setting of 400,000μC/cm^2).

(q) The pattern is developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(r) 50nm of nichrome (60% nickel/40% chromium) is evaporated onto the resist using the Plassys MEB 450.

(r) Lift-off is performed using 3 ten minute dips in nanostrip diluted with water in the ratio 1:1 (this solution should heat up to 120°C), followed by two five minute dips in ultra-pure water and a five minute dip in IPA before being blow dried [App. D and §6.2.3].

(s) The substrate is dry etched in the Plasma Technology BP80 with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 30 minutes, if 500nm of silicon nitride is used, or longer in proportion to the silicon nitride thickness.

(t) The chromium layer is then stripped off using chrome etch [§3.7.3] for approximately 10 minutes or until clear.

At this point two alternative strategies can be followed as described below.

SNOM shield evaporated prior to release etch:

(a) The substrate is cleaned using an inorganic followed by an organic clean but without any ultrasonic agitation [App. D].
(b) 15% equivalent thickness Aldrich is spin coated onto the substrate at 2500 rpm for 60 seconds [Table 3.5.1].
(c) The resist is dried by baking for one hour at 180°C.
(d) 4% Elvacite is spin coated onto the previous resist layer at 2500 rpm for 60 seconds [Table 3.5.1].
(e) The substrate is prebaked at 180°C for at least six hours (or overnight).
(f) The pattern is exposed at 50kV using one of the global EBL markers with a spot size of 400nm and a dose of 500μC/cm^2. The pattern consists of the non-rectangular region of each of the cantilevers (allowing a 1μm inset from the edges), with all corners rounded off (this was found to reduce the likelihood of the shielding metal curling up during the release etch).
(g) The pattern is developed in MIBK:IPA 1:1 at 23°C for 35 seconds.
(h) The SNOM shielding layer (*e.g.* 10nm nichrome/200nm palladium) is evaporated onto the resist using the Plassys MEB 450, or another evaporator if necessary.
(i) Lift-off is performed using acetone, followed by a rinse in IPA.
(j) The substrate is given two five minute dips in ultra-pure water without ultrasonic agitation.
(k) The substrate is wet (release) etched [App. L (§L.1)] in 7 molar (+5%) potassium hydroxide solution at 65°C until release (using the self-calibrating scheme discussed in §4.3.2.3).
(l) If the shielding layer is resistant to it, the substrate is neutralised in sulphuric acid:water 1:4 for 1 minute.
(m) The substrate is rinsed thoroughly in ultra-pure water followed by IPA, before being baked dry (typically 5 minutes at 180°C).

SNOM shield evaporated after release etch:

(a) The substrate is cleaned using an inorganic clean but without any ultrasonic agitation [App. D].
(b) The substrate is wet (release) etched [App. L (§L.1)] in 7 molar (+5%) potassium hydroxide solution at 65°C until release (using the self-calibrating scheme discussed in §4.3.2.3).
(c) The substrate is neutralised in sulphuric acid:water 1:4 for 1 minute.
(d) The substrate is rinsed thoroughly in ultra-pure water followed by IPA, before being baked dry (typically 5 minutes at 180°C).
(e) The SNOM shielding layer (*e.g.* 10nm nichrome/200nm palladium) is evaporated onto the resist using the Plassys MEB 450, or another evaporator if necessary.

Note that with the latter scheme, it is not possible to limit the coverage of the SNOM shielding layer to the end of the cantilever (this would require an additional electron-beam lithography level and lift-off on fully released cantilevers, which is not impossible, but unjustifiably risky in the opinion of the author). This means that if the layer is thick enough, or contains enough residual stress, it may result in the bending of the released cantilever. However, as noted previously, the degree of bending is a function of the metal, the metal layer thickness, the silicon nitride thickness, the residual stress in the silicon nitride layer and the deposition conditions of the metal layer. This means that it may be possible to compensate for the residual stress in the metal layer by choosing the residual stress of the silicon nitride layer deposited. However, the main problem here is that after release and prior to deposition the cantilevers will be bent so that evaporation of the metal layer may end up blocking the aperture (as the undercut profile in the silicon nitride is with respect to the plane of the wafer being perpendicular to the evaporation direction).

Scaling of the above process to whole wafers would be possible if float coating onto whole wafers is developed, probably along the lines of what was discussed previously in §5.2. However, any remaining aspects of process development were now left in the capable hands of Dr. H. Zhou and Dr. J. M. R. Weaver.

Following this stage, the devices would be ready for cleaving and use.

6.5 SIMPLE MODELLING OF LIGHT THROUGHPUT

As noted in §1.4.1, modelling of SNOM probes is not trivial. Here only one aspect, the light throughput, will be considered. A proper and more correct calculation is beyond the scope of this thesis. However, for the purposes of engineering design, a more simplistic and crude model which allows a quick way to determine the effect of any design decisions on the light throughput, in a qualitative, if not quantitative way is useful. It has been found that an analytical solution which indicates how strongly the light throughput varies with various design parameters is possible, provided some assumptions are made:

(1) The waveguide has a perfectly cylindrical cross-section throughout.

(2) The walls of the waveguide are made of an ideal metal, *i.e.* infinite conductivity at all frequencies and therefore there can be no losses in, and no penetration of the electric field into, the walls of the waveguide. The attenuation is therefore principally dominated by the degree by which the wavelength is cut-off.

(3) No mode transitions and back reflections occur. Each mode can be modelled independently of the others.

(4) The radius of the waveguide after the point at which the wavelength has become cut-off can be approximated as linearly decreasing towards the final output diameter.

The z-direction is defined to be as shown in Figure 6.1.1. Then from Equations 1.4.1 and 6.1.7, the intensity for wavelengths longer than the cut-off wavelength is given by:

$$I = I_o \exp\left[-j\frac{4\pi z}{\lambda_g}\right] = I_o \exp\left[-\frac{4\pi z}{\lambda_g'}\right], \tag{6.5.1}$$

where I_o is the intensity at $z = 0$ given by:

$$I_o = \mathrm{Re}\left(\mathbf{E}_o \times \mathbf{H}_o^*\right) \tag{6.5.2}$$

and, using Equation 1.4.2:

$$\frac{1}{\lambda_g'} = \frac{j}{\lambda_g} = \sqrt{-\frac{1}{\lambda_g^{\,2}}} = \sqrt{\frac{1}{\lambda_c^{\,2}} - \frac{n^2}{\lambda_o^{\,2}}} = \frac{n}{\lambda_o}\sqrt{\frac{\lambda_o^{\,2}}{n^2\lambda_c^{\,2}} - 1}, \text{ for } \frac{\lambda_o}{n} > \lambda_c, \tag{6.5.3}$$

where λ_g' is the imaginary part of the group wavelength, λ_g, in the waveguide, λ_o is the wavelength in vacuum of the incident radiation, λ_c is the cut-off wavelength of the waveguide and n is the refractive index of the waveguide medium. All other terms are as defined in Equations 1.4.1, 1.4.2 and 6.1.7. Note that the waveguide medium is assumed to be non-conducting.

For a waveguide tapering in the z-direction, with the radius at position, z, given by $R(z)$, the cut-off wavelength, which is also now a function of z, is given by:

$$\lambda_c^{\,p,q}(z) = \frac{2\pi R(z)}{\gamma_{p,q}},\ \gamma_{p,q} = \begin{cases} \eta_{p,q}, \mathrm{TE_{p,q}\ mode} \\ \eta'_{p,q}, \mathrm{TM_{p,q}\ mode} \end{cases} \tag{6.5.4}$$

using Equation 1.4.3 and the coefficients as given in Table 1.4.1. By substituting Equation 6.5.4 into Equation 6.5.3 the (intensity) decay coefficient, $\alpha_{p,q}(z)$, can be defined by:

$$\alpha_{p,q}(z) = \frac{4\pi n}{\lambda_o}\sqrt{\left(\frac{\gamma_{p,q}\lambda_o}{2\pi n R(z)}\right)^2 - 1}. \tag{6.5.5}$$

Since the decay coefficient is a function of position, the equation determining how the intensity varies with position can be determined from the basic definition of decay constant, *i.e.*:

$$\alpha_{p,q}(z) = -\frac{1}{I(z)}\frac{dI}{dz}\bigg|_z \Rightarrow \int_{I(z_o)}^{I(z)} \frac{dI}{I} = -\int_{z_0}^{z} \alpha_{p,q}(z)dz, \tag{6.5.6}$$

where z_o is an arbitrary reference position where the waveguide is cut-off. Integrating both sides then gives:

$$\ln\frac{I_{p,q}(z)}{I_{p,q}(z_o)} = -\int_{z_0}^{z} \alpha_{p,q}(z)dz \Rightarrow I_{p,q}(z) = I_{p,q}(z_o)\exp\left[-\int_{z_0}^{z} \alpha_{p,q}(z)dz\right]. \tag{6.5.7}$$

This equation assumes that the metallic walls of the waveguide are perfect and does not take account of back reflections or mode transitions. However, it can be used to deduce the general behaviour of the light output in terms of various parameters.

It will now be assumed that the radius of the cut-off region of the waveguide tapers linearly with z, *i.e.*:

$$R(z) = R_c - mz,\ R_c = \frac{\gamma_{p,q}\lambda_o}{2\pi n},\ m = \tan\left(\frac{\theta_c}{2}\right),\ 0 \le z \le \frac{R_c}{m} = R_c \cot\left(\frac{\theta_c}{2}\right), \tag{6.5.8}$$

where R_c is the cut-off radius and θ_c is the full cone angle made by the tapering cylindrical waveguide profile (defined as twice the semi-angle subtended by the cone). [This is simply the equation of a line with gradient, $-m$, and intercept, R_c.] If the radius of the aperture tapers from R_a at $z = z_a$ to R_b at $z = z_b$ for a metallic shield of thickness, $t_m = z_b - z_a$, then:

$$R_a = R(z_a) = R_c - mz_a,\ R_b = R(z_b) = R_c - mz_b \Rightarrow R_a = R_b + mt_m = R_b + t_m \tan\left(\frac{\theta_c}{2}\right), \tag{6.5.9}$$

where both radii must be smaller than the cut-off radius.

Defining the logarithmic attenuation function, $\psi_{p,q}(z_1, z_2)$, by:

$$\psi_{p,q}(z_1, z_2) = \int_{z_1}^{z_2} \alpha_{p,q}(z)dz = \frac{4\pi n}{\lambda_o}\int_{z_1}^{z_2} \sqrt{\left(\frac{\gamma_{p,q}\lambda_o}{2\pi n R(z)}\right)^2 - 1}\, dz, \tag{6.5.10}$$

means that the ratio of light intensity at z_b to that at z_a is given by:

$$\frac{I_b^{\,p,q}}{I_a^{\,p,q}} = \frac{I_{p,q}(z_b)}{I_{p,q}(z_a)} = \exp\left[-\psi_{p,q}(z_a, z_b)\right]. \tag{6.5.11}$$

Then substituting:

$$x = \frac{2\pi n R(z)}{\gamma_{p,q}\lambda_o} \Rightarrow x_a = \frac{2\pi n R_a}{\gamma_{p,q}\lambda_o},\ x_b = \frac{2\pi n R_b}{\gamma_{p,q}\lambda_o}, \tag{6.5.12}$$

so that:

$$\frac{dx}{dz} = \frac{2\pi n}{\gamma_{p,q}\lambda_o}\frac{dR}{dz} = -m\frac{2\pi n}{\gamma_{p,q}\lambda_o} \Rightarrow dz = -\frac{\gamma_{p,q}\lambda_o}{2\pi m n}dx, \tag{6.5.13}$$

gives:

$$\psi_{p,q}(x_a, x_b) = -\frac{2\gamma_{p,q}}{m}\int_{x_a}^{x_b} \sqrt{\frac{1}{x^2} - 1}\, dx = -\frac{2\gamma_{p,q}}{m}\int_{x_a}^{x_b} \frac{\sqrt{1 - x^2}}{x}\, dx. \tag{6.5.14}$$

Notice that x is simply the ratio of the radius at any point to the cut-off radius, R_c.

Now substituting:

$$x = \cos u \Rightarrow u_a = \cos^{-1}\left[\frac{2\pi n R_a}{\gamma_{p,q}\lambda_o}\right],\ u_b = \cos^{-1}\left[\frac{2\pi n R_b}{\gamma_{p,q}\lambda_o}\right], \tag{6.5.15}$$

so that:

$$\frac{dx}{du} = -\sin u \Rightarrow dx = -\sin u\, du, \tag{6.5.16}$$

gives:

$$\psi_{p,q}(u_a, u_b) = \frac{2\gamma_{p,q}}{m}\int_{u_a}^{u_b} \frac{\sin^2 u}{\cos u}\, du = \frac{2\gamma_{p,q}}{m}\int_{u_a}^{u_b} \frac{1 - \cos^2 u}{\cos u}\, du = \frac{2\gamma_{p,q}}{m}\int_{u_a}^{u_b} \frac{1}{\cos u} - \cos u\, du. \tag{6.5.17}$$

Equation 6.5.17 can be integrated directly to give:

$$\psi_{p,q}(u_a,u_b) = \frac{2\gamma_{p,q}}{m}\left[\ln\left|\sec u + \tan u\right| - \sin u\right]_{u_a}^{u_b}$$

$$= \frac{2\gamma_{p,q}}{m}\left[\ln\left|\frac{\sec u_b + \tan u_b}{\sec u_a + \tan u_a}\right| + \sin u_a - \sin u_b\right]$$

$$= \frac{2\gamma_{p,q}}{m}\left[\ln\left|\frac{\cos u_a}{\cos u_b}\bullet\frac{1+\sin u_b}{1+\sin u_a}\right| + \sin u_a - \sin u_b\right] \qquad (6.5.18)$$

$$= \frac{2\gamma_{p,q}}{m}\left[\ln\left|\frac{\cos u_a}{\cos u_b}\bullet\frac{1+\sqrt{1-\cos^2 u_b}}{1+\sqrt{1-\cos^2 u_a}}\right| + \sqrt{1-\cos^2 u_a} - \sqrt{1-\cos^2 u_b}\right]$$

Substituting Equation 6.5.15 back in to this then gives the logarithmic attenuation function in terms of the normalised radii (to the cut-off radius), x_a and x_b, at the input end and output end of the waveguide:

$$\psi_{p,q}(x_a,x_b) = \frac{2\gamma_{p,q}}{m}\left[\ln\left|\frac{x_a}{x_b}\bullet\frac{1+\sqrt{1-x_b^{\,2}}}{1+\sqrt{1-x_a^{\,2}}}\right| + \sqrt{1-x_a^{\,2}} - \sqrt{1-x_b^{\,2}}\right]. \qquad (6.5.19)$$

Upon substituting Equation 6.5.12 back in, this becomes:

$$\psi_{p,q}(R_a,R_b) = \frac{2\gamma_{p,q}}{\tan\left(\frac{\theta_c}{2}\right)}\ln\frac{R_a\left[1+\sqrt{1-\left(\frac{2\pi nR_b}{\gamma_{p,q}\lambda_o}\right)^2}\right]}{R_b\left[1+\sqrt{1-\left(\frac{2\pi nR_a}{\gamma_{p,q}\lambda_o}\right)^2}\right]} \qquad (6.5.20)$$

$$+\frac{2\gamma_{p,q}}{\tan\left(\frac{\theta_c}{2}\right)}\left[\sqrt{1-\left(\frac{2\pi nR_a}{\gamma_{p,q}\lambda_o}\right)^2} - \sqrt{1-\left(\frac{2\pi nR_b}{\gamma_{p,q}\lambda_o}\right)^2}\right]$$

It is more convenient to express this in terms of the input and output aperture diameters, $D_a = 2R_a$ and $D_b = 2R_b$:

$$\psi_{p,q}(D_a,D_b) = \frac{2\gamma_{p,q}}{\tan\left(\frac{\theta_c}{2}\right)}\ln\frac{D_a\left[1+\sqrt{1-\left(\frac{n\pi D_b}{\gamma_{p,q}\lambda_o}\right)^2}\right]}{D_b\left[1+\sqrt{1-\left(\frac{n\pi D_a}{\gamma_{p,q}\lambda_o}\right)^2}\right]} \qquad (6.5.21)$$

$$+\frac{2\gamma_{p,q}}{\tan\left(\frac{\theta_c}{2}\right)}\left[\sqrt{1-\left(\frac{n\pi D_a}{\gamma_{p,q}\lambda_o}\right)^2} - \sqrt{1-\left(\frac{n\pi D_b}{\gamma_{p,q}\lambda_o}\right)^2}\right]$$

Then from Equation 6.5.21 and Equation 6.5.11, the ratio of intensity at the output to that at the input is given by:

$$\frac{I_b^{p,q}}{I_a^{p,q}} = \left[\frac{\exp\left[\frac{2\gamma_{p,q}}{\tan\left(\frac{\theta_c}{2}\right)}\sqrt{1-\left(\frac{n\pi D_b}{\gamma_{p,q}\lambda_o}\right)^2}\right]}{\exp\left[\frac{2\gamma_{p,q}}{\tan\left(\frac{\theta_c}{2}\right)}\sqrt{1-\left(\frac{n\pi D_a}{\gamma_{p,q}\lambda_o}\right)^2}\right]}\right]\left[\frac{D_b\left[1+\sqrt{1-\left(\frac{n\pi D_a}{\gamma_{p,q}\lambda_o}\right)^2}\right]}{D_a\left[1+\sqrt{1-\left(\frac{n\pi D_b}{\gamma_{p,q}\lambda_o}\right)^2}\right]}\right]^{\frac{2\gamma_{p,q}}{\tan\left(\frac{\theta_c}{2}\right)}}, \qquad (6.5.22)$$

or explicitly in terms of the metal thickness:

$$\frac{I_b^{p,q}}{I_a^{p,q}} = \left[\frac{\exp\left[\frac{4t_m\gamma_{p,q}}{(D_a-D_b)}\sqrt{1-\left(\frac{n\pi D_b}{\gamma_{p,q}\lambda_o}\right)^2}\right]}{\exp\left[\frac{4t_m\gamma_{p,q}}{(D_a-D_b)}\sqrt{1-\left(\frac{n\pi D_a}{\gamma_{p,q}\lambda_o}\right)^2}\right]}\right]\left[\frac{D_b\left[1+\sqrt{1-\left(\frac{n\pi D_a}{\gamma_{p,q}\lambda_o}\right)^2}\right]}{D_a\left[1+\sqrt{1-\left(\frac{n\pi D_b}{\gamma_{p,q}\lambda_o}\right)^2}\right]}\right]^{\frac{4t_m\gamma_{p,q}}{(D_a-D_b)}}. \qquad (6.5.23)$$

As an analytical solution, Equations 6.5.22 and 6.5.23 have the usual advantages of being readily calculable with different sets of parameters and allowing graphs of light throughput as a function of one or more parameters to be easily generated, without the need to run lengthy simulations for individual sets of parameters.

However, as can be seen from Equation 6.5.22, the light throughput has a critical dependence on the cone angle of the taper. Therefore it is necessary to be able to estimate this value before any meaningful comparisons with a more conventional SNOM probe can be made. To this end, during development, a test was run to estimate this angle. The starting point was a flat silicon substrate with 500nm of PECVD silicon nitride deposited upon it. The samples were processed as follows:

(a) 4% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].
(b) The resist was dried by baking at 180°C for one hour.
(c) 4% Elvacite was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].
(d) The resist was dried by baking it overnight at 180°C.
(e) The resist was exposed using a spot size of 28nm at 50kV, with no alignment. The pattern consisted of (negatively) exposing [§6.2.3] a variable number of overlaps of 500μm long lines (in a similar manner to those shown in Figures 6.2.37 and 6.2.38) with a dose of 400,000μC/cm^2 per overlap.
(f) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(g) 50nm of nichrome was evaporated onto the resist using the Plassys MEB 450 [§3.6.2].

(h) Nanostrip lift-off using three applications was carried out [§6.2.3].

(i) The sample was dry etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 10 minutes in the Plasma Technology BP80 [§3.6.4].

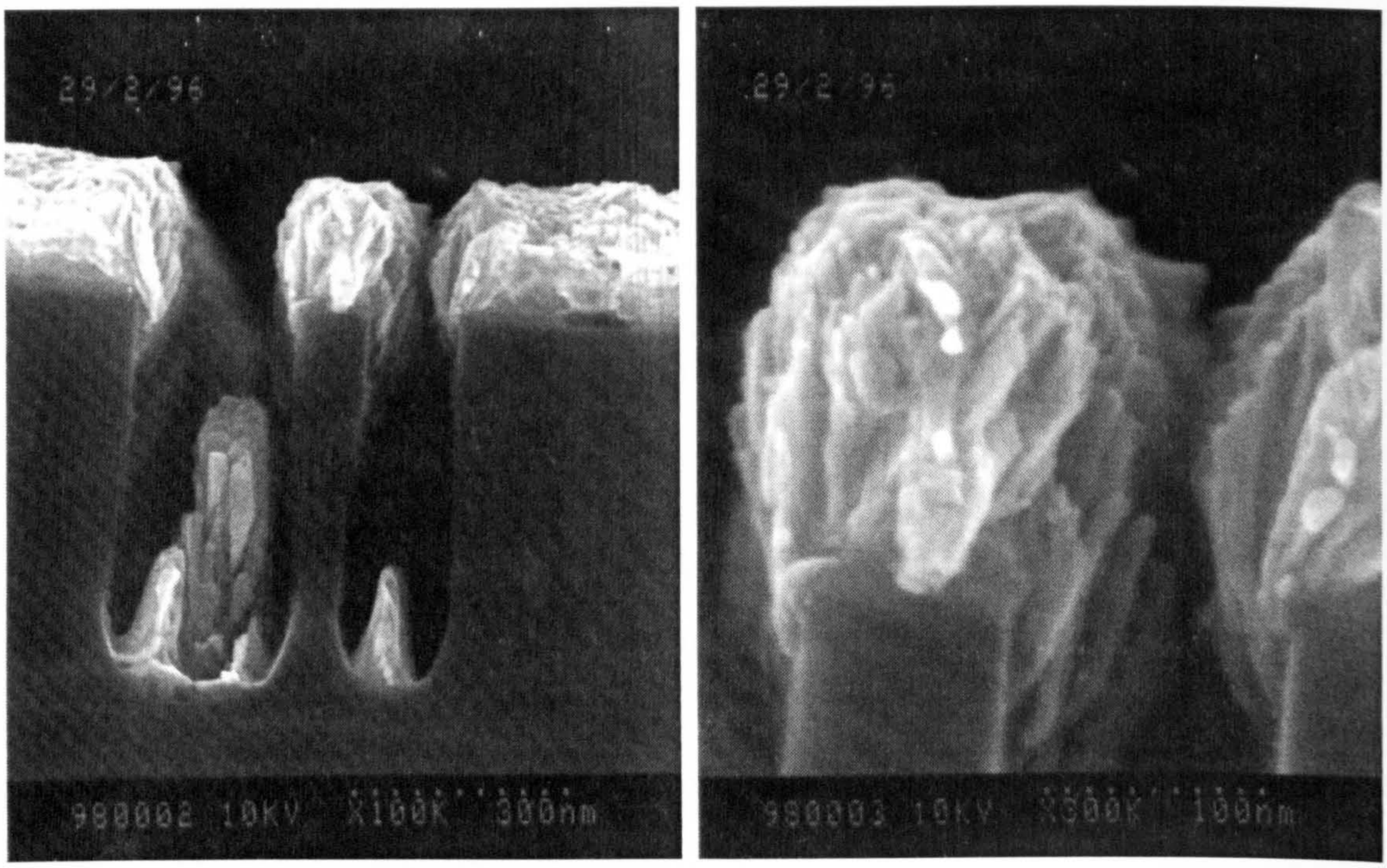


Figure 6.5.1: Micrographs illustrating an attempt to estimate the expected cone angle of a SNOM probe based on 200nm of palladium shielding ('ama0038'). The samples were produced using the cross-linked PMMA process. (a) Left Image: a cleaved-through view of a dry etched wire over which the shielding layer has been deposited; (b) Right Image: a close-up of the grain structure of the shielding layer.

This resulted in a sample similar to that shown in Figure 6.2.39a. However, in this case, the silicon nitride layer had not fully etched through. [This was not a problem for the experiment being conducted.] The nichrome was stripped off using chrome etch [§3.7.3] and 10nm of nichrome and 200nm of palladium were then deposited onto the surface using the Plassys MEB 450. The sample was then cleaved through the lines and examined in cross-section using the Hitachi S900 SEM.

Figure 6.5.1 shows two micrographs of the resulting sample. From Figure 6.5.1b the grain structure of the deposited palladium layer can be seen, although it is difficult to judge the cone angle from this micrograph. Figure 6.5.2 shows one of the more unusual cases where the silicon nitride trench has actually closed up, due to being too narrow for the thickness and closure angle of the shielding layer. The double-sided angle of closure, as determined from Figure 6.5.2b is 33° (to an accuracy of 0.5°). However, this was the exception rather than the rule, and Figure 6.5.3 shows a more typical case. From Figure 6.5.3b the double-sided angle of closure is 21° (to an accuracy of 0.5°). Note that these measurements assume that the image scaling of the SEM in orthogonal directions

is unity to better than 2.5%. This figure matches quite well with the 19° figure estimated from aperture measurements as discussed in §6.2.3. The difference between the results may simply be due to statistical errors alone, or the cone angle may actually exhibit a dependence on the geometry of the aperture. The measurement here has been taken using what is effectively a rectangular aperture, while the other measurement was taken using a circular aperture. However, for the purposes of this discussion it will be assumed that there is no geometric dependence and that the best guess of the cone angle is a simple average of the two results, *i.e.* 20°.

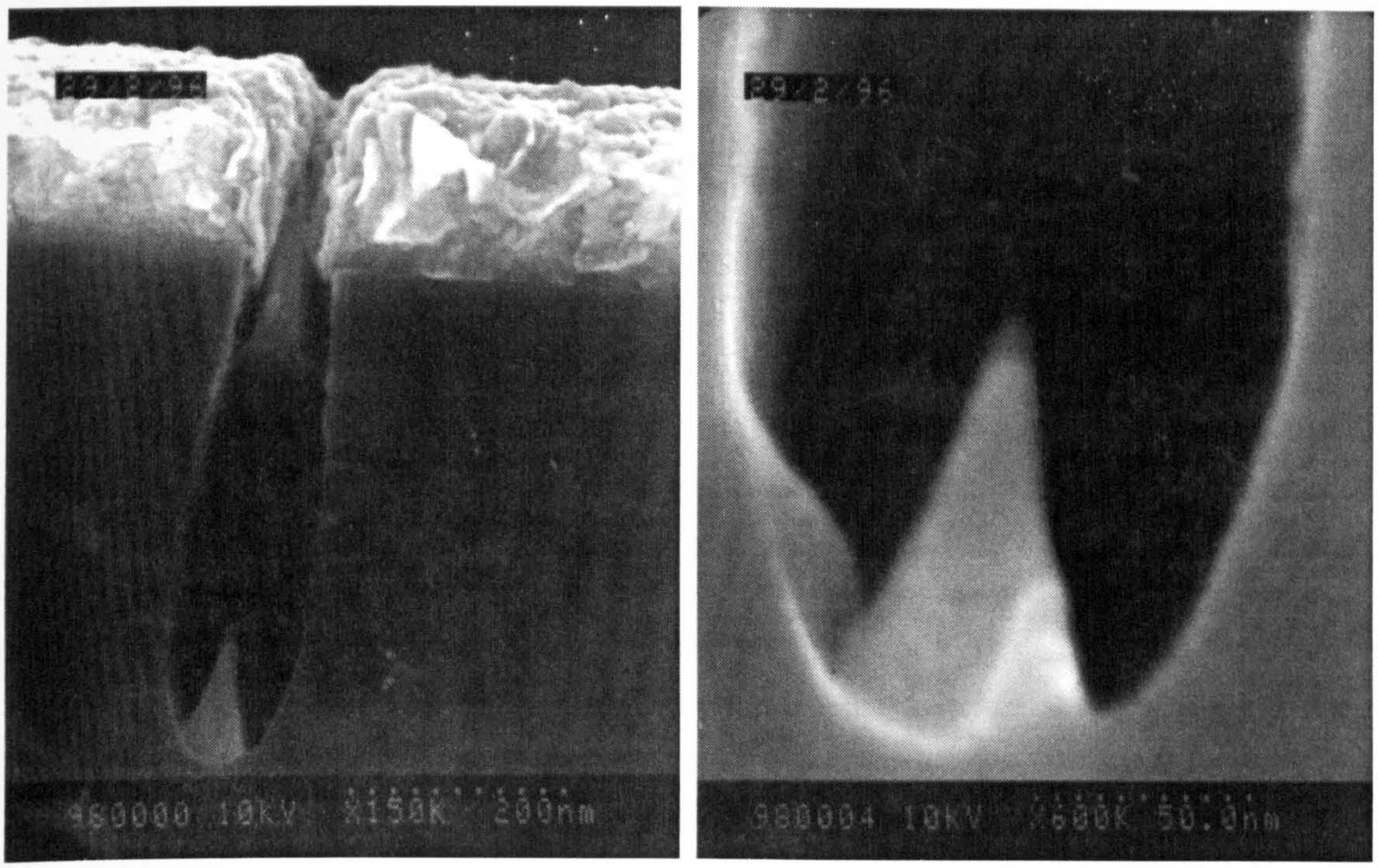


Figure 6.5.2: Micrographs illustrating an attempt to estimate the expected cone angle of a SNOM probe based on 200nm of palladium shielding ('ama0038'). An unusual version of Figure 6.5.1 where the 'aperture' has closed up during evaporation. (a) Left Image: a cleaved-through view of a dry etched wire over which the shielding layer has been deposited; (b) Right Image: a close-up of (a).

The implications of Equations 6.5.22 and 6.5.23 on the optimal choice of various design parameters would seem to be best evaluated graphically, in terms of functions of these parameters. The equations are sufficiently complex to make direct interpretation, particularly comparisons across different parameters, difficult. Thus Figures 6.5.4 to 6.5.37 attempt to give a reasonably comprehensive picture of the various dependencies. In calculating the data for the following, special care had to be given to the fact that Equations 6.5.22 and 6.5.23 are only valid for the region where the light becomes cut-off. Therefore, where the metal shielding thickness and cone angle are such that the probe is not cut-off for the full thickness of the shielding layer, the attenuation is treated from the point at which the light first becomes cut-off, assuming unity gain for the preceding region of the probe.

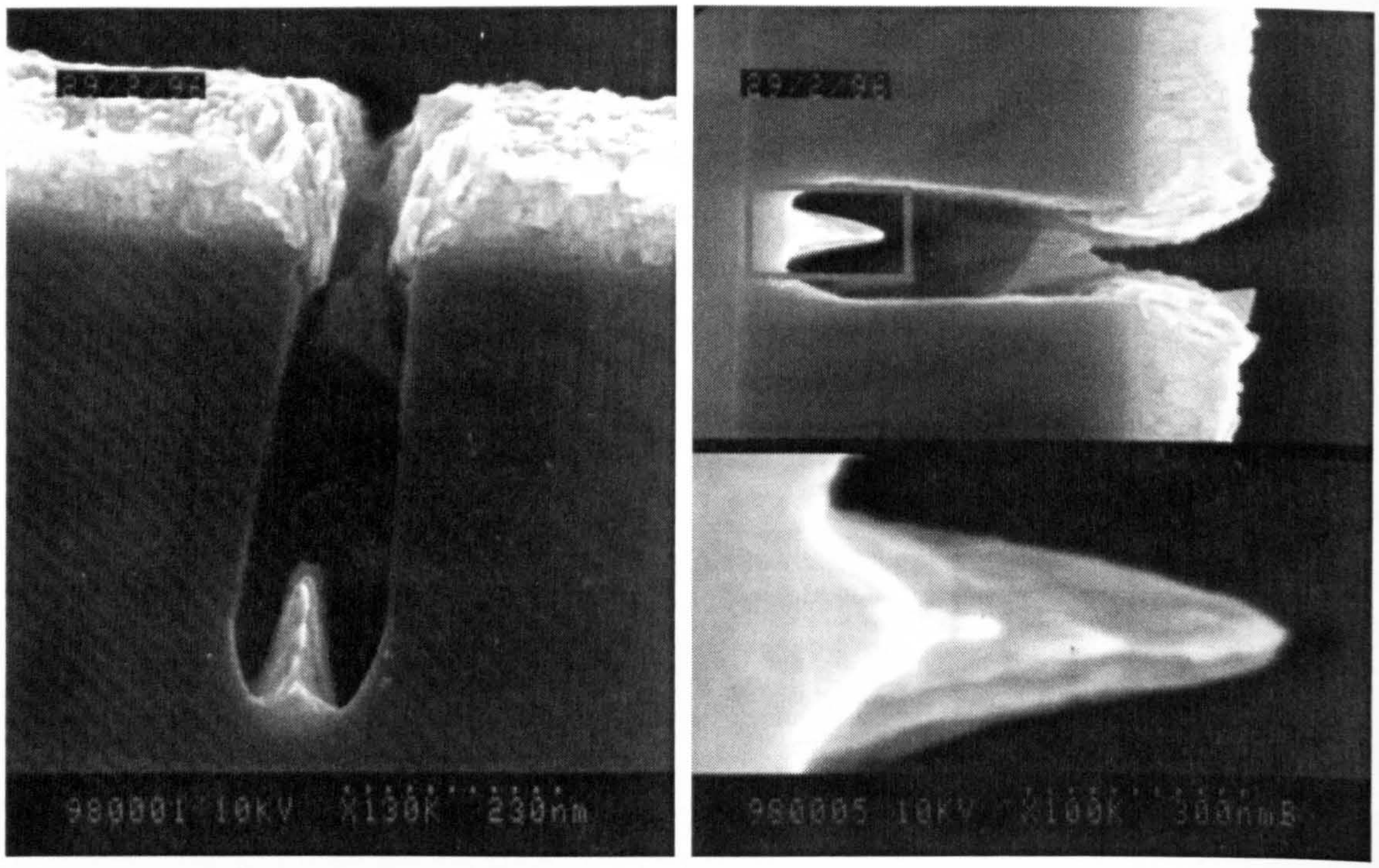


Figure 6.5.3: Micrographs illustrating an attempt to estimate the expected cone angle of a SNOM probe based on 200nm of palladium shielding ('ama0038'). More typical versions of Figure 6.5.1. (a) Left Image: a cleaved-through view of a dry etched wire over which the shielding layer has been deposited; (b) Right Image: an evaporated wire from which an estimate of the cone angle can be made. The image has been rotated 90°. The bottom half is a 5x magnification over the displayed scale.

A crude comparison of the probe designed here with a more conventional fibre probe can also be done by assuming:

(1) The fibre, after cut-off, has the same geometry as the probe here, *i.e.* a linearly decreasing diameter with position.

(2) The medium in which the light propagates has a nominal refractive index of 1.5.

(3) The thickness of the metal shielding layer is infinite. By the statements in the previous paragraph, this means that the thickness of the metallic shielding layer is effectively taken to be equal to the length of the cut-off region of the fibre, any region outwith being treated as having unity gain.

(4) All the other assumptions stated prior to, and during, the derivations of Equations 6.5.22 and 6.5.23 apply.

Figures 6.5.4 and 6.5.5 indicate the cut-off length (of the TE_{11} mode at 488.0nm) as a function of cone angle and output aperture diameter for media with refractive indices of 1.0 and 1.5 respectively. Figure 6.5.5 can be used to quickly determine the effective thickness of the shielding metal for a fibre SNOM probe, assuming a linear taper profile after cut-off. As can be seen from Figure 6.5.5, even for a fibre probe with a cone angle as low as 10° and an aperture size of 50nm, the length of the cut-off region does not exceed 1μm (for the TE_{11} mode at 488.0nm).

The following are useful notes in interpreting the graphs in Figures 6.5.6 to 6.5.37:

(1) The free space wavelength used is 488.0nm. This is one of the most popular of the shorter, and hence less cut-off, wavelengths used for SNOM.

(2) The results displayed are for the TE_{11} or H_{11} mode, hereafter referred to as TE_{11}, except where specified. From Table 1.4.1 it can be seen that this is the least cut-off mode.

(3) All references to aperture size are to the output aperture size and not the input aperture size, unless stated otherwise. The aperture size should be interpreted as the aperture diameter.

(4) Many of the parameters used to generate each graph are given on the graph using the symbolic notation of the derivation of Equations 6.5.22 and 6.5.23. Any remaining parameters are stated in the figure captions below each graph.

(5) When the gain is shown in dB, the interpretation should be:

$$\text{gain (dB)} = 10\log_{10}\left[\frac{I_b^{\,p,q}}{I_a^{\,p,q}}\right] = -\text{attenuation (dB)} \qquad (6.5.24)$$

where the intensity ratio is that given by Equations 6.5.22 or 6.5.23. As an example, a gain of -60dB (attenuation of +60dB) corresponds to an intensity decrease of 6 orders in magnitude or a gain of 10^{-6}.

(6) As noted above, where the metal shielding thickness and cone angle are such that the probe is not cut-off along the full thickness of the metallic shielding layer, the attenuation is treated from the point at which the light first becomes cut-off, assuming unity gain for the preceding region of the probe.

(7) The gain is considered to be the ratio of output intensity to input intensity and not the ratio of power output to power input. To obtain an estimate for the power ratio, by noting that the intensity is simply the power per unit area, the intensity gain (not in dB) should be multiplied by the square of the ratio of the output aperture diameter to the input aperture diameter (*i.e.* the ratio of the output aperture area to the input aperture area). This assumes that all light outwith the output aperture diameter is lost and that the power is distributed uniformly over the area of the input and output apertures. This last assumption is clearly not correct, so that any power gain figures quoted here should be treated as being less accurate than those for the intensity gain. When the light is not cut-off at the input aperture position, the aperture diameter used in a power gain calculation is that when the cut-off first occurs, *i.e.* the cut-off diameter is used in place of the input aperture diameter whenever the input aperture diameter exceeds the cut-off diameter.

Figures 6.5.6 and 6.5.7 show how the gain varies as a function of the shield thickness and aperture size for a probe with a cone angle of 20° in media of refractive index 1.0 and 1.5 respectively. For convenience Table 6.5.1 shows some selected values of the intensity and power gain (the latter being calculated using the above technique) for a medium of refractive index 1.0. It appears that the power gain is typically around 10% less than the intensity gain (in dB). However, it should be noted that this observation is only valid for the choice of parameters used to generate the table. It can also be seen that the light emitted from a 50nm diameter aperture would appear approximately 5 times less bright to the naked eye than that from a 200nm diameter aperture (since the eye responds logarithmically to the absolute light intensity).

Shield Thickness (nm)	Aperture Diameter (nm)	Intensity Gain (dB)	Power Gain (dB)
100	200	-9.55	-10.96
100	150	-15.52	-17.36
100	100	-24.99	-27.61
100	50	-47.07	-51.70
150	200	-12.96	-15.00
150	150	-21.58	-24.20
150	100	-34.59	-38.28
150	50	-63.17	-69.44
200	200	-15.42	-18.04
200	150	-26.66	-30.01
200	100	-42.75	-47.39
200	50	-76.38	-84.02

Table 6.5.1: The intensity and power gains in dB for selected values of the shield thickness and aperture size for the TE_{11} mode at a wavelength of 488.0nm in a medium of refractive index 1.0 using a probe with a cone angle of 20°. Note that the power gain figures here should be treated as less accurate than the intensity gain figures – see text for explanation.

Figures 6.5.8 to 6.5.13 allow comparisons to be made with Figures 6.5.6 and 6.5.7 to determine the effects of changing the cone angles to 30°, 45° and 60°. It is clear that larger cone angles are better because the length of the cut-off region of the probe is reduced. This disproportionately advantages probes with thicker metal shielding layers than probes with thinner metal shielding layers. Note that the white area where thicker metal shielding layer data is expected indicates that those thicknesses of shielding layer are not altering the cut-off region of the probe (*i.e.* the input aperture diameter has exceeded the cut-off diameter for the probe) and so the gain is equal to the lowest value (greatest attenuation) shown for that aperture size.

Figures 6.5.14 to 6.5.21 show more directly the effect of the cone angle on the intensity gain as a function of aperture size but only for selected shielding metal thicknesses and a refractive index of 1.0 and 1.5. In particular, Figures 6.5.20 and 6.5.21 show the effect of having the maximum cut-off thickness of shielding layer, as in the case with more conventional fibre SNOM probes. Figure 6.5.21 is more relevant as it is for a medium of refractive index 1.5. Note that for all these graphs where the refractive index is 1.5, no attenuation occurs for apertures with diameters larger than 190.7nm as this is the cut-off diameter (to one decimal place) for the TE_{11} cylindrical mode at an operating wavelength of 488.0nm (in vacuo).

Figures 6.5.22 to 6.5.29 show the effect of operating wavelength (in vacuo) on the intensity gain as a function of aperture diameter but for selected shielding metal thicknesses, a refractive index of 1.0 and 1.5 and a fixed cone angle of 20°. Clearly, shorter wavelengths are attenuated less as they are cut-off for shorter distances (at a fixed cone angle). Note that the ripples in these figures at the edges of the white regions corresponding to shorter wavelengths and larger aperture sizes are caused by the discrete nature of the data used to construct these graphs (aperture size increments in 5nm steps, wavelength increments in 15nm steps) and are not physically meaningful.

Figures 6.5.30 to 6.5.37 show the effect of refractive index on the intensity gain as a function of aperture diameter but for selected shielding metal thicknesses and a fixed cone angle of 20° for both the TE_{11} and TM_{11} cylindrical modes at a vacuum wavelength of 488.0nm. The main significance of the TM_{11} mode is that the normal mode launched in a fibre is the HE_{11} mode which is a hybrid mode, made up of a superposition of the TE_{11} and TM_{11} modes. As can be seen, the TM_{11} mode is severely attenuated in comparison to the TE_{11} mode so that the output from the TM_{11} mode can be regarded as being negligible.

It is clear from these figures that a high refractive index is better as the cut-off diameter becomes smaller and so for a particular aperture size, metal thickness, cone angle and wavelength, any given mode is cut-off for a shorter distance.

In summary the model developed here suggests the following design rules for SNOM probes:

(1) The metal thickness should be minimised as far as possible, being principally determined by the degree of background light leakage tolerable (layer opacity).

(2) The cone angle should be as large as possible to minimise the cut-off length of the mode.

(3) The wavelength used should be as short as possible, but will usually be limited by the availability of relatively inexpensive, compact and stable lasers, as well as by the availability of shielding materials offering a sufficiently large decay constant at these wavelengths.

(4) The refractive index of the material within the aperture (which should be purely 'real') should be made as large as possible.

(5) If possible, light should be launched at the probe so as to encourage the TE_{11} mode to be excited within the probe.

None of these rules required the derivation of Equations 6.5.22 and 6.5.23 as some simple physics would have revealed the same principal dependencies. However, the model here allows a more quantitative comparison of various compromises to be made quickly and simply and Figures 6.5.4 to 6.5.37 provide a 'ready reckoner' which can be used to compare various designs prior to fabrication. In particular Equations 6.5.22 and 6.5.23 can be used to see how light throughput scales with any given parameter.

Finally Tables 6.5.2 and 6.5.3 allow a crude comparison to be made between the more conventional fibre probe and the probe developed here. The principal problem with comparing these probes is that the profile of the more conventional fibre SNOM probes is not well defined or known. However, it is likely that the assumption that the profile of the probe after cut-off can be approximated by a linear taper is not unreasonable, as this is the final 1 to 2μm of the probe. The critical unknown parameter is that of the cone angle of the probe in this region. Thus Table 6.5.3 displays data for a variety of different cone angles. The other issue is that of what happens to the TM_{11} component of the normal HE_{11} fibre mode within the cut-off region. It is possible that it is attenuated as suggested by Figure 6.5.37, however it is also possible that some or most of the energy in this mode is transferred to the TE_{11} mode. As stated earlier, the model here does not take account of mode transitions or sidewall reflections, and the values given in Table 6.5.3 are only those of a pure launched TE_{11} mode. With these qualifications the intensity gain for a conventional fibre SNOM probe with a 100nm diameter aperture only exceeds that of the equivalent probe designed here (based on 100nm thick shielding metal) for cone angles exceeding 29.40° (as calculated from the model, correct to two decimal places). For a 50nm diameter aperture the situation is even less favourable for the more conventional probe, with the threshold cone angle being 39.24° (as calculated from the model, correct to two decimal places).

	25	50	75	100	125	150	175	200	225	250
200	-2.7	-5.2	-7.5	-9.6	-11.4	-13.0	-14.3	-15.4	-16.2	-16.5
190	-3.0	-5.8	-8.3	-10.6	-12.7	-14.5	-16.2	-17.6	-18.7	-19.5
180	-3.3	-6.4	-9.2	-11.7	-14.1	-16.2	-18.0	-19.7	-21.1	-22.3
170	-3.6	-7.0	-10.1	-12.9	-15.5	-17.9	-20.0	-21.9	-23.6	-25.0
160	-4.0	-7.7	-11.0	-14.2	-17.0	-19.7	-22.0	-24.2	-26.2	-27.9
150	-4.4	-8.4	-12.1	-15.5	-18.7	-21.6	-24.2	-26.7	-28.9	-30.8
140	-4.8	-9.2	-13.3	-17.0	-20.5	-23.7	-26.6	-29.3	-31.7	-34.0
130	-5.3	-10.1	-14.5	-18.7	-22.4	-26.0	-29.2	-32.1	-34.9	-37.4
120	-5.8	-11.1	-16.0	-20.5	-24.7	-28.5	-32.0	-35.3	-38.3	-41.0
110	-6.4	-12.3	-17.6	-22.6	-27.1	-31.3	-35.2	-38.8	-42.1	-45.1
100	-7.1	-13.6	-19.6	-25.0	-30.0	-34.6	-38.8	-42.8	-46.4	-49.7
90	-8.0	-15.2	-21.8	-27.8	-33.3	-38.4	-43.0	-47.3	-51.3	-54.9
80	-9.1	-17.2	-24.5	-31.2	-37.2	-42.8	-47.9	-52.6	-56.9	-60.9
70	-10.4	-19.6	-27.8	-35.2	-42.0	-48.1	-53.8	-58.9	-63.7	-68.1
60	-12.1	-22.7	-32.0	-40.4	-47.9	-54.7	-61.0	-66.6	-71.9	-76.7
50	-14.5	-26.8	-37.6	-47.1	-55.5	-63.2	-70.1	-76.4	-82.1	-87.4
40	-17.8	-32.6	-45.2	-56.2	-65.8	-74.4	-82.2	-89.2	-95.6	-101.4
30	-23.2	-41.5	-56.7	-69.6	-80.7	-90.5	-99.2	-107.1	-114.2	-120.6
20	-33.0	-57.1	-75.9	-91.4	-104.6	-115.9	-125.9	-134.8	-142.7	-149.9
10	-57.3	-92.0	-117.0	-136.4	-152.3	-165.7	-177.3	-187.4	-196.4	-204.5

Table 6.5.2: The intensity gain in dB for selected values of the shield thickness (top row in nm) and aperture size (left column in nm) for the TE_{11} mode at a wavelength of 488.0nm in a medium of refractive index 1.0 using a probe with a cone angle of 20°.

	12.5	15	17.5	20	22.5	25	27.5	30	45	60
200	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
190	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
180	-1.9	-1.6	-1.3	-1.2	-1.0	-0.9	-0.8	-0.8	-0.5	-0.4
170	-5.2	-4.3	-3.7	-3.2	-2.8	-2.6	-2.3	-2.1	-1.4	-1.0
160	-9.6	-8.0	-6.8	-6.0	-5.3	-4.7	-4.3	-3.9	-2.5	-1.8
150	-15.1	-12.6	-10.7	-9.4	-8.3	-7.5	-6.8	-6.2	-4.0	-2.9
140	-21.6	-18.0	-15.4	-13.4	-11.9	-10.7	-9.7	-8.8	-5.7	-4.1
130	-29.3	-24.4	-20.8	-18.2	-16.1	-14.5	-13.1	-12.0	-7.7	-5.6
120	-38.1	-31.7	-27.1	-23.7	-21.0	-18.8	-17.1	-15.6	-10.1	-7.2
110	-48.2	-40.1	-34.3	-30.0	-26.6	-23.8	-21.6	-19.7	-12.8	-9.1
100	-59.9	-49.8	-42.6	-37.2	-33.0	-29.6	-26.8	-24.5	-15.8	-11.4
90	-73.2	-60.9	-52.1	-45.5	-40.3	-36.2	-32.8	-29.9	-19.4	-13.9
80	-88.6	-73.7	-63.0	-55.0	-48.8	-43.8	-39.7	-36.2	-23.4	-16.8
70	-106.5	-88.6	-75.8	-66.2	-58.6	-52.6	-47.7	-43.5	-28.2	-20.2
60	-127.7	-106.2	-90.9	-79.3	-70.3	-63.1	-57.1	-52.2	-33.8	-24.2
50	-153.2	-127.4	-109.0	-95.1	-84.3	-75.7	-68.6	-62.6	-40.5	-29.1
40	-184.8	-153.8	-131.5	-114.8	-101.8	-91.3	-82.7	-75.6	-48.9	-35.1
30	-226.1	-188.1	-160.9	-140.5	-124.5	-111.7	-101.2	-92.4	-59.8	-42.9
20	-284.8	-237.0	-202.7	-176.9	-156.8	-140.7	-127.5	-116.4	-75.3	-54.0
10	-385.8	-320.9	-274.5	-239.6	-212.4	-190.6	-172.7	-157.7	-102.0	-73.2

Table 6.5.3: The intensity gain in dB for selected values of the cone angle (top row in °) and aperture size (left column in nm) for the TE_{11} mode at a vacuum wavelength of 488.0nm in a medium of refractive index 1.5 using a fibre probe.

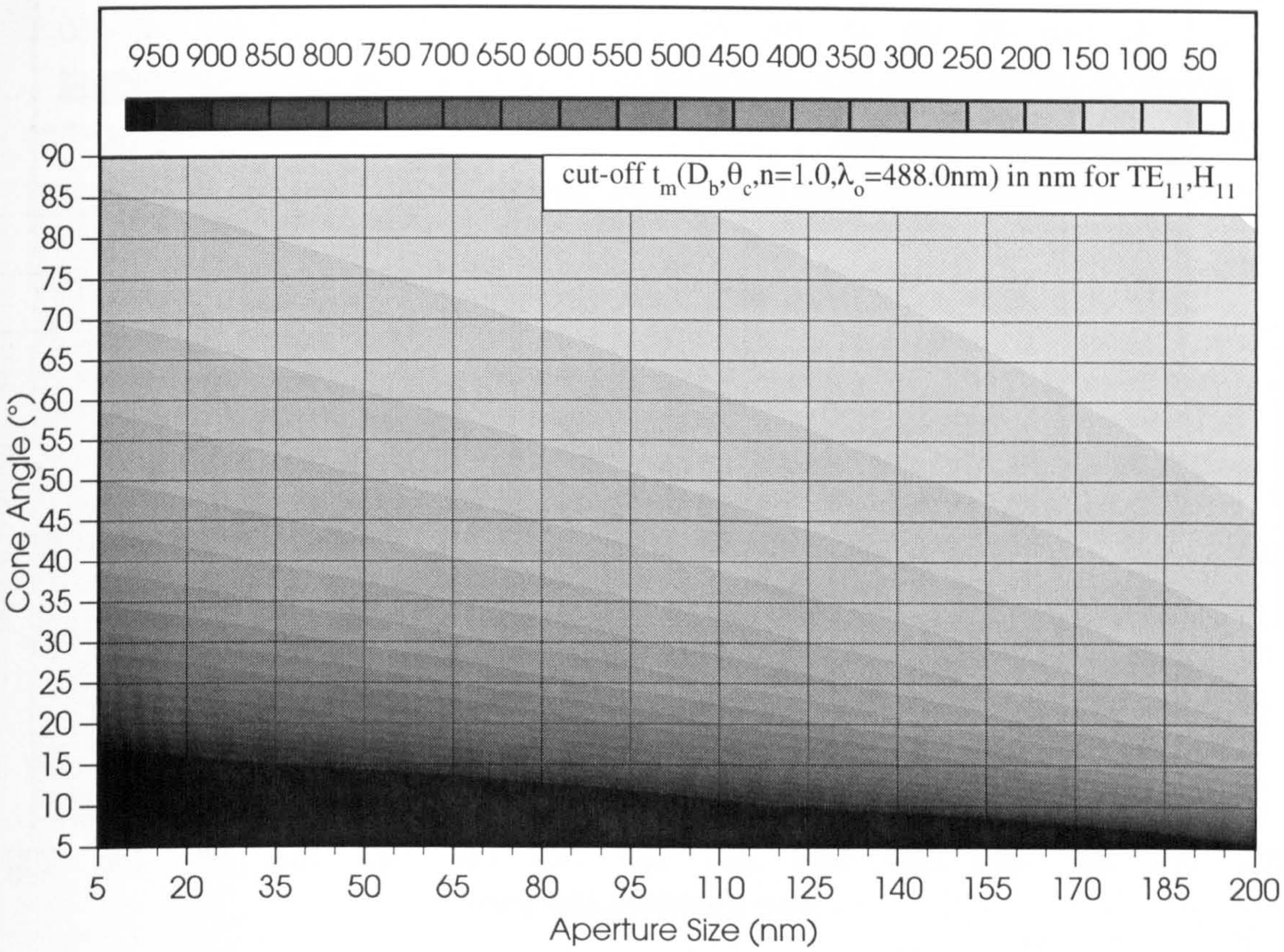


Figure 6.5.4: The maximum cut-off shielding layer thickness in nm as a function of cone angle and aperture size at 488.0nm in a medium of refractive index 1.0.

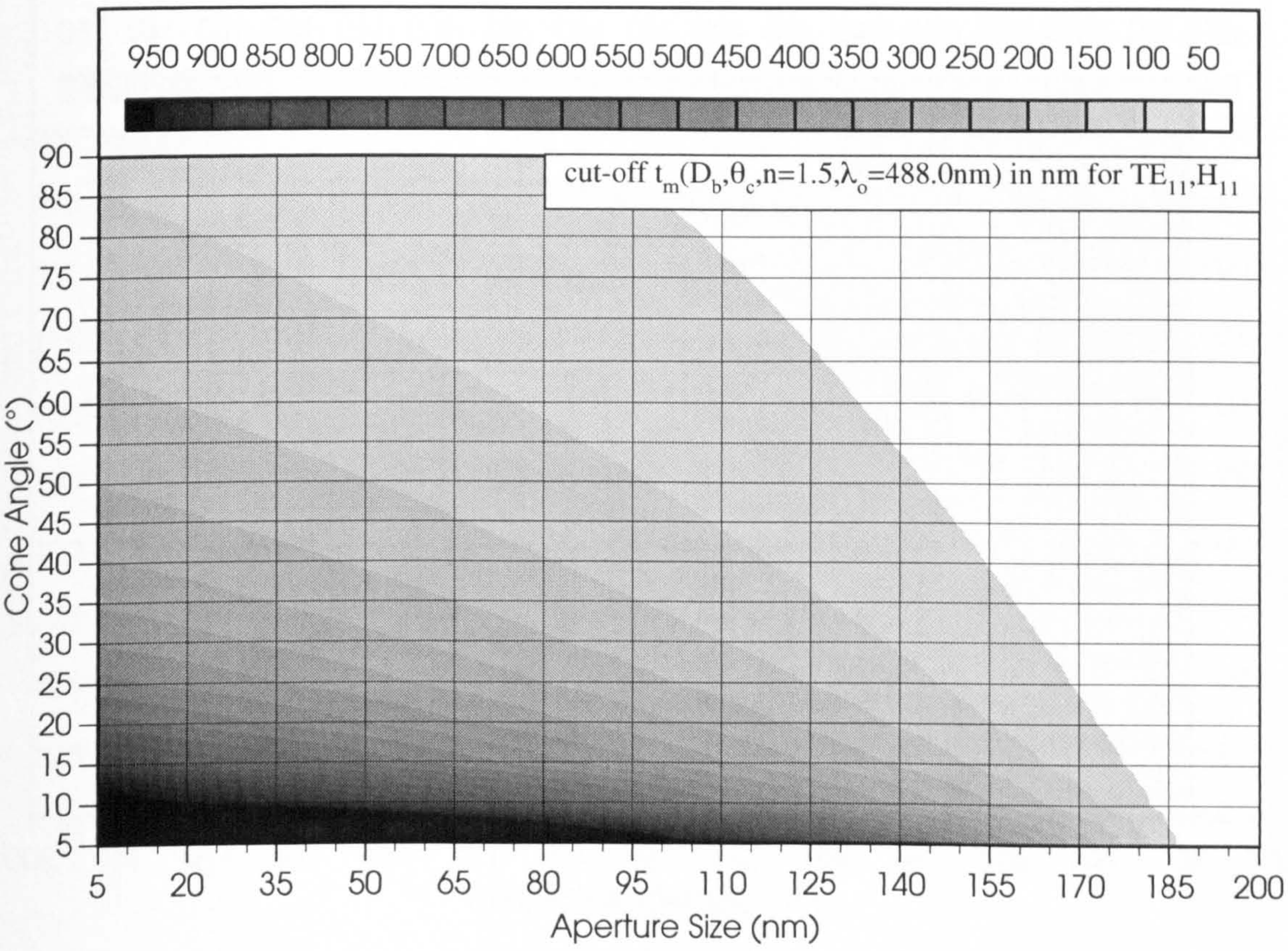


Figure 6.5.5: The maximum cut-off shielding layer thickness in nm as a function of cone angle and aperture size at 488.0nm in a medium of refractive index 1.5.

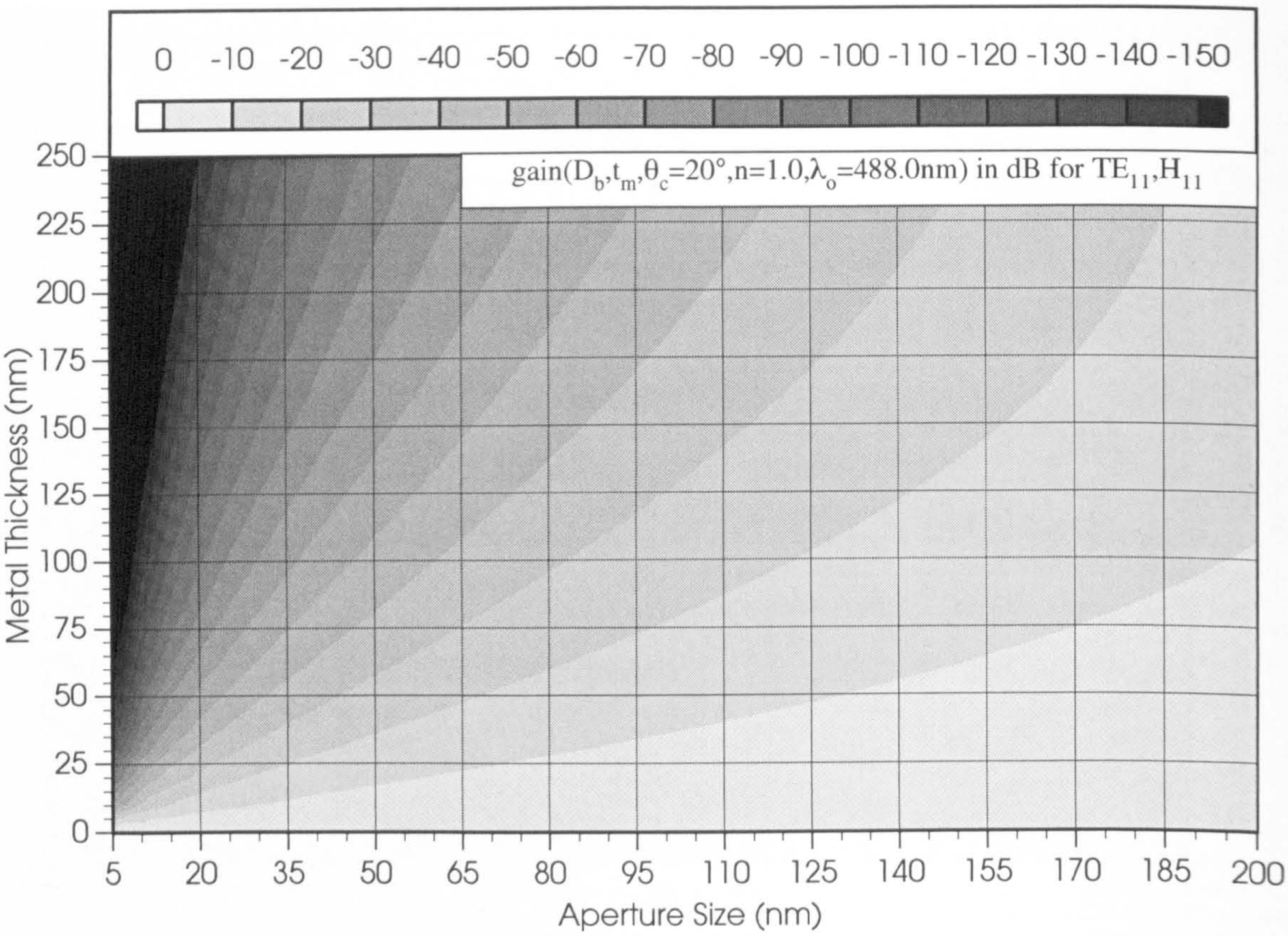


Figure 6.5.6: The gain in dB as a function of the shield thickness and aperture size for the TE_{11} mode with parameters as given on the graph.

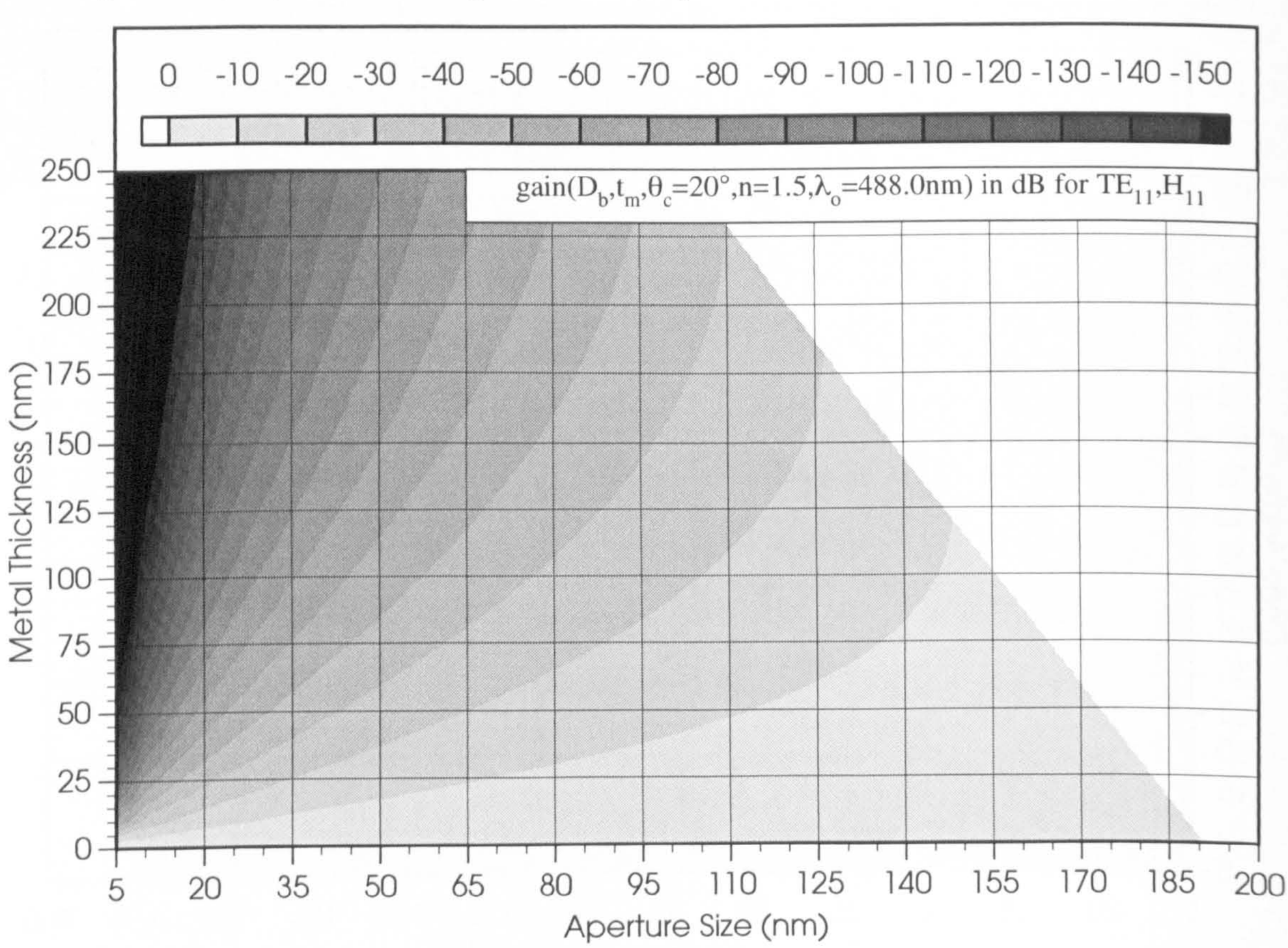


Figure 6.5.7: The gain in dB as a function of the shield thickness and aperture size for the TE_{11} mode with parameters as given on the graph.

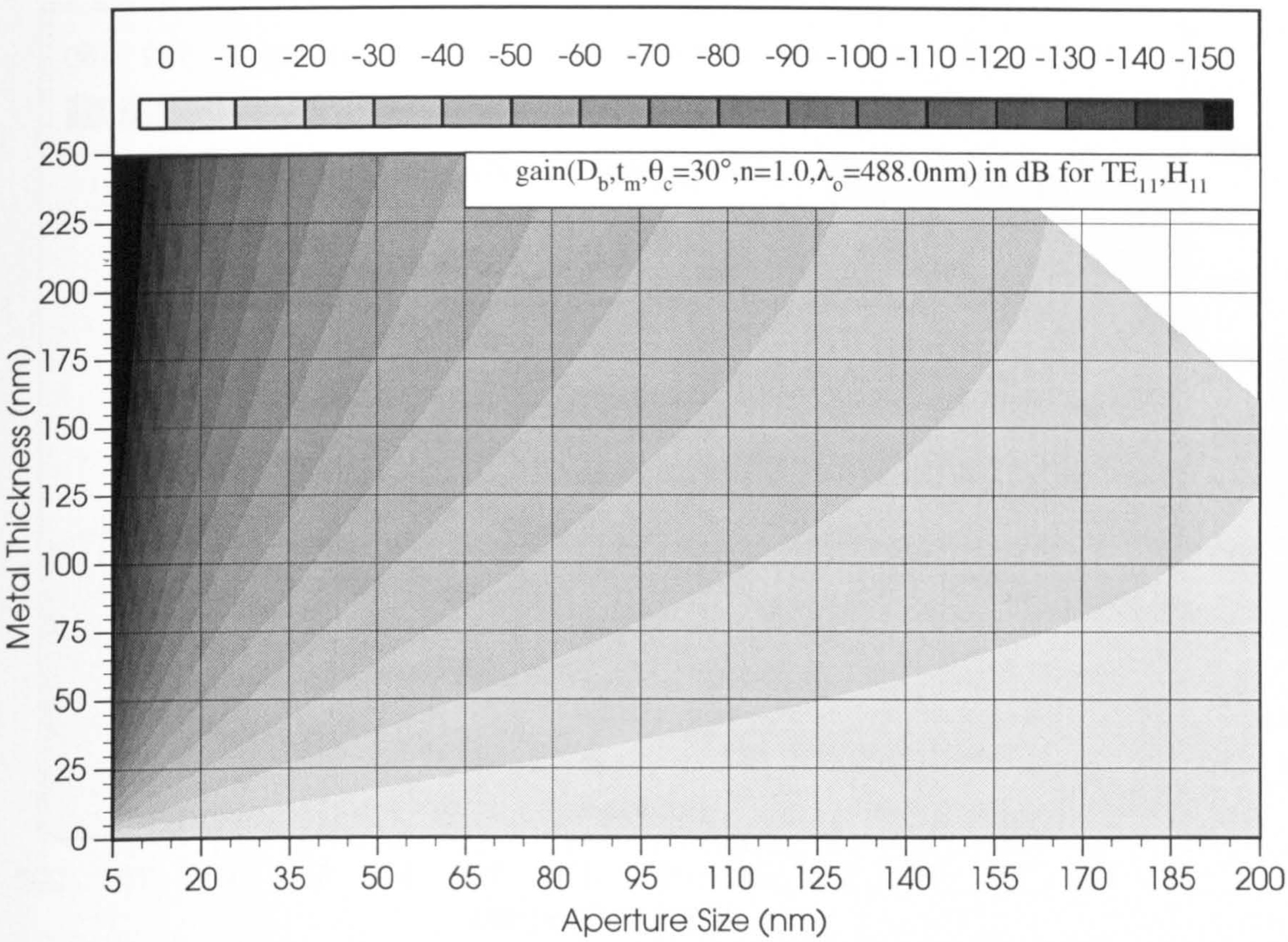


Figure 6.5.8: The gain in dB as a function of the shield thickness and aperture size for the TE_{11} mode with parameters as given on the graph.

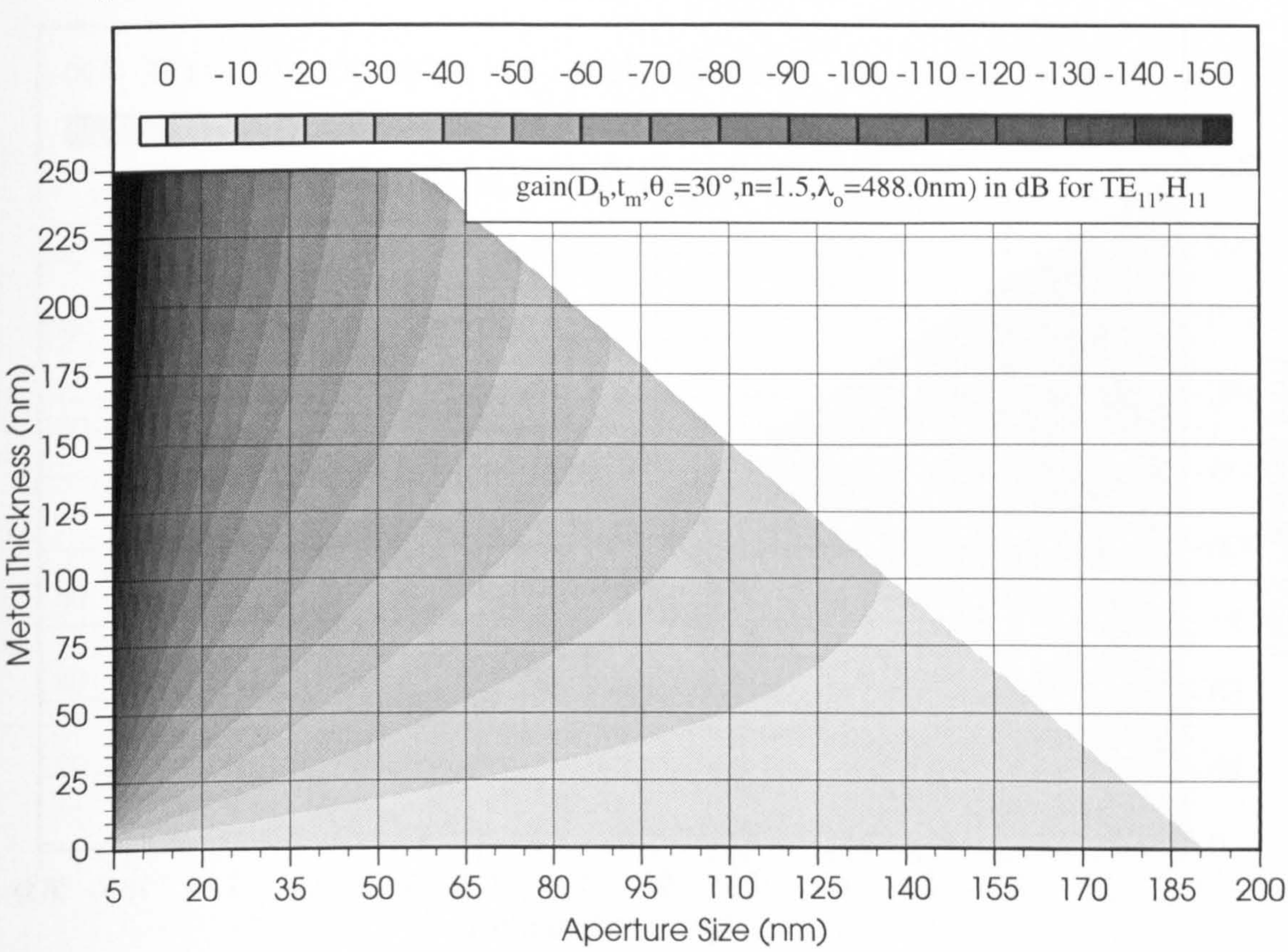


Figure 6.5.9: The gain in dB as a function of the shield thickness and aperture size for the TE_{11} mode with parameters as given on the graph.

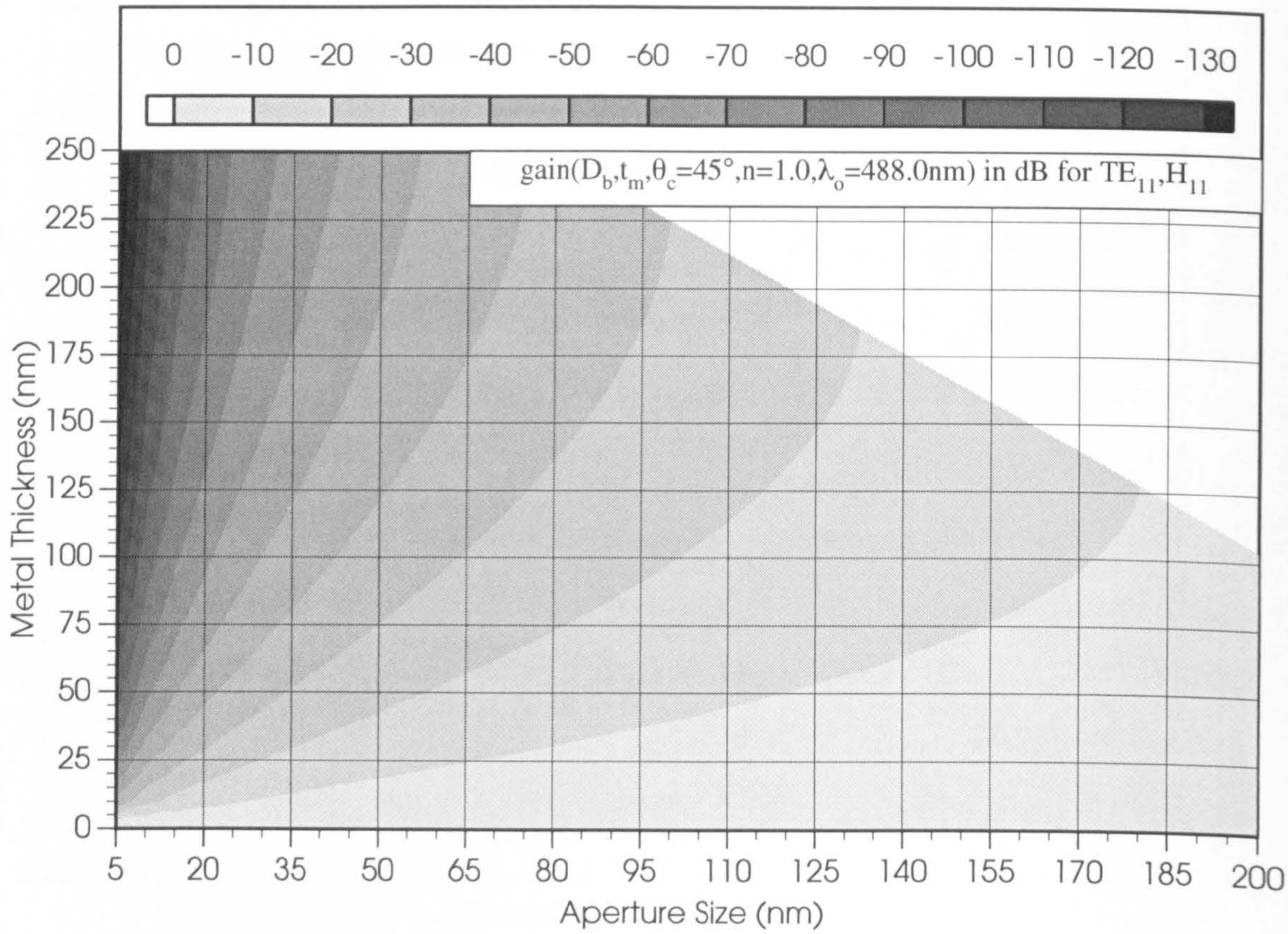


Figure 6.5.10: The gain in dB as a function of the shield thickness and aperture size for the TE_{11} mode with parameters as given on the graph.

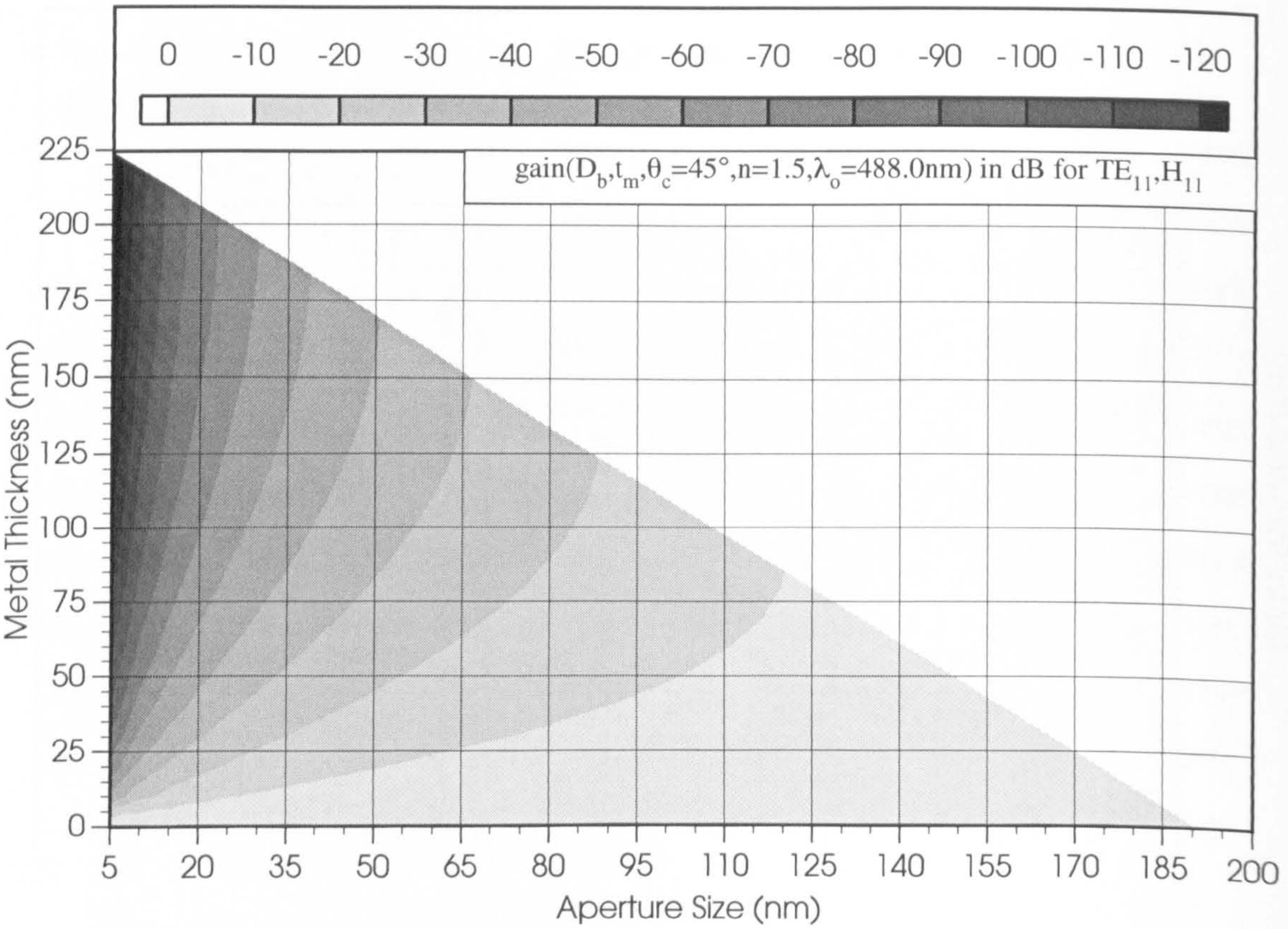


Figure 6.5.11: The gain in dB as a function of the shield thickness and aperture size for the TE_{11} mode with parameters as given on the graph.

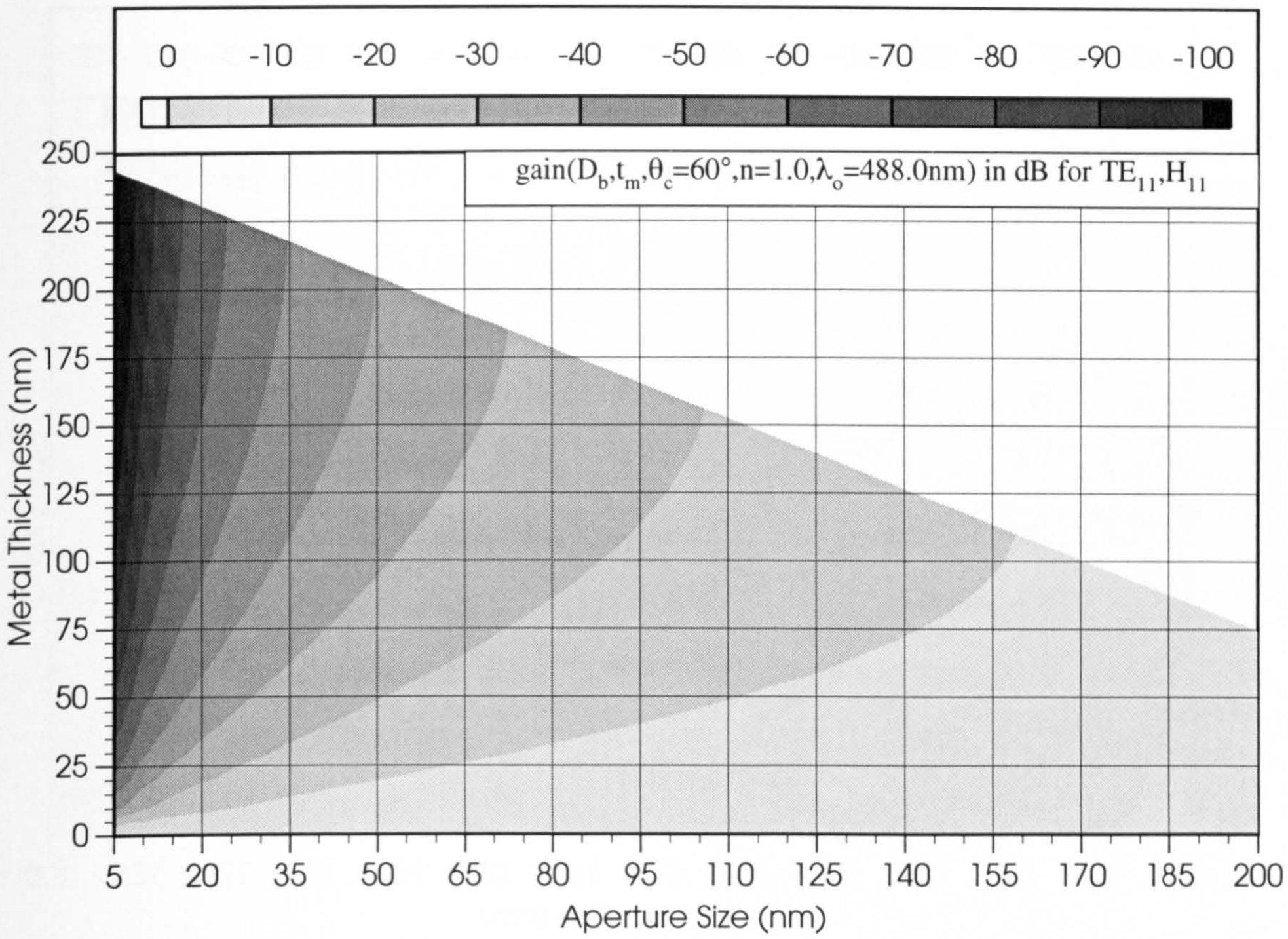


Figure 6.5.12: The gain in dB as a function of the shield thickness and aperture size for the TE_{11} mode with parameters as given on the graph.

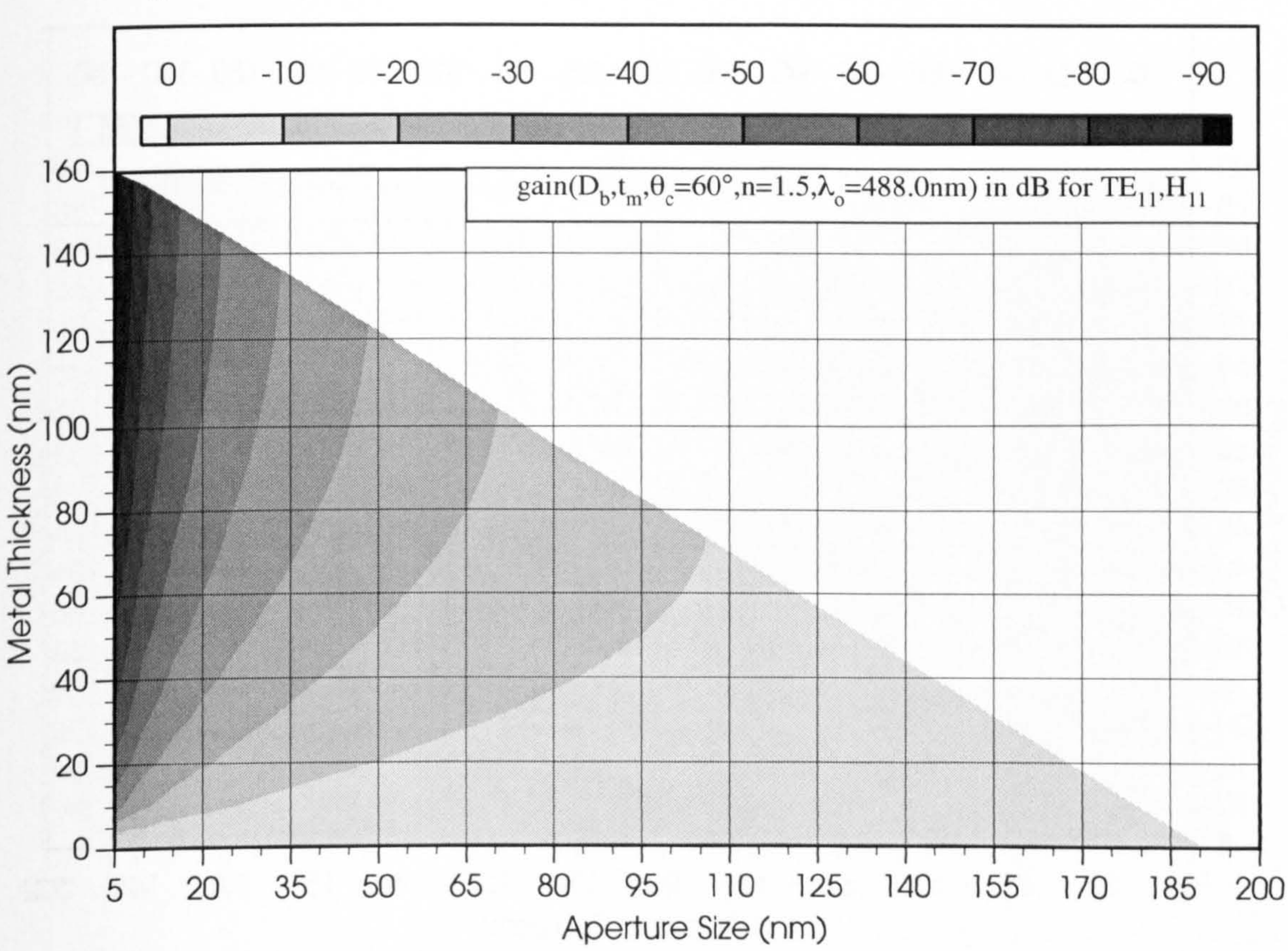


Figure 6.5.13: The gain in dB as a function of the shield thickness and aperture size for the TE_{11} mode with parameters as given on the graph.

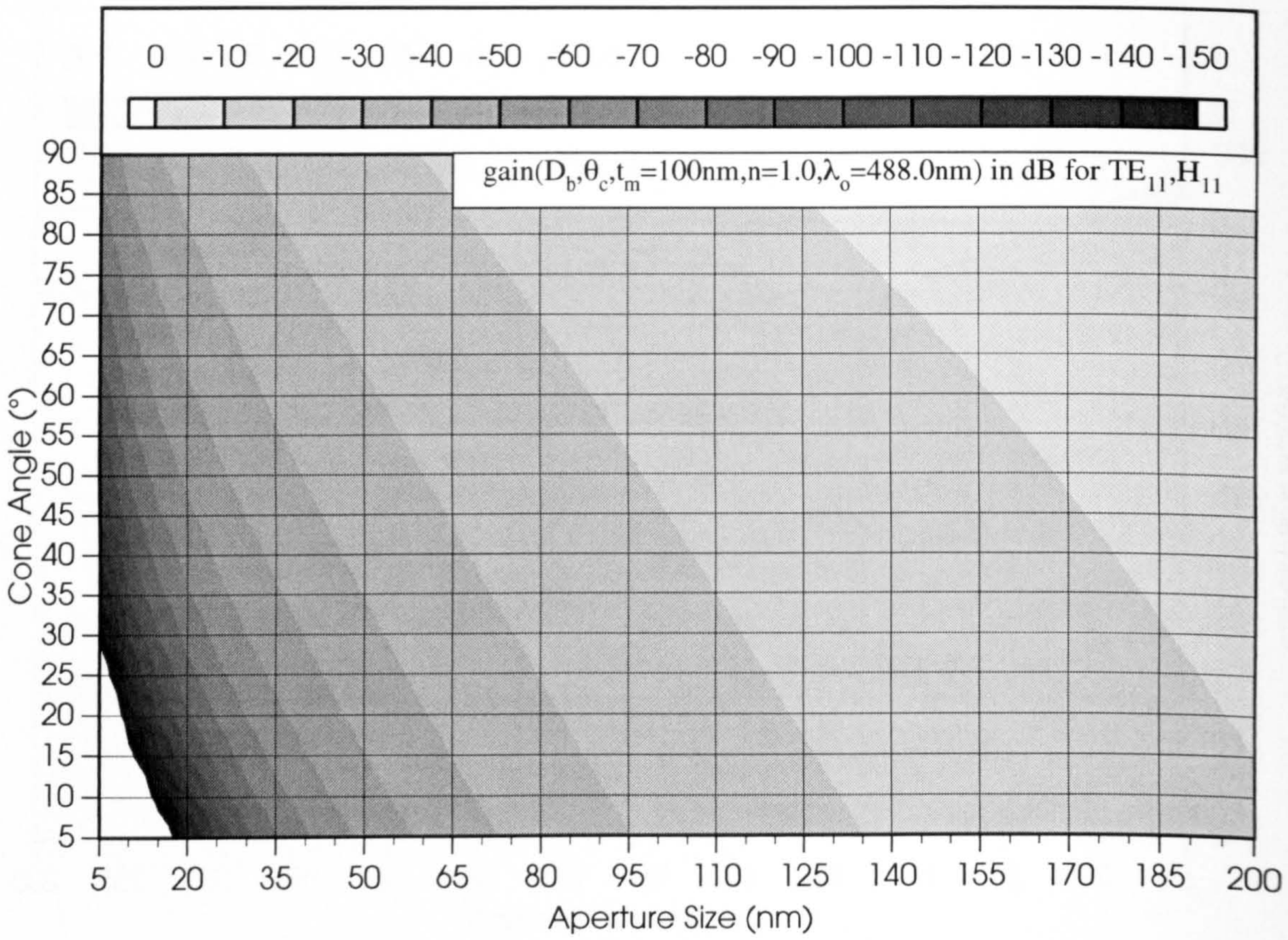


Figure 6.5.14: The gain in dB as a function of the cone angle and aperture size for the TE_{11} mode with parameters as given on the graph.

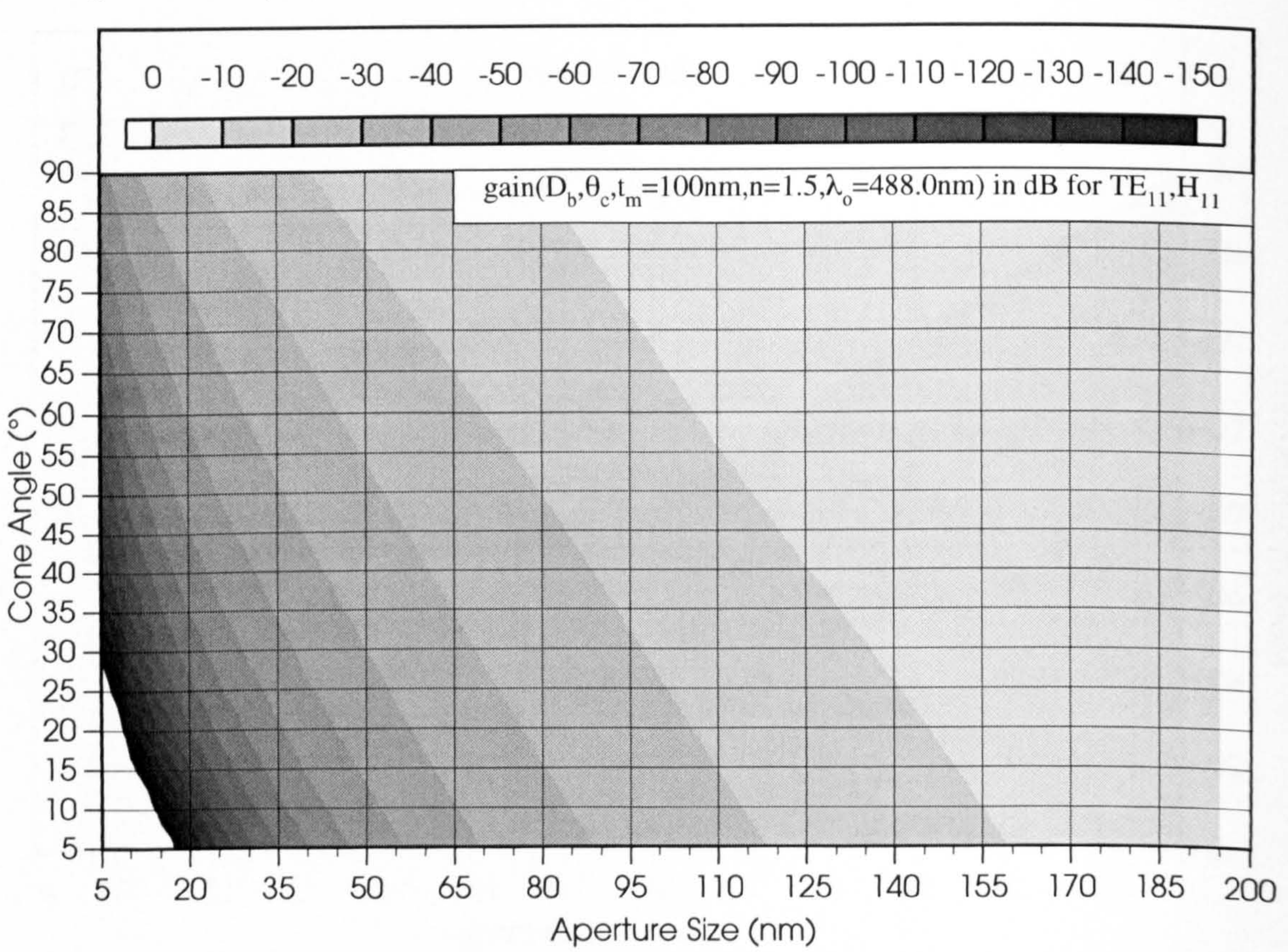


Figure 6.5.15: The gain in dB as a function of the cone angle and aperture size for the TE_{11} mode with parameters as given on the graph.

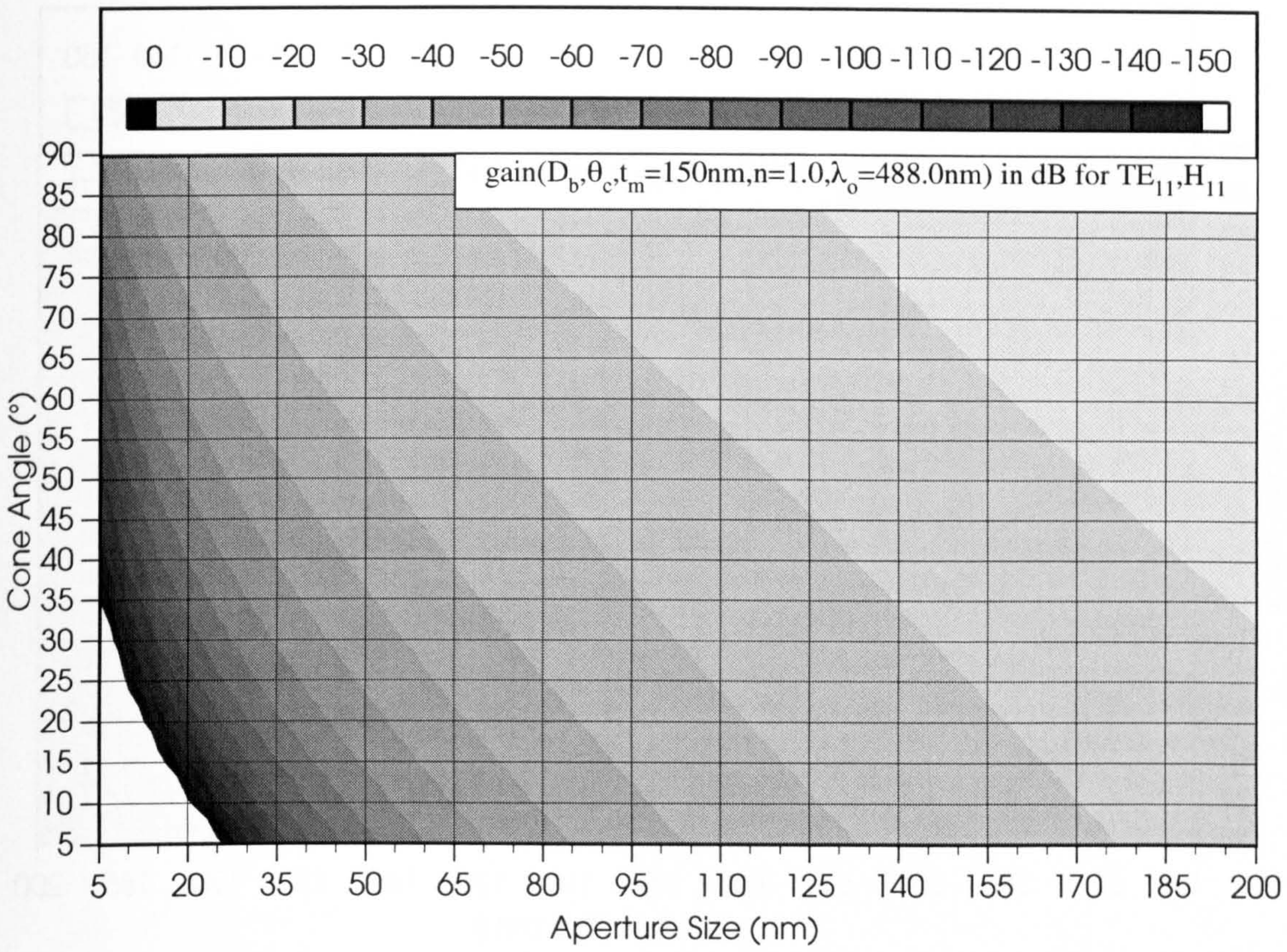


Figure 6.5.16: The gain in dB as a function of the cone angle and aperture size for the TE_{11} mode with parameters as given on the graph.

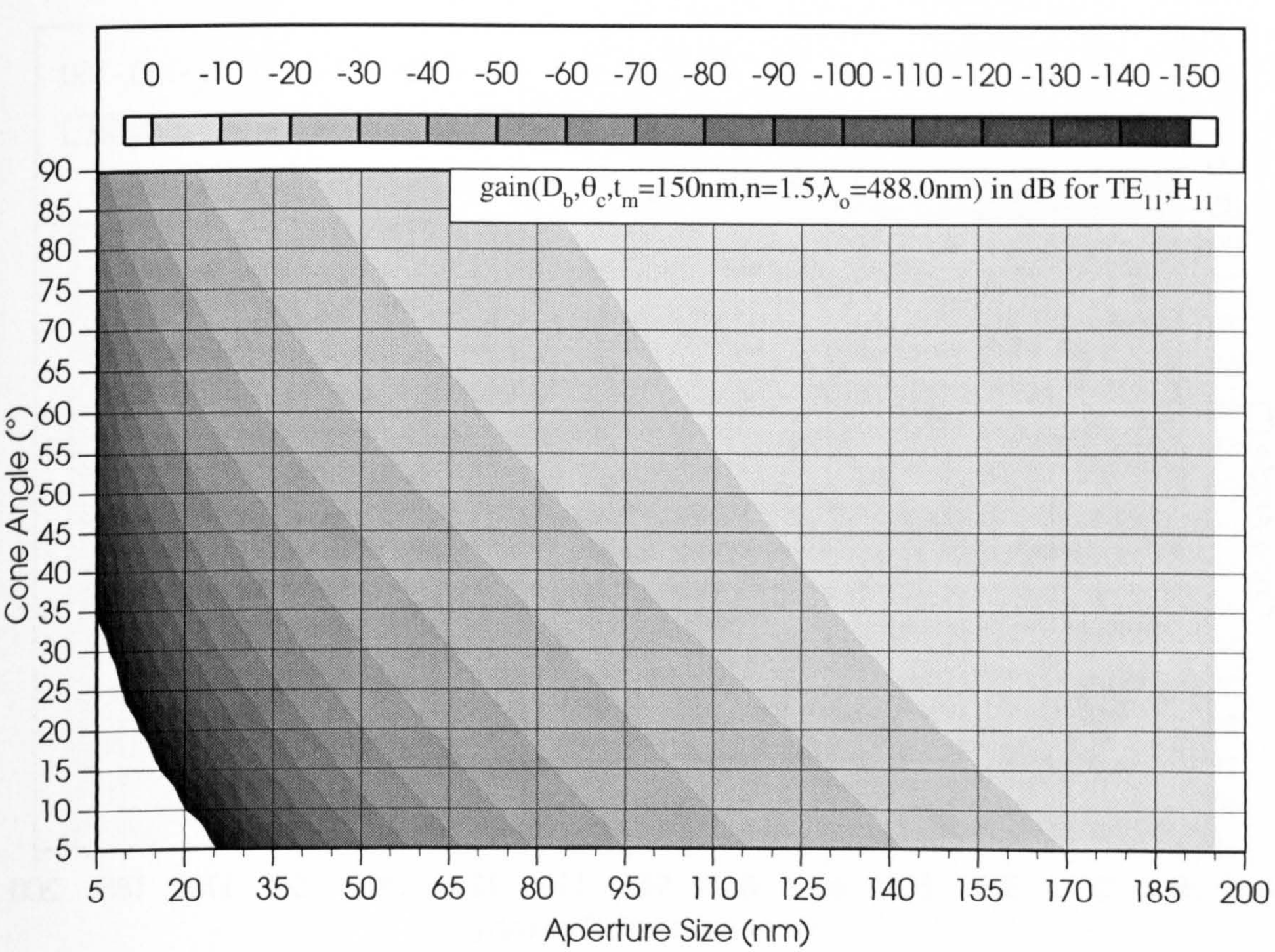


Figure 6.5.17: The gain in dB as a function of the cone angle and aperture size for the TE_{11} mode with parameters as given on the graph.

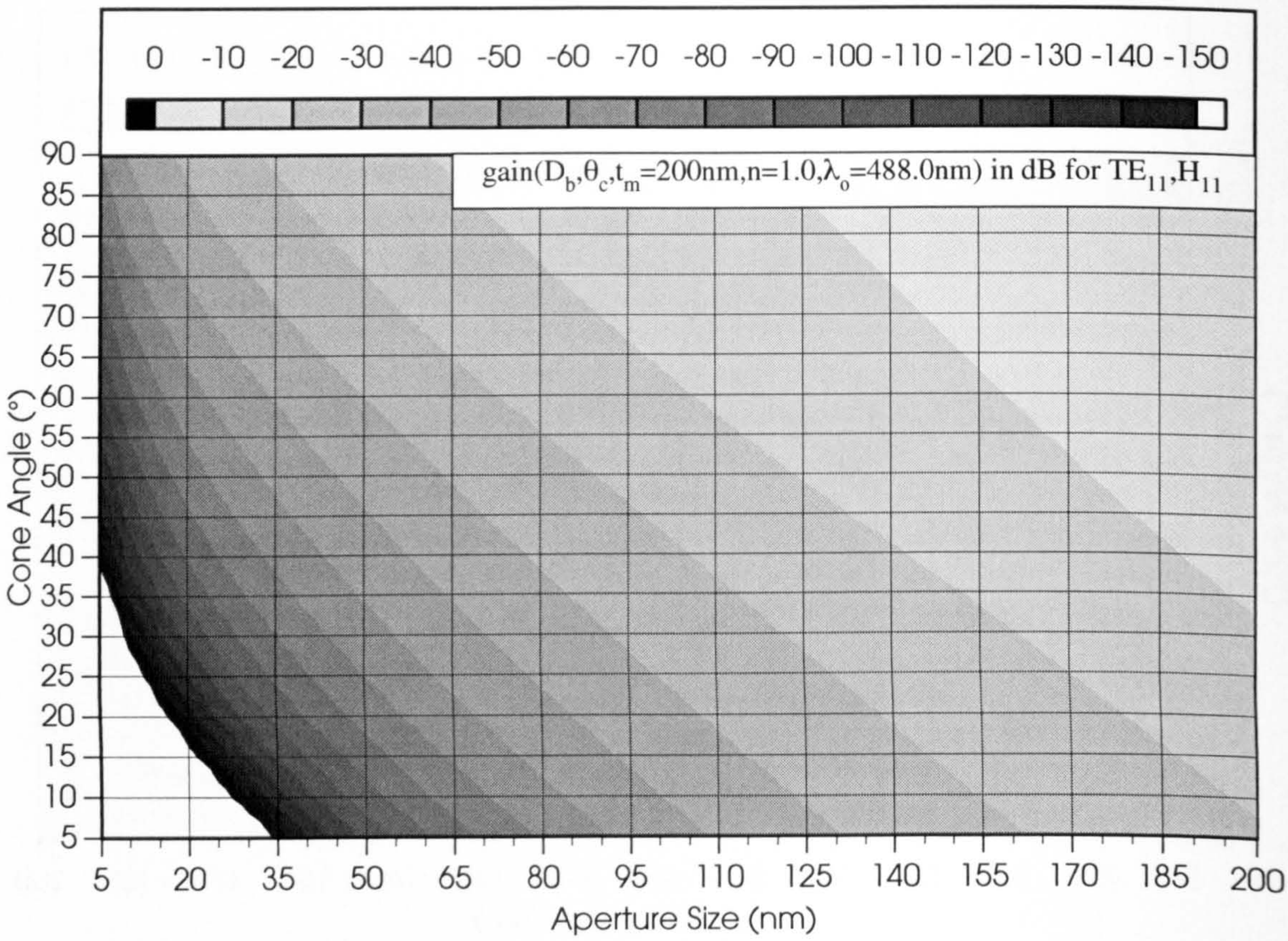


Figure 6.5.18: The gain in dB as a function of the cone angle and aperture size for the TE_{11} mode with parameters as given on the graph.

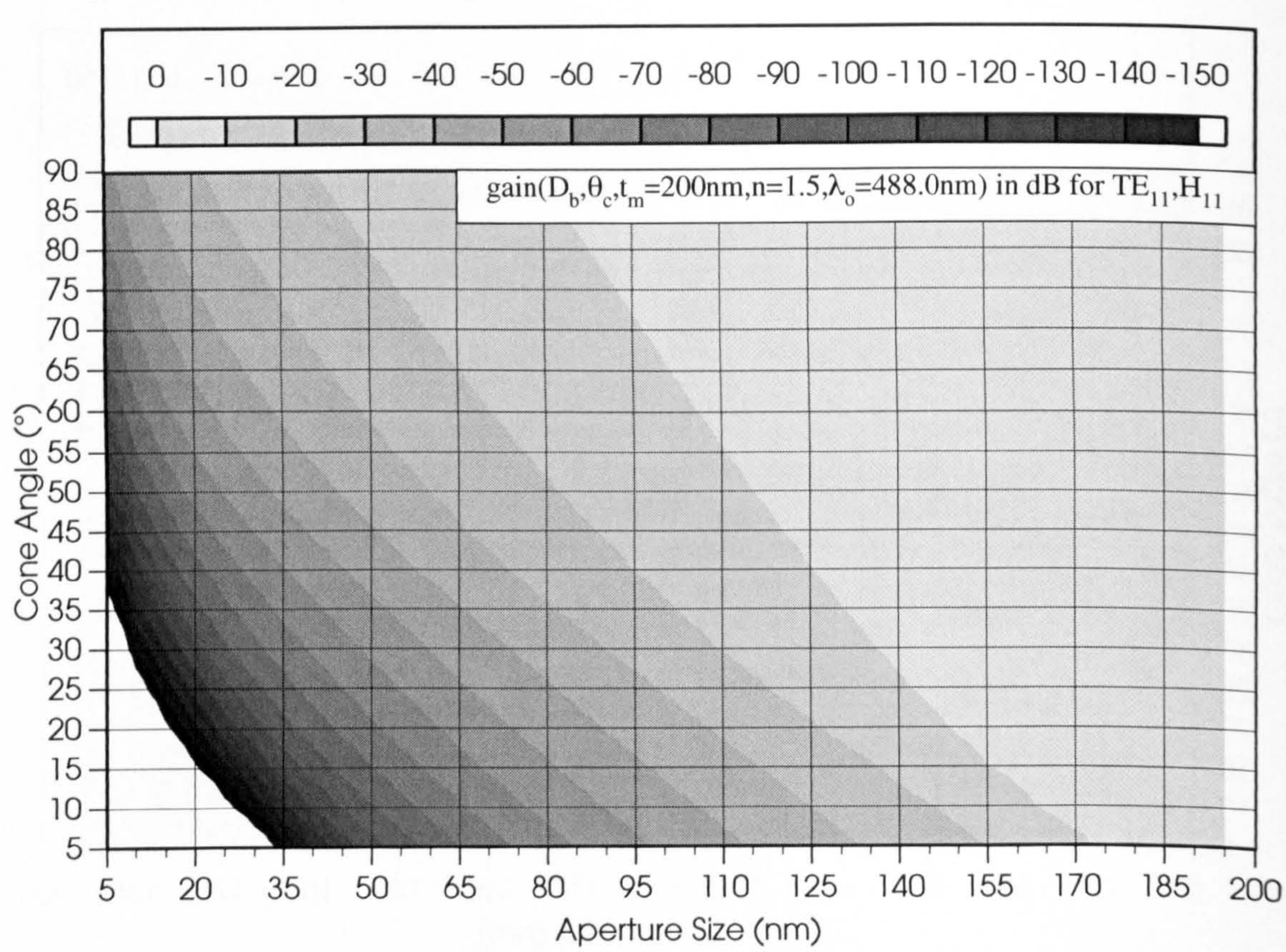


Figure 6.5.19: The gain in dB as a function of the cone angle and aperture size for the TE_{11} mode with parameters as given on the graph.

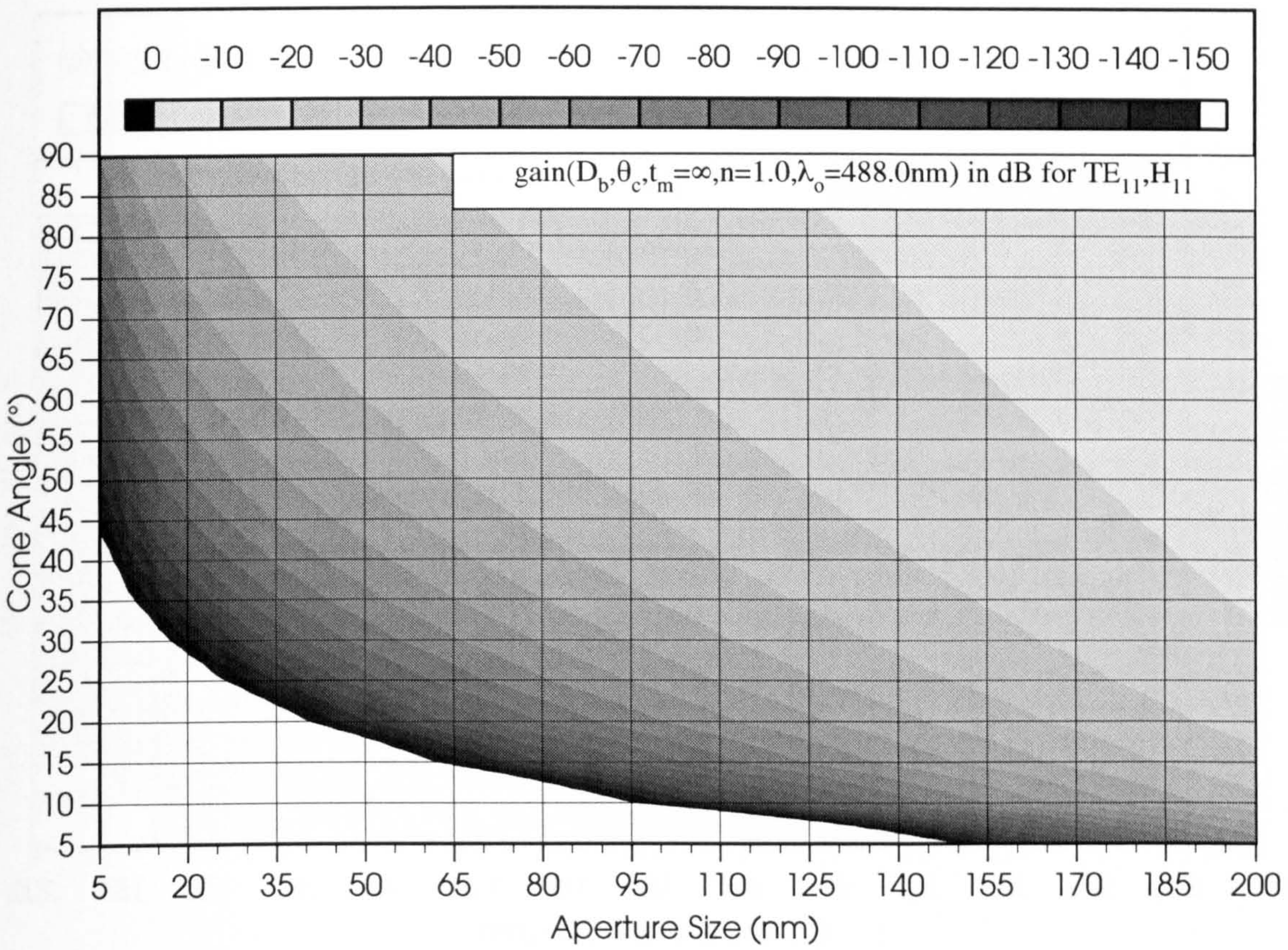


Figure 6.5.20: The gain in dB as a function of the cone angle and aperture size for the TE_{11} mode with parameters as given on the graph.

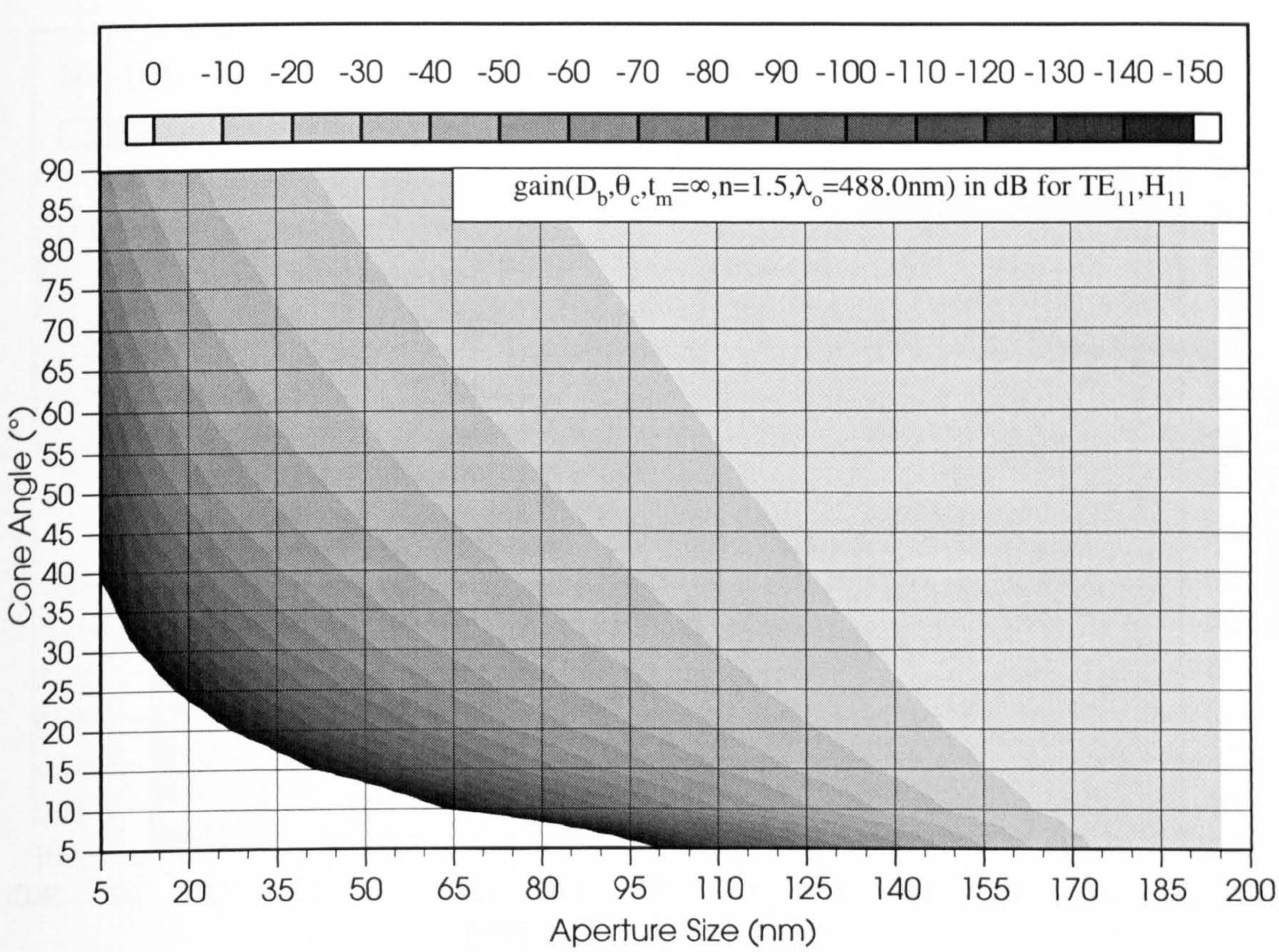


Figure 6.5.21: The gain in dB as a function of the cone angle and aperture size for the TE_{11} mode with parameters as given on the graph.

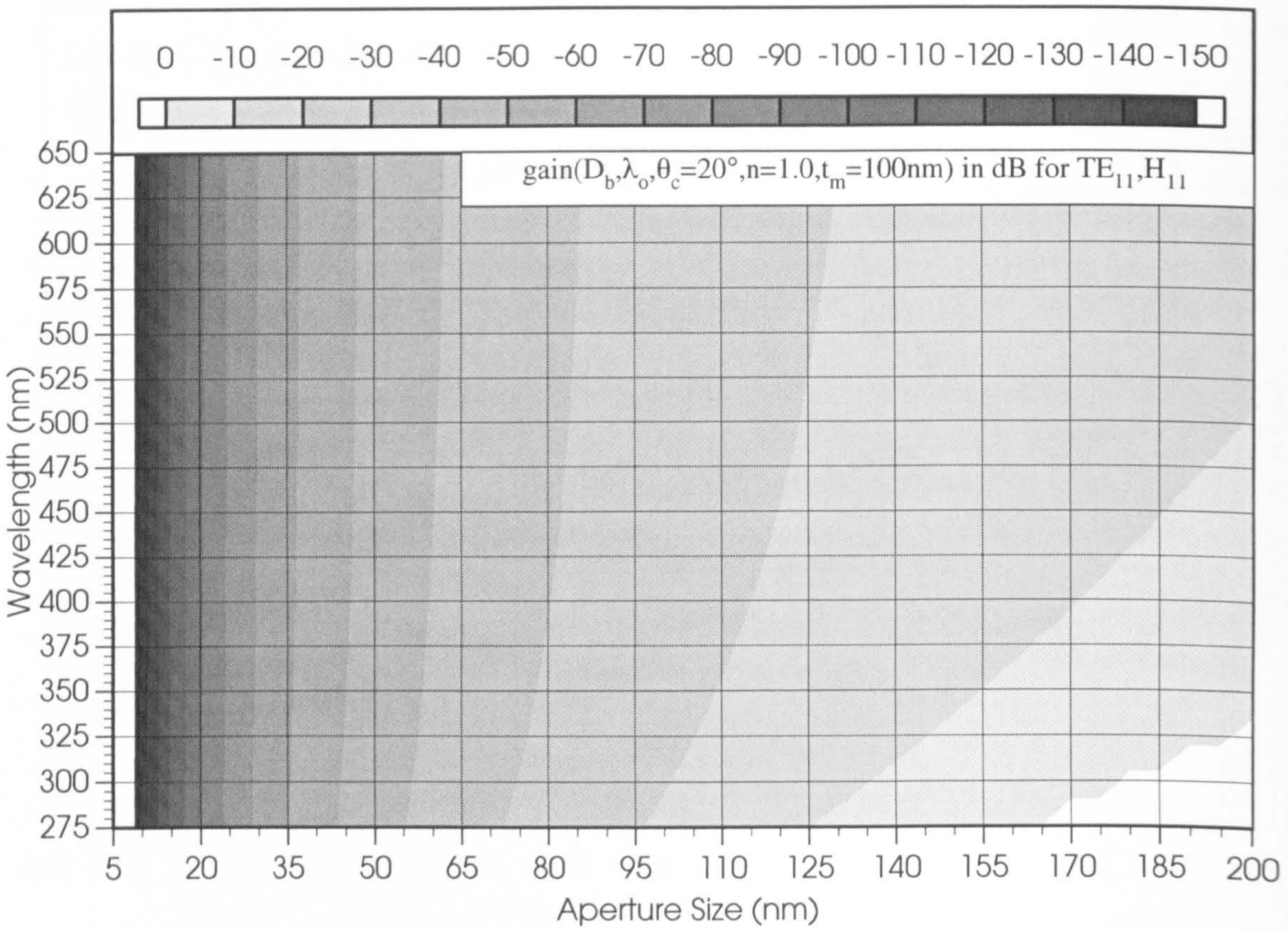


Figure 6.5.22: The gain in dB as a function of the vacuum wavelength and aperture size for the TE_{11} mode with parameters as given on the graph.

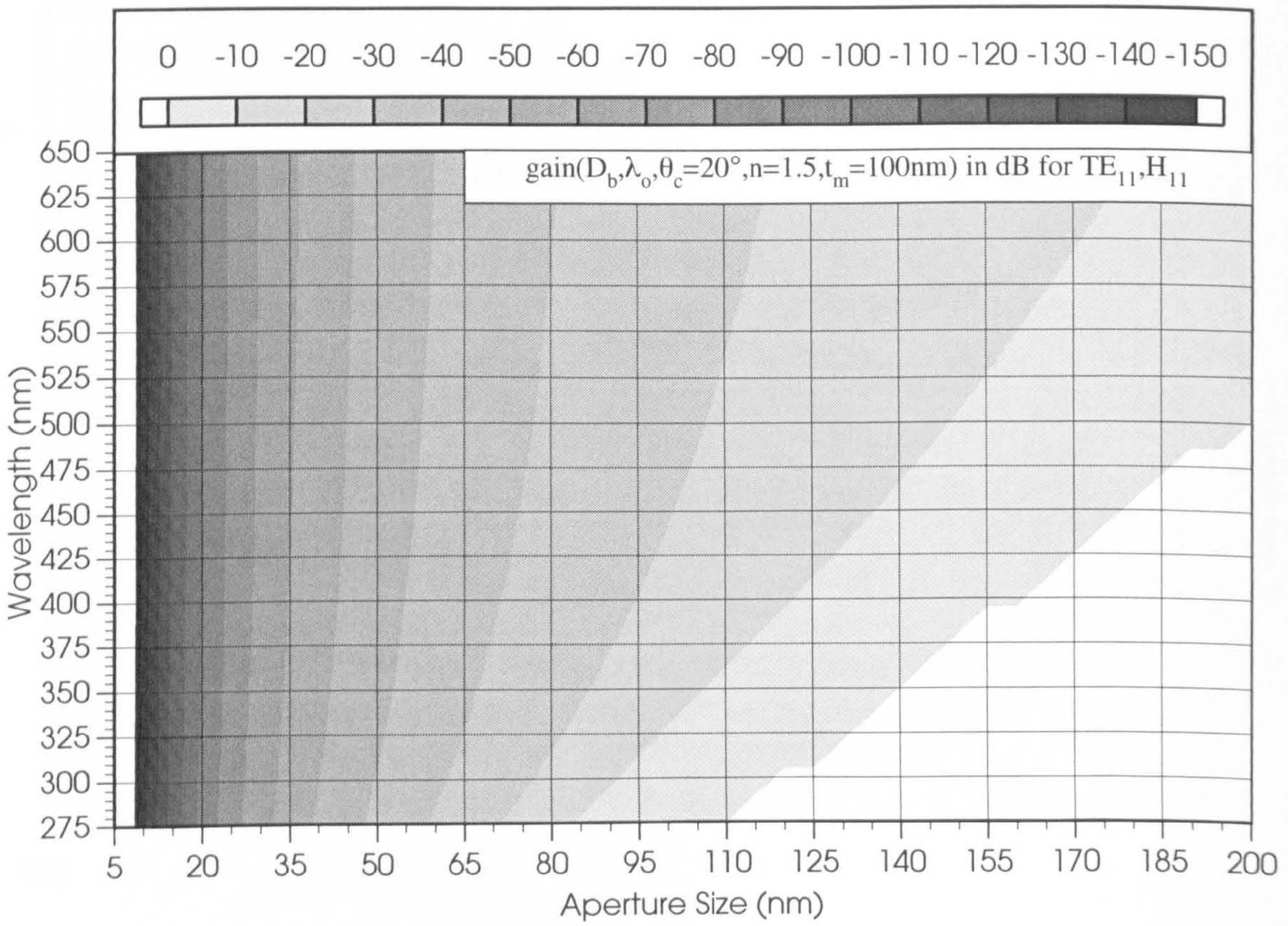


Figure 6.5.23: The gain in dB as a function of the vacuum wavelength and aperture size for the TE_{11} mode with parameters as given on the graph.

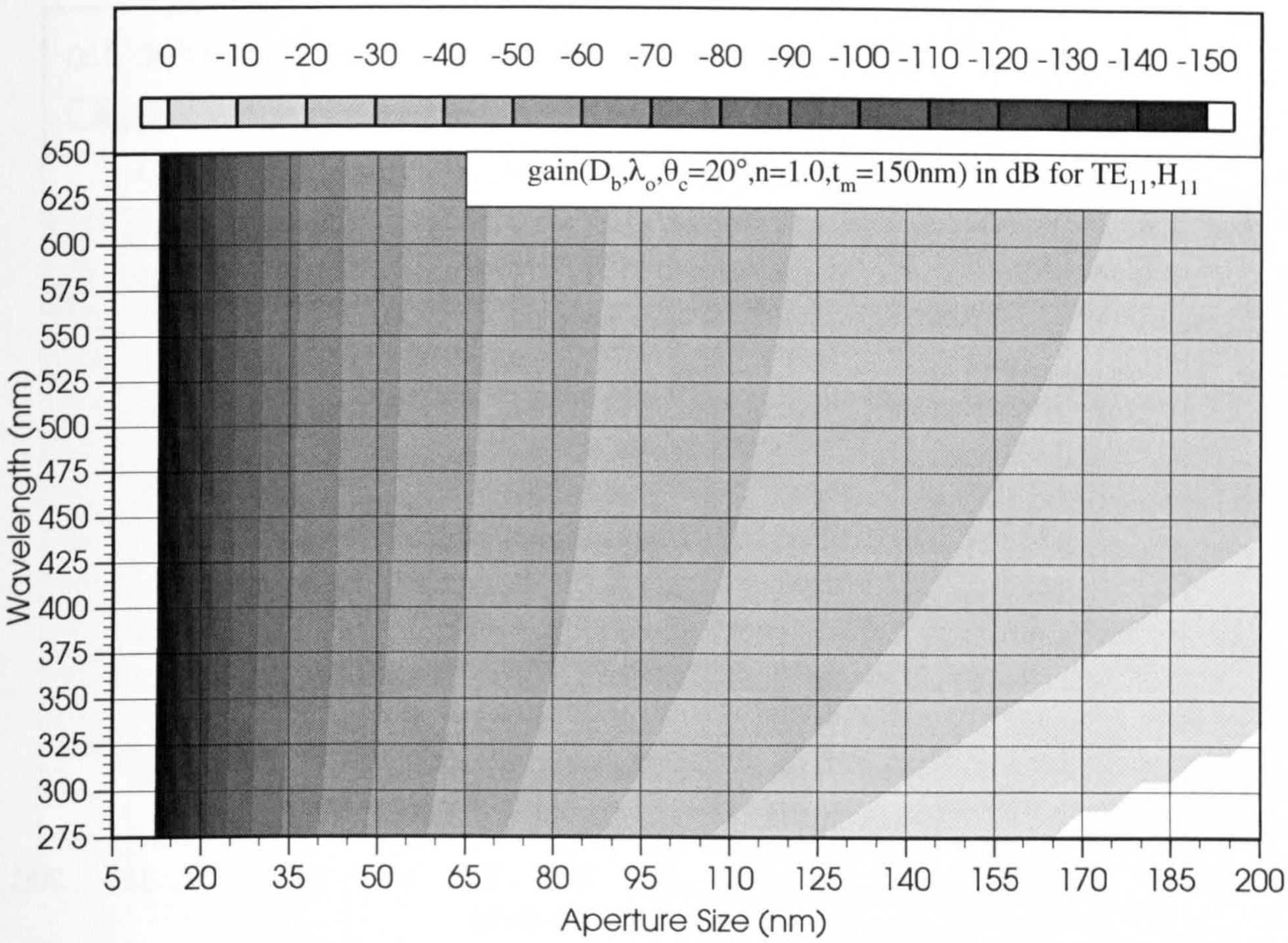


Figure 6.5.24: The gain in dB as a function of the vacuum wavelength and aperture size for the TE_{11} mode with parameters as given on the graph.

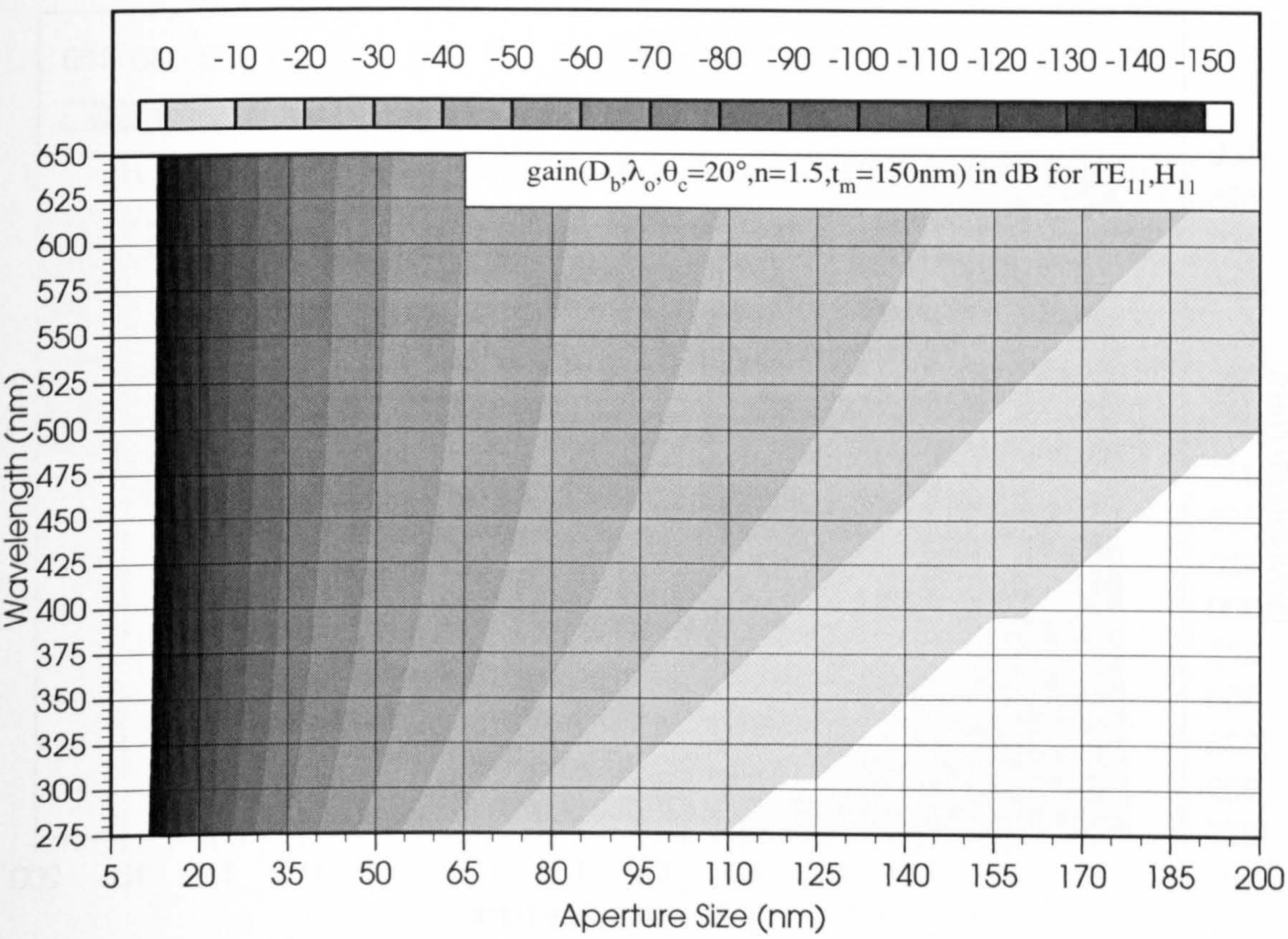


Figure 6.5.25: The gain in dB as a function of the vacuum wavelength and aperture size for the TE_{11} mode with parameters as given on the graph.

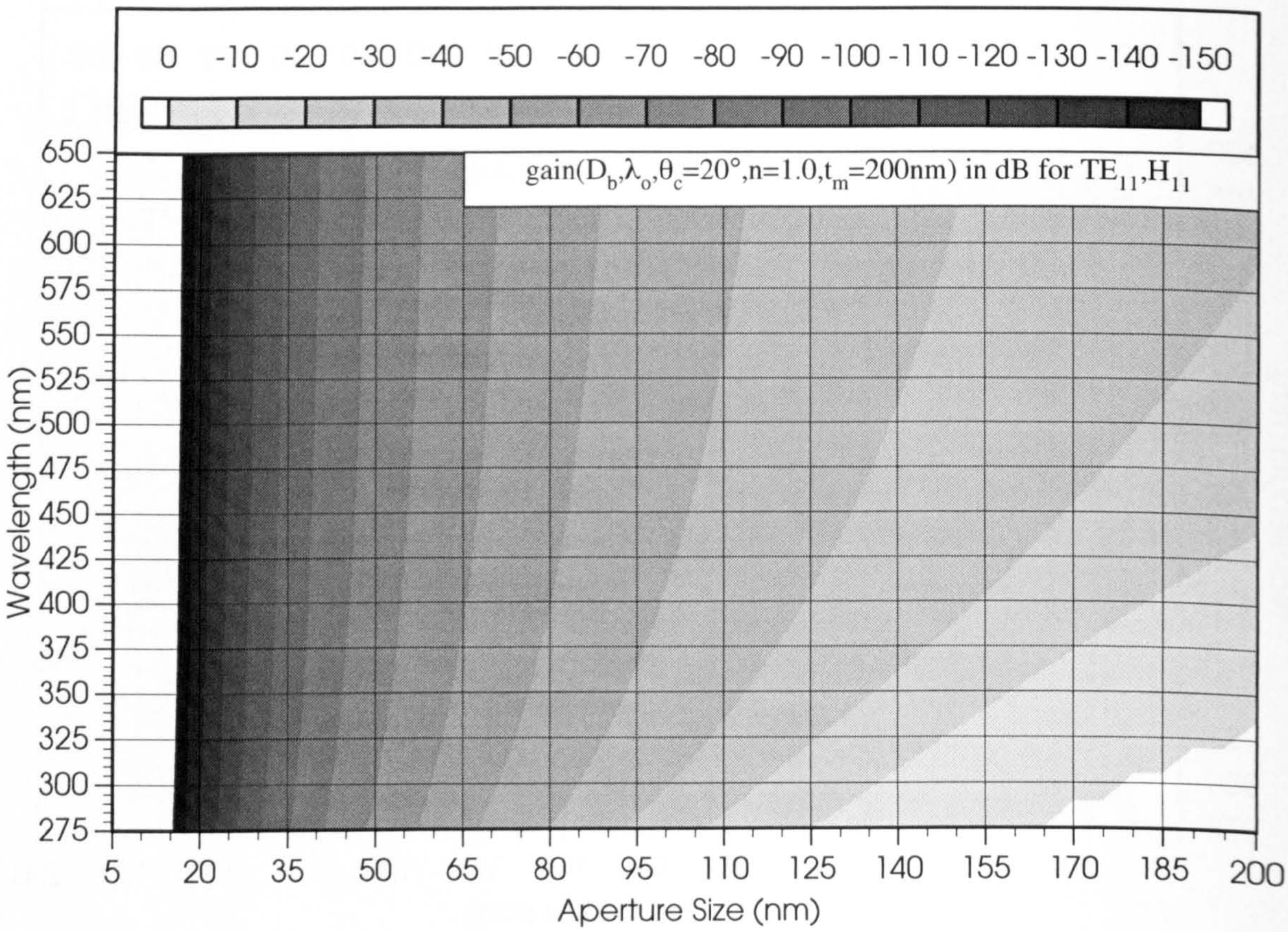


Figure 6.5.26: The gain in dB as a function of the vacuum wavelength and aperture size for the TE_{11} mode with parameters as given on the graph.

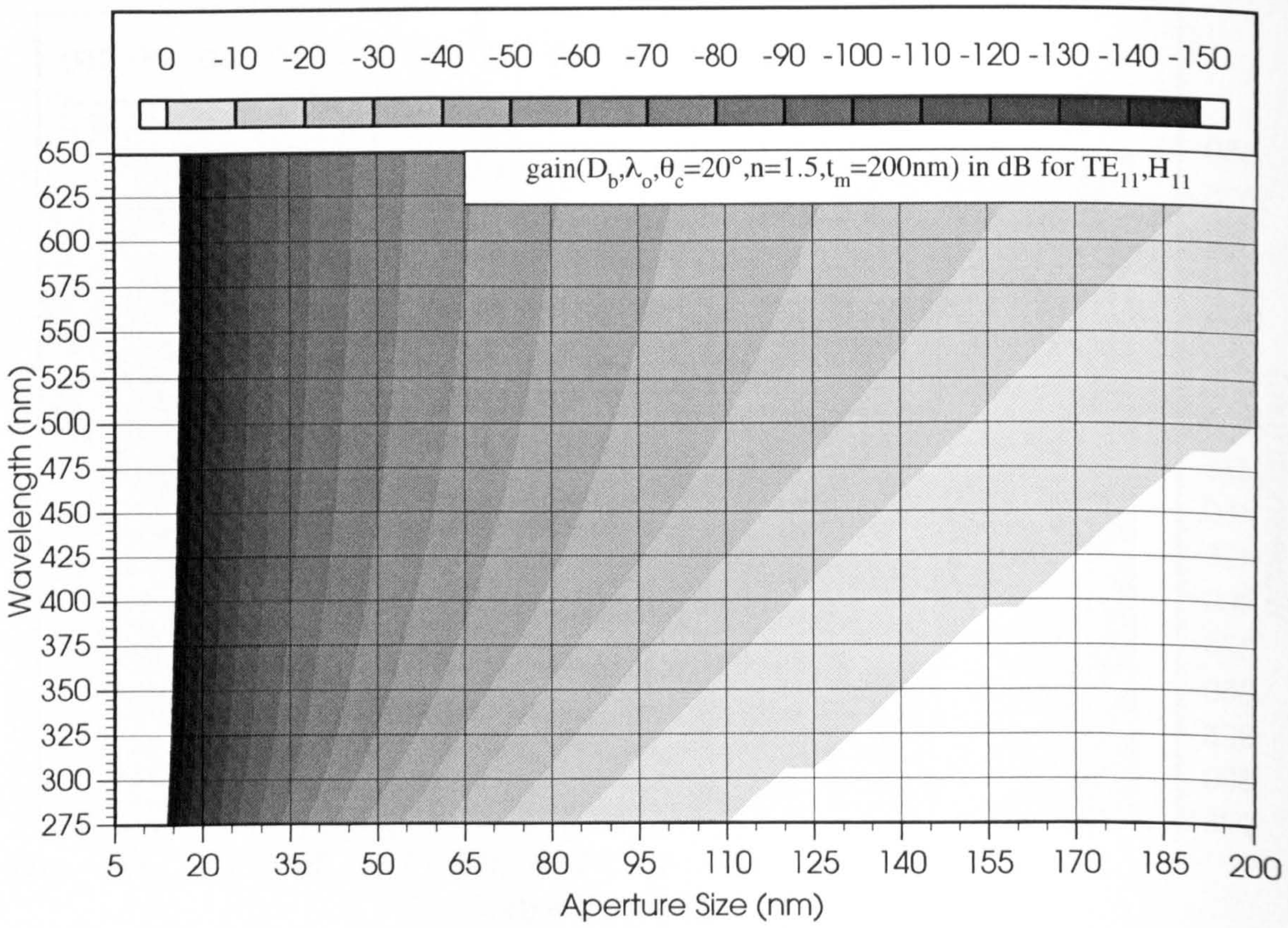


Figure 6.5.27: The gain in dB as a function of the vacuum wavelength and aperture size for the TE_{11} mode with parameters as given on the graph.

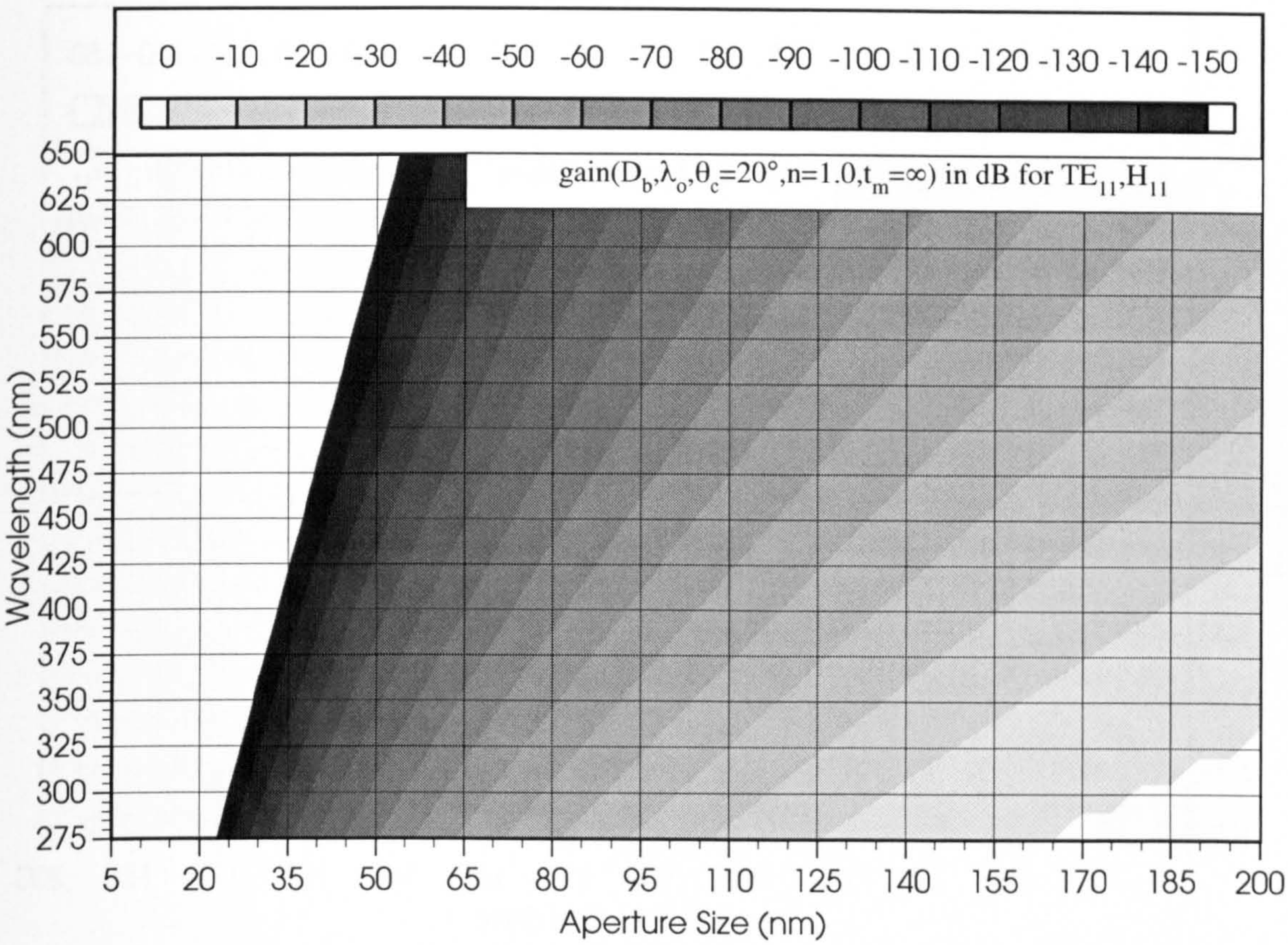


Figure 6.5.28: The gain in dB as a function of the vacuum wavelength and aperture size for the TE_{11} mode with parameters as given on the graph.

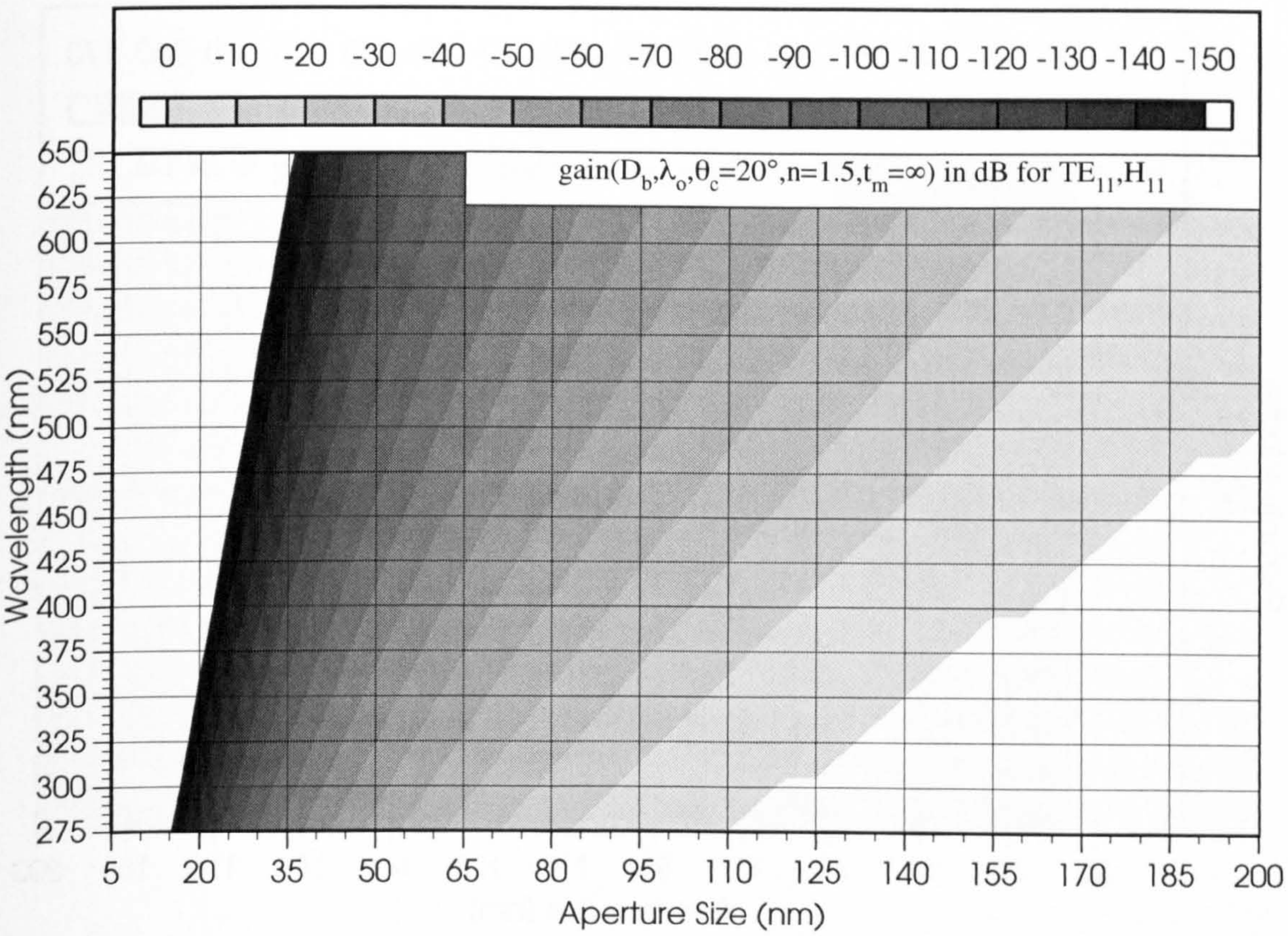


Figure 6.5.29: The gain in dB as a function of the vacuum wavelength and aperture size for the TE_{11} mode with parameters as given on the graph.

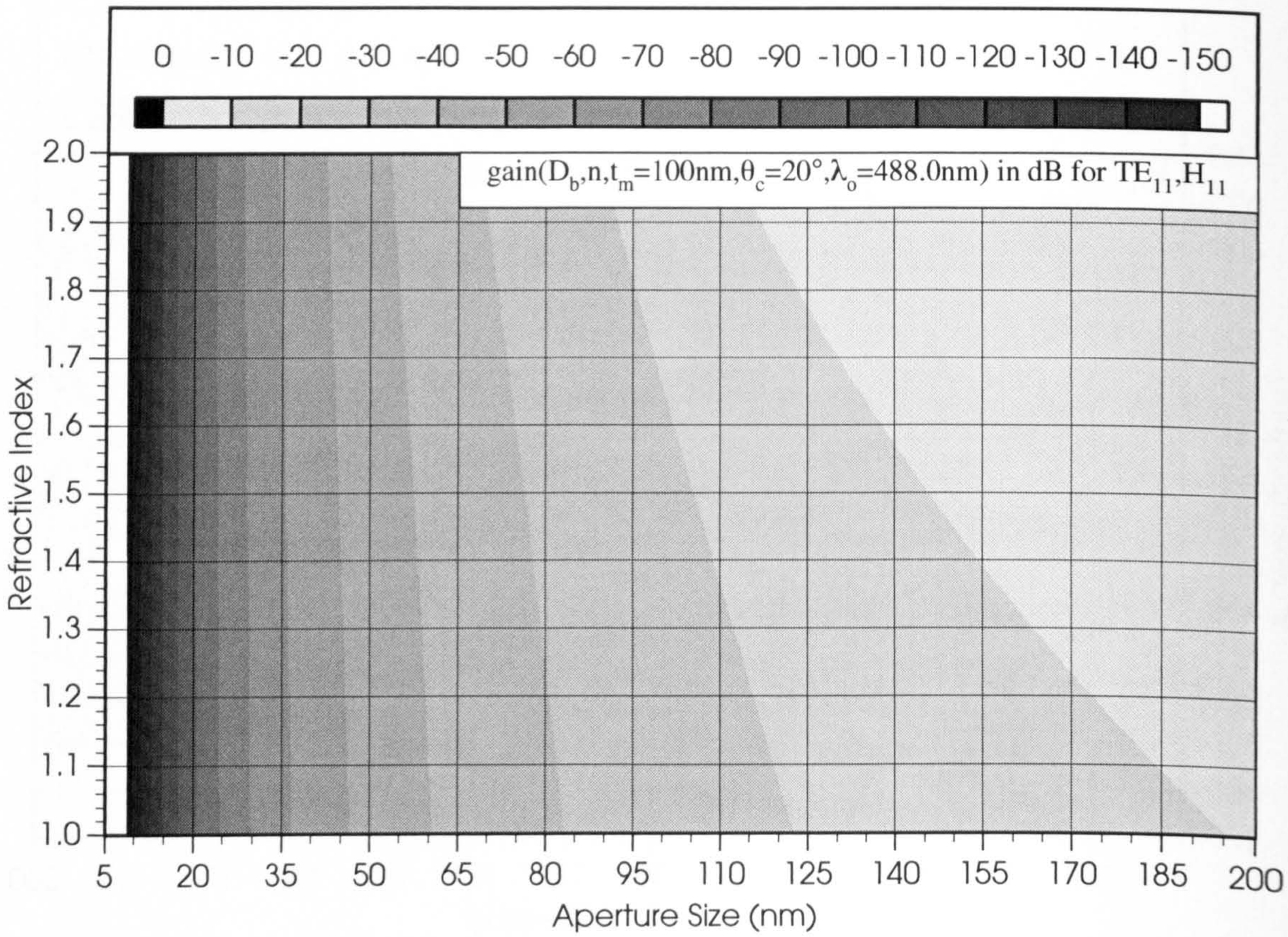


Figure 6.5.30: The gain in dB as a function of the refractive index and aperture size for the TE_{11} mode with parameters as given on the graph.

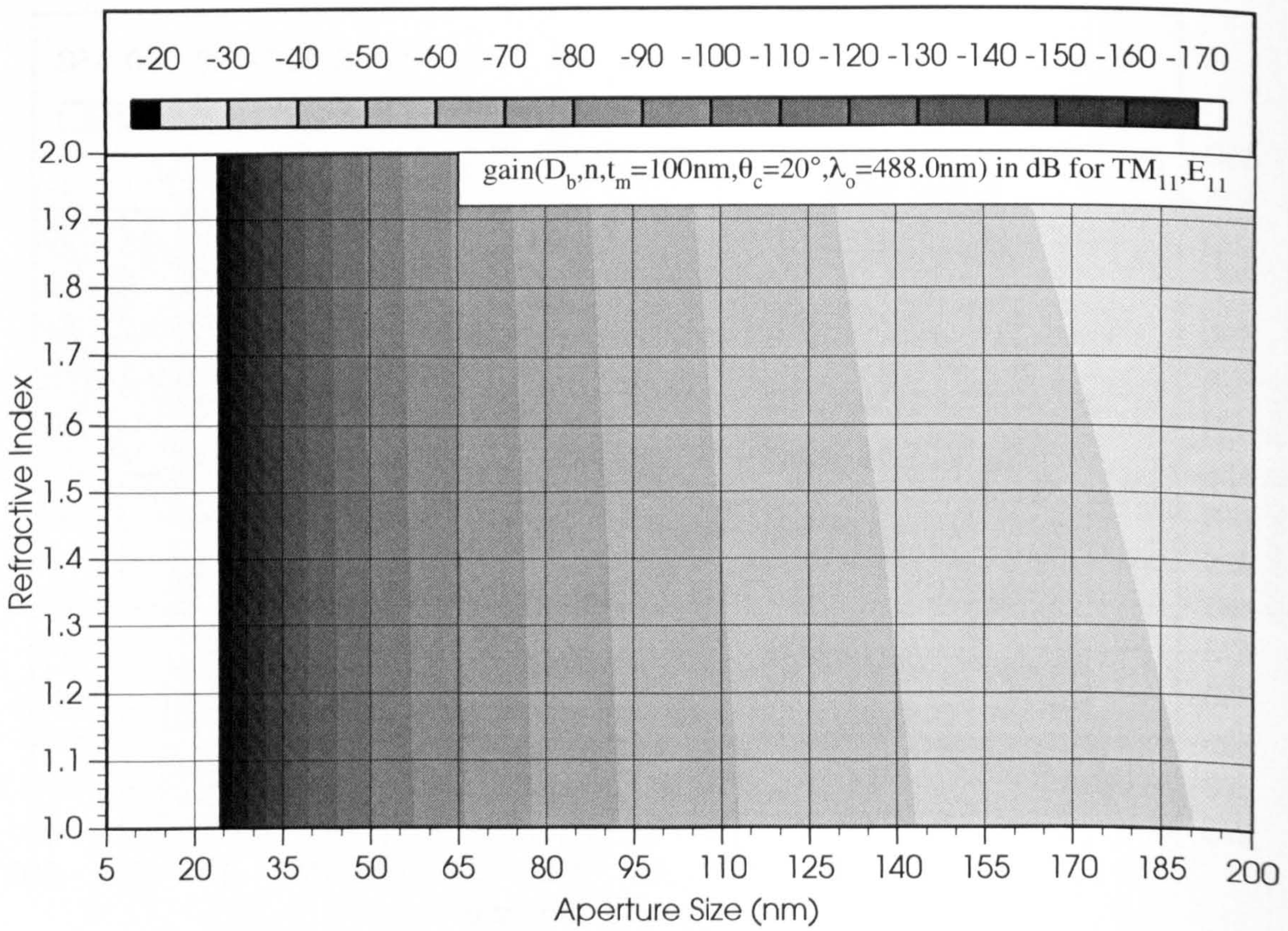


Figure 6.5.31: The gain in dB as a function of the refractive index and aperture size for the TM_{11} mode with parameters as given on the graph.

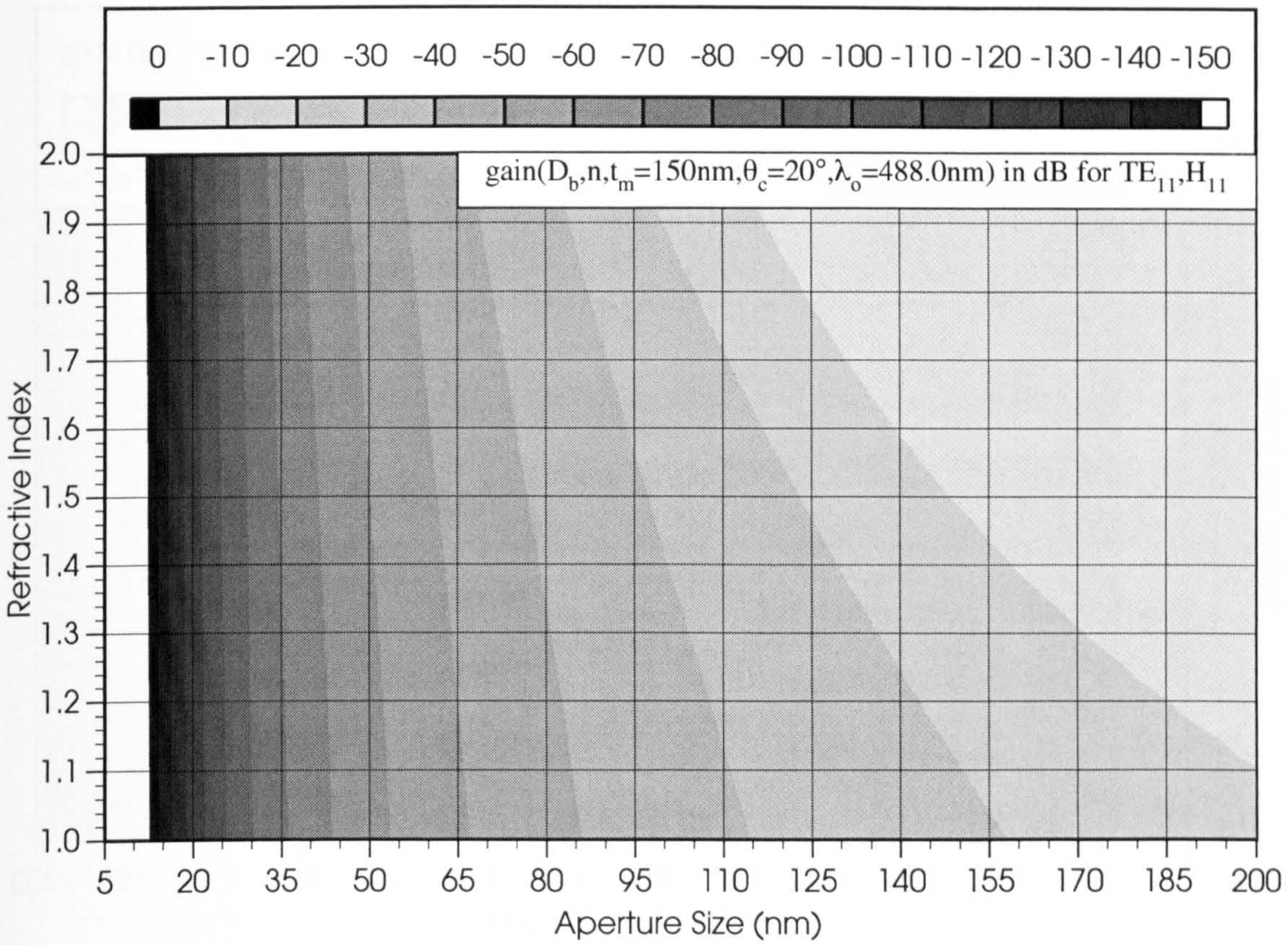


Figure 6.5.32: The gain in dB as a function of the refractive index and aperture size for the TE_{11} mode with parameters as given on the graph.

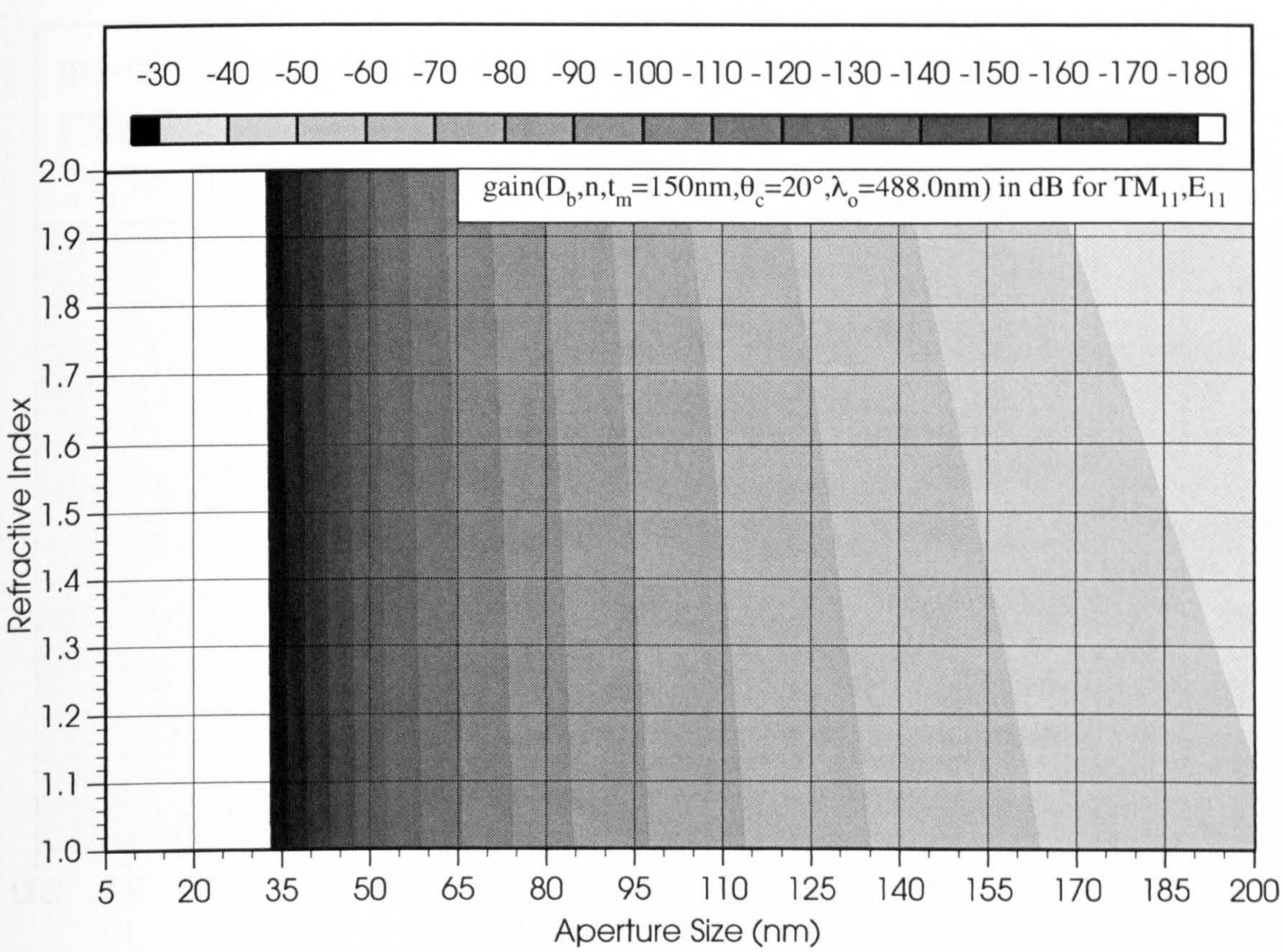


Figure 6.5.33: The gain in dB as a function of the refractive index and aperture size for the TM_{11} mode with parameters as given on the graph.

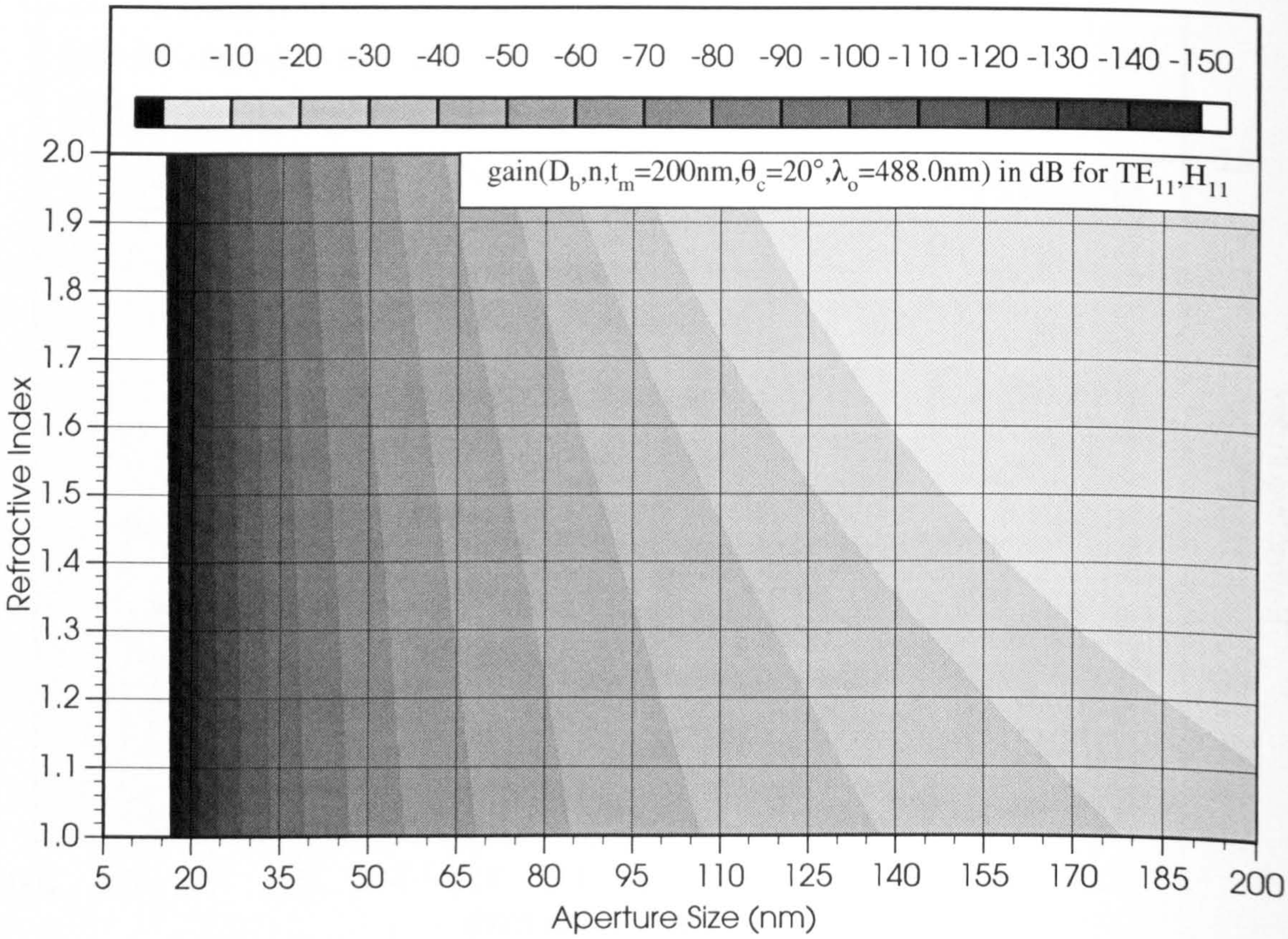


Figure 6.5.34: The gain in dB as a function of the refractive index and aperture size for the TE_{11} mode with parameters as given on the graph.

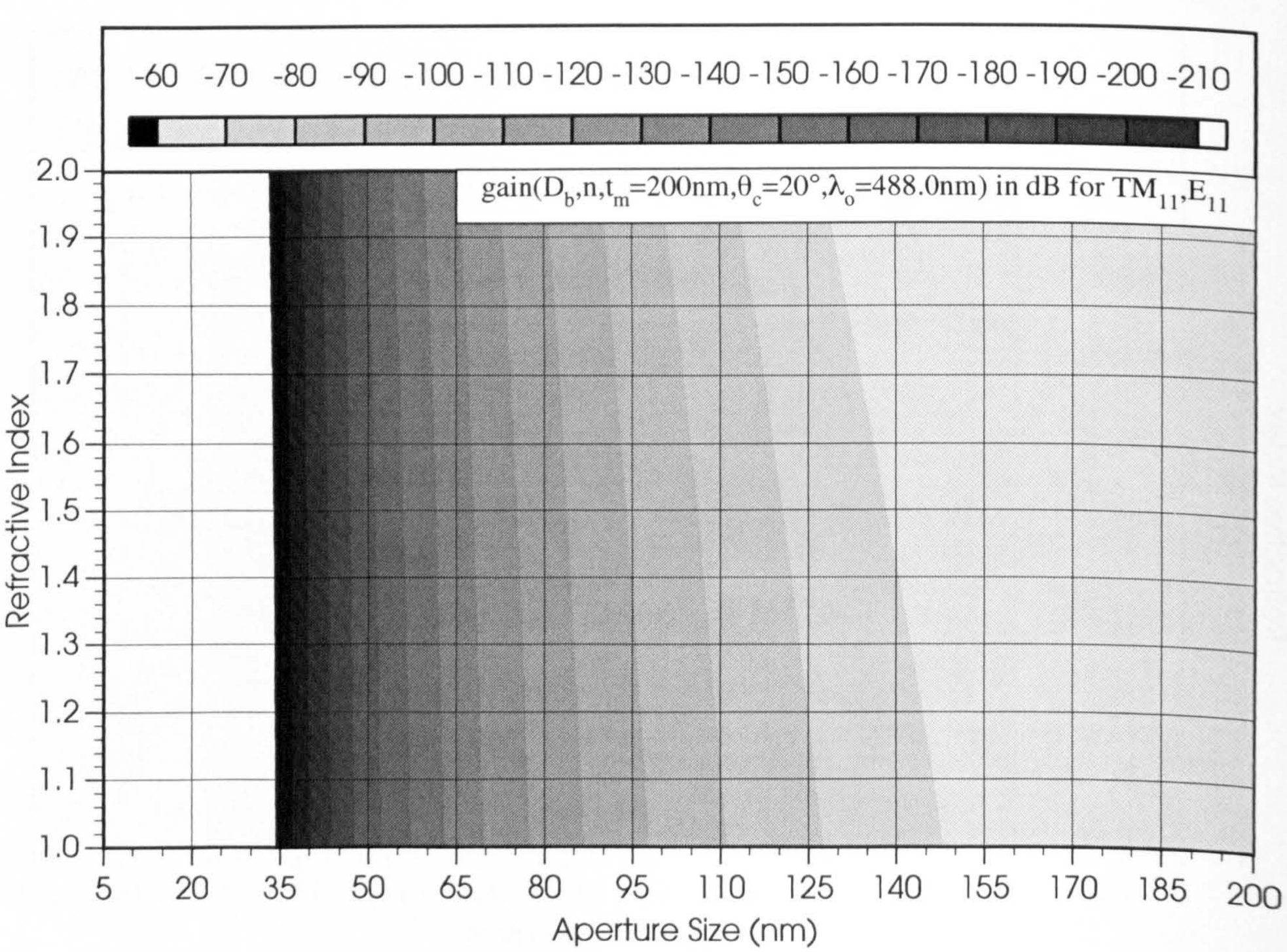


Figure 6.5.35: The gain in dB as a function of the refractive index and aperture size for the TM_{11} mode with parameters as given on the graph.

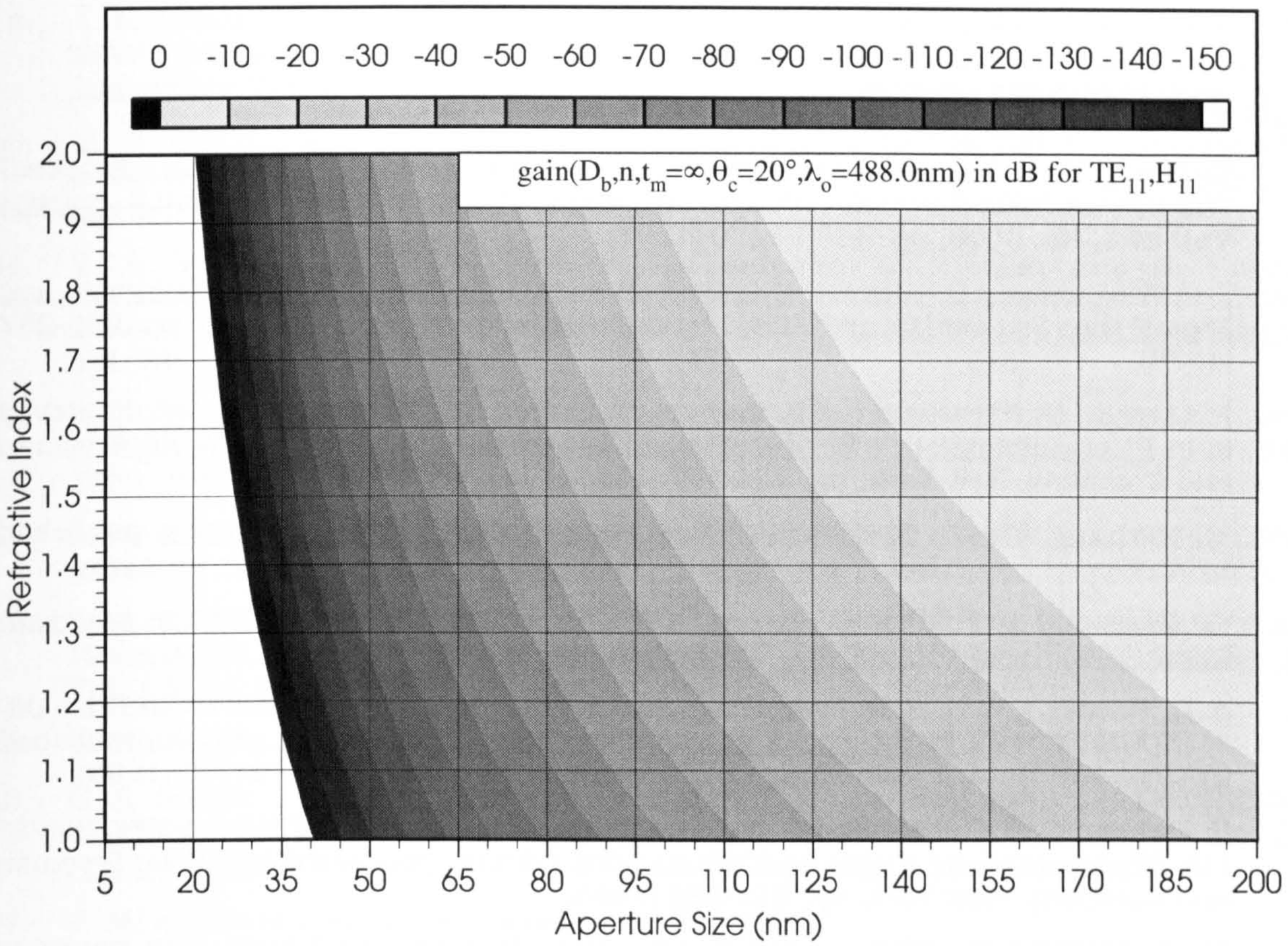


Figure 6.5.36: The gain in dB as a function of the refractive index and aperture size for the TE_{11} mode with parameters as given on the graph.

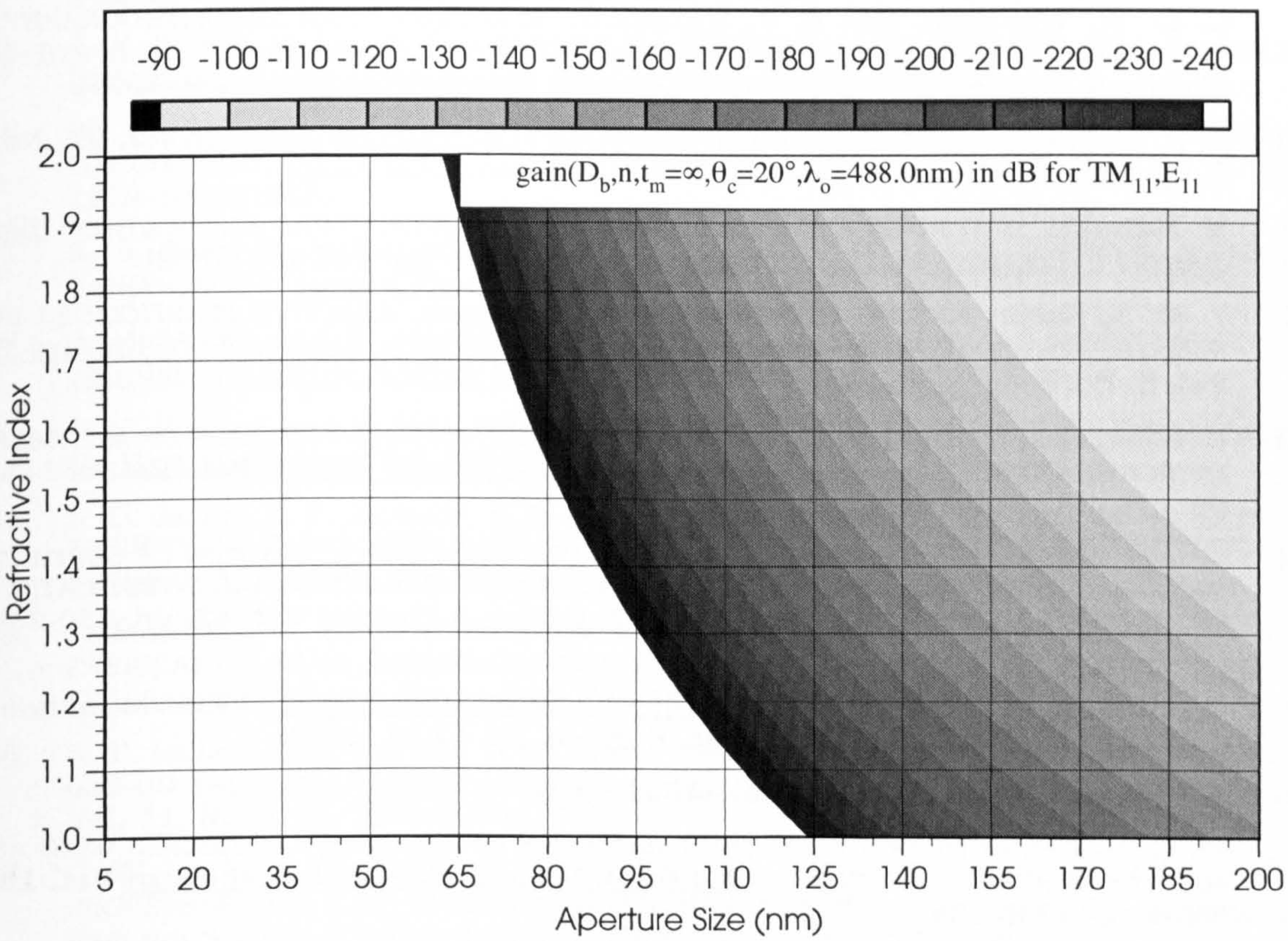


Figure 6.5.37: The gain in dB as a function of the refractive index and aperture size for the TM_{11} mode with parameters as given on the graph.

6.6 REFERENCES

1. E. Betzig, J. K. Trautman, T. D. Harris, J. S. Weiner and R. L. Kostelak, *BREAKING THE DIFFRACTION BARRIER - OPTICAL MICROSCOPY ON A NANOMETRIC SCALE*, Science **Vol. 251,** No. 5000, pp. 1468-1470 (1991).

2. B. I. Bleaney and B. Bleaney, *Chapter 8: ELECTROMAGNETIC WAVES* in Electricity and Magnetism, 3rd edition, published by Oxford University Press, Oxford, pp. 225-257 (1976).

3. P. Lorrain, D. P. Corson and F. Lorrain, *Chapter 30: PLANE ELECTROMAGNETIC WAVES III* in Electromagnetic Fields and Waves, 3rd edition, published by W. H. Freeman and Company, New York, pp. 554-577 (1988).

4. M. Garbuny, *Chapter 5: PHENOMENA OF PROPAGATION* in Optical Physics, published by Academic Press, New York, pp. 240-301 (1965).

5. G. Hass and L. Hadley, *Section 6g: OPTICAL PROPERTIES OF METALS* in American Institute of Physics Handbook, 3rd edition, pp. 118-123 (1972).

6. J. H. Weaver and H. P. R. Frederikse, *Section 12: OPTICAL PROPERTIES OF METALS AND SEMICONDUCTORS* in Handbook of Chemistry and Physics, 73rd edition, published by Chemical Rubber Publishing Co, Cleveland, Ohio, USA, pp. 109-132 (1996).

7. P. Lorrain, D. P. Corson and F. Lorrain, *Chapter 28: PLANE ELECTROMAGNETIC WAVES I* in Electromagnetic Fields and Waves, 3rd edition, published by W. H. Freeman and Company, New York, pp. 514-535 (1988).

8. M. A. McCord, A. Wagner and T. Donohue, *RESOLUTION LIMITS AND PROCESS LATITUDE OF X-RAY MASK FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2958-2963 (1993).

9. C. D. W. Wilkinson and S. P. Beaumont, *ELECTRON BEAM NANOLITHOGRAPHY*, published by Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1980-1993).

10. R. F. W. Pease, *ELECTRON BEAM LITHOGRAPHY*, Contemporary Physics **Vol. 22,** No. 3, pp. 265-290 (1981).

11. M. Hatzakis, *MATERIALS AND PROCESSES FOR MICROSTRUCTURE FABRICATION*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 441-453 (1988).

12. Y. M. N. Namaste, S. K. Obendorf and F. Rodriguez, *REACTIVE PLASTICIZERS IN NEGATIVE ELECTRON BEAM RESISTS*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2245-2249 (1988).

13. D. Kobayashi, C. Kim and H. Fujita, *PHOTORESIST-ASSISTED RELEASE OF MOVABLE MICROSTRUCTURES*, Japanese Journal of Applied Physics - Part 2 **Vol. 32,** No. 11A, pp. L1642-L1644 (1993).

14. R. Legtenberg, A. C. Tilmans, J. Elders and M. Elwenspoek, *STICTION OF SURFACE MICROMACHINED STRUCTURES AFTER RINSING AND DRYING: MODEL AND INVESTIGATION OF ADHESION MECHANISMS*, Sensors and Actuators A-Physical **Vol. 43,** pp. 230-238 (1994).

15. Goodfellow, *METAL DATA TABLE* in The Goodfellow Catalogue, 1993/94 edition, published by Goodfellow, Cambridge, England, pp. 17-31 (1993).

16. S. K. Murad, *Private Communication*, (1993-1996).

17. M. Cantagrel, *COMPARISON OF THE PROPERTIES OF DIFFERENT MATERIALS USED AS MASKS FOR ION-BEAM ETCHING*, Journal Of Vacuum Science & Technology **Vol. 12,** No. 6, pp. 1340-1343 (1975).

18. P. C. Karulkar and M. A. Wirzbicki, *CHARACTERIZATION OF ETCHING OF SILICON DIOXIDE AND PHOTORESIST IN A FLUOROCARBON PLASMA*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 5, pp. 1595-1599 (1988).

19. J. R. Lothian, F. Ren and S. J. Pearton, *MASK EROSION DURING DRY ETCHING OF DEEP FEATURES IN III-V SEMICONDUCTOR STRUCTURES*, Semiconductor Science & Technology **Vol. 7,** pp. 1199-1209 (1992).

20. M. A. McCord and R. F. W. Pease, *THE EFFECT OF REFLECTED AND SECONDARY ELECTRONS ON LITHOGRAPHY WITH THE SCANNING TUNNELING MICROSCOPE*, Surface Science **Vol. 181,** pp. 278-284 (1987).

21. V. A. Yunkin, V. F. Lukichev, K. V. Rudenko, D. Fischer and E. Voges, *EXPERIMENTAL STUDY AND COMPUTER SIMULATION OF ASPECT RATIO DEPENDENT EFFECTS OBSERVED IN SILICON REACTIVE ION ETCHING*, Microelectronic Engineering **Vol. 30,** pp. 345-348 (1996).

22. R. A. Gottscho, *ION TRANSPORT ANISOTROPY IN LOW PRESSURE, HIGH DENSITY PLASMAS*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 5, pp. 1884-1889 (1993).

23. J. M. F. Zachariasse and A. N. Broers, *THE DEPENDENCE OF NANOSTRUCTURE ETCH RATES ON FEATURE SIZE*, Microelectronic Engineering **Vol. 30,** pp. 349-352 (1996).

24. I. W. Rangelow and H. Loschner, *REACTIVE ION ETCHING FOR MICROELECTRICAL MECHANICAL SYSTEM FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 2394-2399 (1995).

25. S. Thoms, *Private Communication*, (1993-1996).

26. C. A. Spindt, I. Brodie, L. Humphrey and E. R. Westerburg, *PHYSICAL PROPERTIES OF THIN-FILM FIELD EMISSION CATHODES WITH MOLYBDENUM CONES*, Journal Of Applied Physics **Vol. 47,** No. 12, pp. 5248-5263 (1976).

27. J. M. R. Weaver, *Private Communication*, (1993-1996).

28. S. Thoms, S. P. Beaumont and C. D. W. Wilkinson, *A JEOL 100 CXII CONVERTED FOR USE AS AN ELECTRON-BEAM LITHOGRAPHY SYSTEM*, Journal Of Vacuum Science & Technology B **Vol. 7,** No. 6, pp. 1823-1826 (1989).

29. A. N. Broers, *RESOLUTION LIMITS FOR ELECTRON-BEAM LITHOGRAPHY*, IBM Journal of Research and Development **Vol. 32,** No. 4, pp. 502-513 (1988).

30. W. Liu, J. Ingino and R. F. Pease, *RESIST CHARGING IN ELECTRON BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 5, pp. 1979-1983 (1995).

31. K. S. Ralls, R. A. Buhrman and R. C. Tiberio, *FABRICATION OF THIN-FILM METAL NANOBRIDGES*, Applied Physics Letters **Vol. 55,** No. 23, pp. 2459-2461 (1989).

32. J. M. Moran and D. Mayden, *HIGH-RESOLUTION, STEEP PROFILE RESIST PATTERNS*, Journal Of Vacuum Science & Technology **Vol. 16,** p. 1620 (1980).

33. R. E. Howard, *MULTILEVEL RESIST LITHOGRAPHY BELOW 100 NM*, IEEE Transactions On Electron Devices **Vol. ED-28,** No. 11, pp. 1378-1381 (1981).

34. L. D. Jackel, R. E. Howard, E. L. Hu, D. M. Tennant and P. Grabbe, *50-NM SILICON STRUCTURES FABRICATED WITH TRILEVEL ELECTRON BEAM RESIST AND REACTIVE-ION ETCHING*, Applied Physics Letters **Vol. 39,** No. 3, pp. 268-270 (1981).

35. T. Matthies, C. David and J. Thieme, *TRILEVEL REACTIVE ION PROCESSES FOR FABRICATION OF 60 NM GERMANIUM STRUCTURES WITH HIGH ASPECT RATIO*, Journal Of Vacuum Science & Technology **Vol. 11,** No. 5, pp. 1873-1878 (1993).

36. K. P. Muller, N. K. Eib and T. B. Faure, *ETCHING ON SILICON MEMBRANES FOR SUB-0.25-UM X-RAY MASK MANUFACTURING*, Journal Of Vacuum Science & Technology B **Vol. 11,** No. 6, pp. 2270-2274 (1993).

37. C. W. Lo, M. J. Rooks, W. K. Lo, M. Isaacson and H. G. Craighead, *RESISTS AND PROCESSES FOR 1 KV ELECTRON BEAM MICROCOLUMN LITHOGRAPHY*, Journal Of Vacuum Science & Technology **Vol. 13,** No. 3, pp. 812-820 (1995).

38. M. B. Stern, S. C. Palmateer, M. W. Horn, M. Rothschild, B. E. Maxwell and J. E. Curtin, *PROFILE CONTROL IN DRY DEVELOPMENT OF HIGH-ASPECT-RATIO RESIST*

STRUCTURES, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 3017-3021 (1995).

39. P. Unger, V. Bogli, H. Beneking, B. Niemann and P. Guttman, *IMAGING ZONE PLATES FOR X-RAY MICROSCOPY FABRICATED BY ELECTRON-BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 1, pp. 323-327 (1988).

40. W. H. Juan and S. W. Pang, *HIGH ASPECT RATIO POLYIMIDE ETCHING USING AN OXYGEN PLASMA GENERATED BY ELECTRON CYCLOTRON RESONANCE SOURCE*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 1, pp. 422-426 (1994).

41. G. Schneider, T. Schliebe and H. Aschoff, *CROSS-LINKED POLYMERS FOR NANOFABRICATION OF HIGH-RESOLUTION ZONE PLATES IN NICKEL AND GERMANIUM*, Journal Of Vacuum Science & Technology B **Vol. 13,** No. 6, pp. 2809-2812 (1995).

42. W. Patrick, W. S. Mackie, S. P. Beaumont and C. D. W. Wilkinson, *POLYIMIDE AS A NEGATIVE RESIST AND ITS APPLICATION IN CROSSOVERS AND METAL ON POLYMER MASK FABRICATION*, Journal Of Vacuum Science & Technology B **Vol. 4,** No. 1, pp. 390-393 (1985).

43. R. Dagostino, F. Fracassi, C. Pacifico and P. Capezzuto, *PLASMA ETCHING OF TI IN FLUORINE-CONTAINING FEEDS*, Journal Of Applied Physics **Vol. 71,** No. 1, pp. 462-471 (1992).

44. R. Dagostino, F. Fracassi and C. Pacifico, *DRY ETCHING OF TI IN CHLORINE CONTAINING FEEDS*, Journal Of Applied Physics **Vol. 72,** No. 9, pp. 4351-4357 (1992).

45. G. Turban, *BASIC PHENOMENA IN REACTIVE LOW-PRESSURE PLASMAS USED FOR DEPOSITION AND ETCHING*, Pure & Applied Chemistry **Vol. 56,** No. 2, pp. 215-230 (1984).

46. S. G. Ingram, *THE INFLUENCE OF SUBSTRATE TOPOGRAPHY ON ION BOMBARDMENT IN PLASMA ETCHING*, Journal Of Applied Physics **Vol. 68,** No. 2, pp. 500-504 (1989).

47. T. Schliebe, G. Schneider and H. Aschoff, *NANOSTRUCTURING HIGH RESOLUTION PHASE ZONE PLATES IN NICKEL AND GERMANIUM USING CROSS-LINKED POLYMERS*, Microelectronic Engineering **Vol. 30,** pp. 513-516 (1996).

48. A. N. Broers, J. M. E. Harper and W. W. Molzen, *250 Å LINEWIDTHS WITH PMMA ELECTRON RESIST*, Applied Physics Letters **Vol. 33,** p. 392 (1978).

49. T. Tada and T. Kanayama, *FABRICATION OF SILICON NANOSTRUCTURES WITH A POLY(METHYLMETHACRYLATE) SINGLE LAYER PROCESS*, Journal Of Vacuum & Science Technology B **Vol. 13,** No. 6, pp. 2801-2804 (1995).

50. M. Yasuda, H. Kawata, K. Murata, K. Hashimoto, Y. Hirai and N. Nomura, *RESIST HEATING EFFECT IN ELECTRON BEAM LITHOGRAPHY*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 3, pp. 1362-1366 (1994).

51. J. O. Choi, J. A. Moore, J. C. Corelli, J. P. Silverman and H. Bakhru, *DEGRADATION OF POLY(METHYLMETHACRYLATE) BY DEEP ULTRAVIOLET, X-RAY, ELECTRON BEAM, AND PROTON BEAM IRRADIATIONS*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2286-2289 (1988).

52. N. Cameron, *Private Communication*, (1993-1996).

53. K. G. Chiong, S. Wind and D. Seeger, *EXPOSURE CHARACTERISTICS OF HIGH-RESOLUTION NEGATIVE RESISTS*, Journal Of Vacuum Science & Technology B **Vol. 8,** No. 6, pp. 1447-1453 (1990).

54. D. W. Keith, R. J. Soave and M. J. Rooks, *FREE-STANDING GRATINGS AND LENSES FOR ATOM OPTICS*, Journal Of Vacuum Science & Technology B **Vol. 9,** No. 6, pp. 2846-2850 (1991).

55. T. K. S. Wong and S. G. Ingram, *FABRICATION OF SUB-20 NM TRENCHES IN SILICON NITRIDE USING CHF_3/O_2 REACTIVE ION ETCHING AND OBLIQUE METALLIZATION*, Journal Of Vacuum Science & Technology B **Vol. 10,** No. 6, pp. 2393-2397 (1992).

56. A. Franciosi, D. J. Peterman and J. H. Weaver, *SILICON-REFRACTORY METAL INTERFACES: EVIDENCE OF ROOM-TEMPERATURE INTERMIXING FOR SI-CR*, Journal Of Vacuum Science & Technology **Vol. 19,** No. 3, pp. 657-660 (1981).

57. P. J. Grunthaner, F. J. Grunthaner, A. Madhuker and J. W. Mayer, *METAL/SILICON INTERFACE FORMATION: THE NI/SI AND PD/SI SYSTEMS*, Journal Of Vacuum Science & Technology **Vol. 19,** No. 3, pp. 649-656 (1981).

58. S. M. Sze, *Chapter 9: OXIDATION AND FILM DEPOSITION* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 341-380 (1985).

59. S. Thoms, *NOTES ON EBL*, published by Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1993-1996).

CHAPTER 7

THERMAL PROBE FABRICATION

This chapter describes the attempts made to design and fabricate probes for scanning thermal microscopy. First some theoretical and design issues specific to scanning thermal microscope probes are discussed. Then the problems encountered when trying to fabricate these probes and the measures taken in resolving these problems are described in some detail. This is followed by a description of the design and fabrication of a new kind of thermal probe requested as part of a collaboration with another university, and serves to highlight some of the advantages of the generic probe fabrication strategy being pursued. Finally, the performance and capabilities that might be expected of these thermal probes are discussed.

7.1 THEORETICAL CONSIDERATIONS

As mentioned in §3.2, the probe being designed for use in thermal microscopy applications consists of a contact mode AFM cantilever and pyramidal tip [§4.3.3] upon which a sub-micron thermocouple has been defined with the crossover of metal wires being coincident with the apex of the tip.

To understand the design issues involved in a thermocouple probe requires some understanding of the origin of the *Seebeck* effect by which the thermocouple signal is generated. This effect is one of several material properties known collectively as thermoelectric properties [Bleaney and Bleaney[1] (1976)]. Unfortunately a detailed understanding of thermoelectric properties includes a detailed understanding of electronic properties and as such is beyond the scope of this thesis. In any case, electronic properties are still not thoroughly understood and no adequate models exist to predict the electrical conductivity of different metals with reasonable accuracy. This is a basic prerequisite for calculating the values of thermoelectric coefficients for each metal. That this should be the case is not immediately obvious, however, it is fortunate that such a detailed understanding is not necessary in order to determine the main considerations for the device being developed here. For this reason, only a brief discussion of the origin of the Seebeck effect will be given here. Due to the complexity of the subject, the discussion here will attempt to enlighten at the expense of rigour.

Consider first two different metals, A and B, at absolute zero. The *Pauli Exclusion Principle* forbids more than one electron from occupying any state with the same set of quantum numbers [Sandin[2] (1989)]. As a consequence of this, at absolute zero the electrons will occupy all the lowest states in each metal up to the *Fermi* level of each metal, ${E_F}^A$ and ${E_F}^B$, respectively [Bleaney, *et al.*[1] (1976)]. These energies are normally measured with respect to the lowest energy states of each metal and cannot be compared numerically with one another. Another reference level is required before such a comparison can be made. Taking this reference to be the vacuum level, *i.e.* making the vacuum level the zero of electron energy, ${E_F}^A$ and ${E_F}^B$ can be placed below zero by the minimum amount of energy required to excite an electron to escape from the metal. This energy is also known as the *work function* of the metal [Bleaney, *et al.*[1] (1976)]. Letting the work functions of each metal be denoted by ϕ_A and ϕ_B respectively, the situation when both metals are brought together can be analysed. Note that the work function as used here ignores the contribution made by the image charge potential that an electron experiences outside a metal [Myers[3] (1990)]. Then:

$$ {E_F}^A + \phi_A = {E_F}^B + \phi_B = 0. \qquad (7.1.1) $$

The Fermi levels (and work functions) will equalise when both materials are brought together [Bleaney, *et al.*[1] (1976)]. This is so that there is no energy gradient between the

two metals, *i.e.* the energy of the highest occupied state in both materials will be the same. This is equivalent to saying that the new system can now be characterised by a single work function. The mechanism by which this equalisation takes place is the same as that in a *p-n* junction, *i.e.* electrons are transferred from the material with the highest Fermi energy to that with the lowest [Sze[4] (1985)]. Without loss of generality, it will be assumed that:

$$E_F^{\ A} > E_F^{\ B} \Leftrightarrow \phi_A < \phi_B. \tag{7.1.2}$$

Then the transfer of electrons from metal A to metal B results in metal B becoming negatively charged and metal A positively charged. These charges set up an electrostatic field which oppose further electron transfer. In equilibrium, equal numbers of electrons travel from A to B due to the chemical potential gradient as travel from B to A due to the electrostatic field. In semiconductor theory, the width of the region in each metal which is affected by charge transfer depends on the number density of donors and acceptors in the material. This region is known as the depletion region [Sze[4] (1985)]. In the situation depicted here, the number of 'donors' and 'acceptors' are equal (total charge neutrality), although the number densities may differ due to each metal having a different number density of states [Myers[5] (1990)]. However, these number densities will typically be several orders of magnitude greater than that for a semiconductor. Thus the 'depletion region' of the metal junction will be much shorter, and, by virtue of the fact that any electrostatic field inside each conductor would be destroyed (condition of high conductivity), the transferred charges essentially lie on the surfaces of the junction faces so that the situation is very close to that of an ideal double-sided abrupt junction, as illustrated by the top graph in Figure 7.1.1 [Sze[4] (1985)].

Assuming that this 'depletion region' is not infinitesimally small, the electrostatic field distribution can be calculated. Clearly any volume that completely encloses the metal junction contains zero net charge and so the electric flux in both metals far away from the junction must be equal by Gauss' law, *i.e.* the same number of electric field lines enter the volume as leave it [Bleaney and Bleaney[6] (1976)]. This 'background' field level can be taken as zero without loss of generality. However, any volume that encloses one half of the metal junction but not the other will contain a non-zero net charge and this shows that electric field lines must enter or leave this volume. This electric field does not pervade through both metals because it is screened by the field generated by the other half of the metal junction (in the same way as in a charged capacitor).

Since the electric field strength is proportional to the integral of the charge density with respect to position (Gauss' law in one dimension), taking a route from the free end of metal A to the free end of metal B, a qualitative electric field distribution can be derived: the electric field strength is zero from the free end of metal A to the beginning of the 'depletion region' in metal A, then increases until the centre of the junction (metal A is positively charged) and decreases again to zero from the centre of the junction to the end

of the 'depletion region' in metal B (metal B is equally negatively charged), from where it continues to be zero until reaching the free end of metal B. This is illustrated by the middle graph in Figure 7.1.1. Due to the very small width of the 'depletion region' the electric field distribution can be thought of as being close to a delta function centred at the junction.

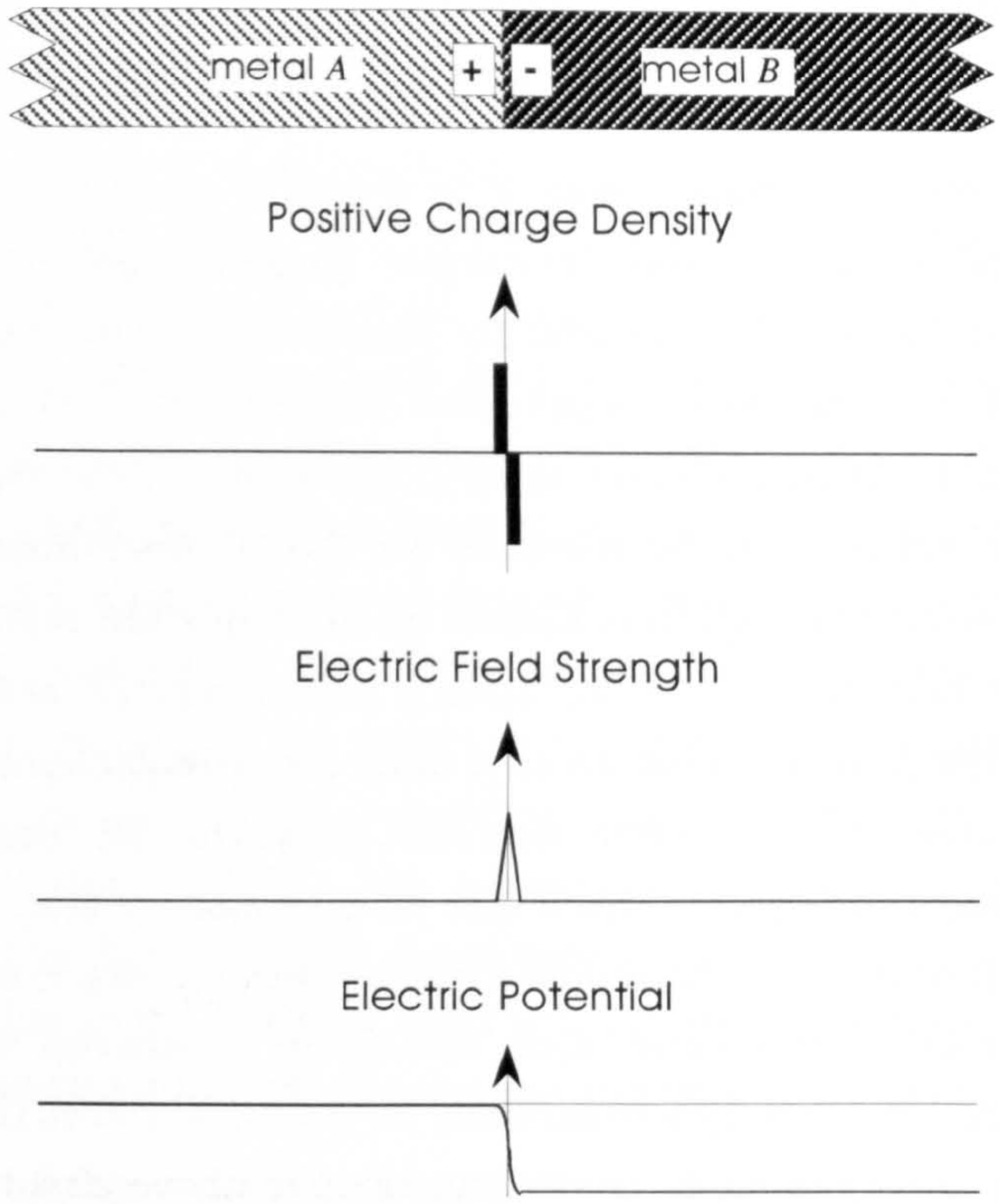


Figure 7.1.1: The effect of Fermi level equalisation at the junction of two metals. The representation shown here treats the system as an abrupt junction.

This electrostatic field distribution gives rise to an electric potential. The electric potential is the (negative of the) integral of the electric field distribution with position. Taking the free end of metal A as the reference for this potential, it is clear that the potential remains constant until reaching the beginning of the 'depletion region' in metal A, from where it decreases monotonically through the junction and until reaching the end of the 'depletion region' in metal B, after which it remains at this new value until reaching the free end of metal B. This is illustrated by the bottom graph in Figure 7.1.1. [Note that the integral of a delta function is a step function.]

The preceding argument clearly shows that a potential difference is developed between the two metals after they have been brought into contact. This potential difference is known as the *contact potential difference* of the two metals in question and has a definite sign [Bleaney, *et al.*[1] (1976)]. The magnitude of this potential is equal to the difference between the Fermi energies (or work functions) of the two metals divided by the electronic charge:

$$q_e V_{BA} = \phi_B - \phi_A, \tag{7.1.3}$$

where q_e is the (magnitude of the) electronic charge and V_{BA} is the contact potential difference measured from B to A. For the example above, metal B is at a lower electric potential (higher electron potential) than metal A. This electric potential opposes the chemical potential gradient between the two metals, resulting in zero net energy gradient as required for the condition of equilibrium. The reason for describing the Fermi equalisation process in such great detail will become apparent in what follows.

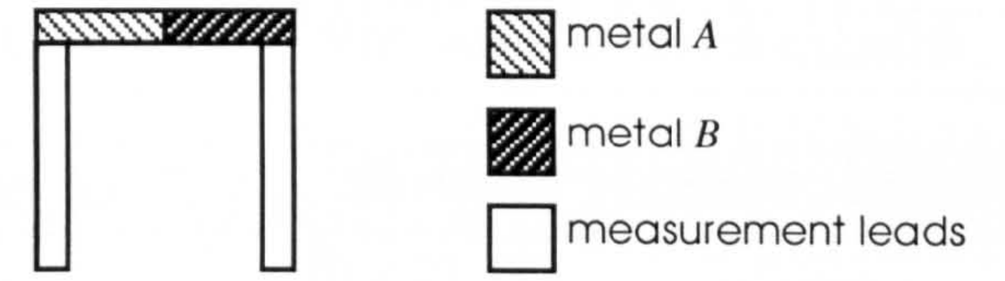


Figure 7.1.2: Measurement of the contact potential difference between two metals.

Consider now the measurement of such a contact potential difference at absolute zero as illustrated in Figure 7.1.2. Assume that the work function of each measurement lead is ϕ_m. Then the potential measured in the direction from B to A is:

$$\Delta V = V_{mB} + V_{BA} + V_{Am}, \tag{7.1.4}$$

where $V_{mB} = \left(\phi_m - \phi_B\right)/q_e$ is the contact potential difference between the first measurement lead and metal B and $V_{Am} = \left(\phi_A - \phi_m\right)/q_e$ is the contact potential difference between metal A and the other measurement lead. Expanding Equation 7.1.4:

$$q_e\Delta V = \left(\phi_m - \phi_B\right) + \left(\phi_B - \phi_A\right) + \left(\phi_A - \phi_m\right) = 0, \tag{7.1.5}$$

from which it is clear that the contact potential difference cannot be measured directly.

Now consider the same measurement at a temperature above absolute zero. Since the Fermi levels (and work functions) are a function of temperature [Gardner[7] (1994)], the contact potential difference must also be a function of temperature. However, under isothermal conditions, Equation 7.1.5 is still valid as each contribution cancels itself out.

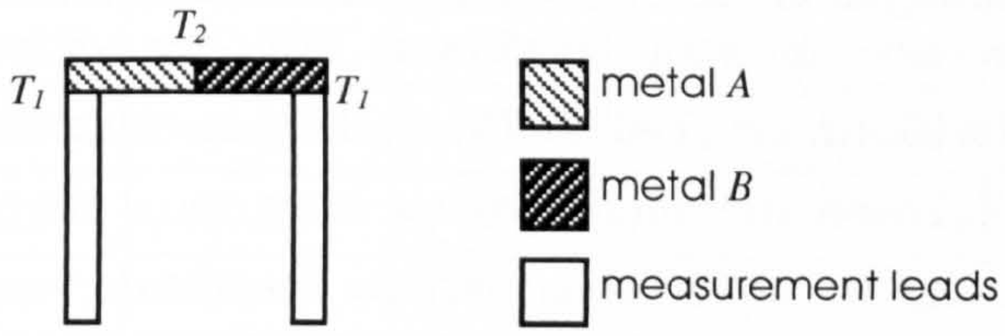


Figure 7.1.3: Measurement of the contact potential difference between two metals.

Finally consider the measurement illustrated in Figure 7.1.3 where the junction between metal A and B is at a temperature, T_2, and the junctions with the measurement leads are both at temperature, T_1. Then the measured potential difference is given by:

$$\begin{aligned} q_e\Delta V &= \left[\phi_m(T_1) - \phi_B(T_1)\right] + \left[\phi_B(T_2) - \phi_A(T_2)\right] + \left[\phi_A(T_1) - \phi_m(T_1)\right] \\ &= \left[\phi_B(T_2) - \phi_A(T_2)\right] - \left[\phi_B(T_1) - \phi_A(T_1)\right] = q_e\left[V_{BA}(T_2) - V_{BA}(T_1)\right] \neq 0 \end{aligned} \tag{7.1.6}$$

Notice how the effects of the measurement leads cancel out. This is always true if both measurement lead junctions are at the same temperature. Then defining:

$$\Delta V_{BA}(T_2, T_1) = V_{BA}(T_2) - V_{BA}(T_1), \tag{7.1.7}$$

the quantity, $\Delta V_{BA}(T_2, T_1)$, is known as the Seebeck potential. The Seebeck coefficient (also known as the thermopower) for a thermocouple junction of materials A and B measured from B to A is then defined by [MacDonald[8] (1962)]:

$$S_{BA}(T) = \lim_{\Delta T \to 0} \frac{\Delta V_{BA}(T + \Delta T, T)}{\Delta T} = \frac{dV_{BA}}{dT}, \tag{7.1.8}$$

so that:

$$\Delta V_{BA}(T_1 + \Delta T, T_1) \approx S_{BA}(T_1)\Delta T. \tag{7.1.9}$$

Material	Seebeck Coefficient	Material	Seebeck Coefficient
Antimony	+50.7	Pt-10% Rh	+7.20
Chromel	+29.8	Aluminium	+5.30
Iron	+17.7	Tantalum	+4.65
Molybdenum	+16.0	Platinum	0.00
Tungsten	+13.1	Calcium	-2.55
Cadmium	+11.8	Palladium	-6.15
Gold	+9.20	Alumel	-10.85
Copper	+9.15	Cobalt	-15.40
Silver	+8.85	Nickel	-15.50
Rhodium	+8.05	Constantan	-37.25
Pt-13% Rh	+7.35	Bismuth	-67.85

Table 7.1.1: The Seebeck coefficients or thermopowers of various metals and alloys in μV/K at 200°C with reference to platinum at 0°C from Gardner[7] (1994).

It is this potential which is measured in the normal operation of a thermocouple. For small temperature changes, the Seebeck coefficient can be taken as constant, however, the Seebeck coefficient is generally a function of temperature and so the accuracy of any extrapolated values decrease for large deviations from the temperature at which the coefficient was measured [Gardner[7] (1994)]. The Seebeck coefficients for several materials are shown in Table 7.1.1 and the compositions of various alloys commonly used in thermocouples is tabulated in Table 7.1.2. Various standards for thermocouples exist. The UK standard metal/alloy thermocouples are described by BS 4937 [Gardner[7] (1994)]. Table 7.1.3 tabulates some of these. By rearranging Equation 7.1.6:

$$\begin{aligned} q_e \Delta V_{BA} &= \left[\phi_B(T_2) - \phi_B(T_1)\right] - \left[\phi_A(T_2) - \phi_A(T_1)\right] \\ &\approx \left[\left.\frac{d\phi_B}{dT}\right|_{T_1} - \left.\frac{d\phi_A}{dT}\right|_{T_1}\right]\Delta T, T_2 = T_1 + \Delta T, \Delta T \text{ small} \end{aligned}, \tag{7.1.10}$$

it can be seen that it is possible to define the thermopower (or Seebeck coefficient) for a single material according to:

$$S_B(T) = \frac{dV_B}{dT}, V_B(T) = \frac{\phi_B(T)}{q_e}. \tag{7.1.11}$$

Then it is also clear that:

$$S_{BA}(T) = S_B(T) - S_A(T). \tag{7.1.12}$$

Alloy	Composition
Chromel	90% Nickel, 10% Chromium
Constantan	55% Copper, 45% Nickel
Alumel	95% Nickel, 2% Aluminium, 2% Manganese, 1% Silicon
Nicrosil	71-86% Nickel, 14% Chromium, 0-15% Iron
Nisil	mainly 95% Nickel, 4.5% Silicon
Manganin	86% Copper, 12% Manganese, 2% Nickel

Table 7.1.2: The elemental composition of various alloys used in thermocouples.

So far no consideration has been given to the effect of scaling down the dimensions of these thermocouples. At first glance there does not appear to be any particular problem. If the metal wire junction is used in a standard thermocouple arrangement, then the measurement is one of potential difference using a voltage measuring circuit which is nominally of a very high impedance. Thus negligible current flows and the circuit is analogous to that of measuring the potential across a capacitor or battery. It then appears that the dimensions of each metal used to form the thermocouple has no effect on the measured potential difference provided that the resistance of the measuring circuit much exceeds that of the leads, thermocouple junction and lead junctions (just as the dimensions of the leads used to measure the potential across a battery or capacitor do not affect the measurement under similar circumstances).

Type	Min. Temp.	Max. Temp.	Accuracy (Class 2)	Temp. Range (Class 2)	Positive Arm	Negative Arm
J	-40°C	+850°C	±2.5°C	(-40,+750)°C	Iron	41% Ni, 55% Al
K	-200°C	+1100°C	±2.5°C	(-40,+750)°C	Chromel	Alumel
N	-230°C	+1230°C	±2.5°C	(-40,+750)°C	Nicrosil	Nisil
T	-250°C	+400°C	±1°C	(-40,+750)°C	Copper	Constantan
R	-50°C	+1350°C	±2°C	(-15,+1760)°C	Platinum	87% Pt, 13% Rh

Table 7.1.3: The specifications of selected standard UK thermocouples according to BS 4937.

However, from the preceding discussion, if the dimensions of any of the thermocouple constituents becomes smaller than the 'depletion region' of the junction, then the electric field distribution would be perturbed and alter the measured potential difference (*i.e.* the Seebeck coefficient would be affected). Since the 'depletion region' of a metal junction would seem to be on an atomic or near atomic scale, it would then appear that the thermocouple would work in exactly the same way down to comparable dimensions. This has been demonstrated using the STM [Weaver, Walpita and Wickramasinghe[9] (1989)].

Nevertheless, while it is the case that the thermocouple should work down to near-atomic scale, the Seebeck coefficient does not remain the same down to such scales because of *localisation effects* [Cutler and Mott[10] (1969)] and because electron transport properties also contribute to the measured Seebeck coefficient [MacDonald[8] (1962); Barnard[11] (1972)]. The first manifestation of localisation effects will appear when the dimensions of any of the conductors approach the *mean free path* of electrons in the conductor. The mean free path is defined as the average distance electrons can travel in the conductor before being scattered [Sze[12] (1985)]. As dimensions approach this length, the electron wave functions become 'aware' of the walls of the conductor. These localisation effects mean that the metal no longer behaves as a semi-infinite lattice since the available states are gradually reduced to wave functions with shorter wavelengths, as determined by the dimensions of the conductor in each direction. This then modifies the spectral density of states and raises the average state energy in the more confined regions of the conductor. In one dimension this situation would be analogous to treating the conductor as a potential well: as the width is decreased, the wavelengths of the allowable states decrease and their energies increase [Sandin[13] (1989)].

The way the density of state function affects the Seebeck coefficient is through its effect on the first derivative of the Fermi energy with temperature. This can be seen from the following argument. As a consequence of the Pauli Exclusion Principle, the probability of an electron occupying any particular state of energy, E, at any temperature, T, is given by the *Fermi-Dirac distribution function*, $f\left(E,E_F,T\right)$, defined by [Sandin[14] (1989)]:

$$f\left(E,E_F,T\right)=\frac{1}{1+\exp\left[\dfrac{E-E_F}{k_BT}\right]}, \tag{7.1.13}$$

where k_B is the Boltzmann constant and E_F is the Fermi energy, defined as the energy at which the probability of occupancy is exactly 0.5 (or 50%). This function always has a value between 0 and 1 and has the property that the probability of a state ΔE above the Fermi energy being occupied is equal to the probability of a state ΔE below the Fermi energy not being occupied.

At absolute zero all the states up to the Fermi energy have a 100% probability of occupancy and all those above have a zero probability of occupancy. For temperatures above absolute zero the electrons have energies of the order of k_BT available to occupy states above the Fermi energy, leaving some of those below unoccupied [Sandin[14] (1989)]. Table 7.1.4 lists some elements and their work functions (as derived from photoelectric effect data). It can be seen that these are very large in comparison to the value of k_BT at room temperature (25.8meV at 300K). However the occupancy function is not symmetric about the Fermi energy so that the Fermi energy does not remain at a fixed value independent of temperature [Ziman[15] (1972)]. Also, because the number of states with a particular energy tends to vary with energy and is not symmetric about the Fermi energy

(at absolute zero), the Fermi energy must itself be a function of temperature [Sandin[14] (1989)].

Element	Work Function	Element	Work Function	Element	Work Function
Antimony	4.1eV	Iron	4.4eV	Rhodium	4.6eV
Bismuth	4.4eV	Lead	4.0eV	Rubidium	2.1eV
Cadmium	4.0eV	Lithium	2.4eV	Ruthenium	4.71eV
Calcium	2.87eV	Magnesium	3.66eV	Silicon	4.2eV
Carbon	4.8eV	Manganese	3.8eV	Silver	4.7eV
Chromium	4.4eV	Molybdenum	4.2eV	Strontium	2.59eV
Cobalt	5.0eV	Nickel	4.9eV	Tantalum	4.1eV
Copper	4.5eV	Niobium	4.3eV	Tin	4.3eV
Gallium	4.2eV	Palladium	5.0eV	Titanium	4.1eV
Germanium	4.8eV	Platinum	5.3eV	Tungsten	4.55eV
Gold	4.8eV	Potassium	2.2eV	Vanadium	4.3eV
Iridium	4.6eV	Rhenium	5.0eV	Zinc	4.3eV

Table 7.1.4: The work function for selected elements as derived from photoelectric effect data. As such these include a small contribution from the image charge potentials generated outside the conductor by an escaping electron. Data from Goodfellow[16] (1993).

In general, at higher energies, the electron wavefunctions have a shorter wavelength and higher spatial frequency and are therefore more localised. This means there tends to be more states at higher energies than lower energies, so that if the total number of electrons is constant, an asymmetry is introduced which generally results in the Fermi energy decreasing with increasing temperature [Ziman[15] (1972)]. This lowering of the Fermi energy alters the contact potential developed between that metal and any other metal. As this applies to both metals forming the thermocouple, the contribution to the Seebeck effect arises from the differences in the rate of change of these Fermi energies with respect to temperature (and the reference level). This means that although the sign of the contact potential difference may be from one metal to another, by virtue of the fact that the difference in the rates of change of the Fermi energies of the two metals are not dependent on the values of the Fermi energy in those metals, the sign of the Seebeck coefficient could be in the opposite sense to that of the contact potential difference [Bleaney, *et al.*[1] (1976)]. The exact magnitude of the effect will depend on the precise variation and degeneracy in the density of states with energy [Ziman[15] (1972)].

However, this change of the Fermi energy with temperature is only one contribution to the observed Seebeck potential. It is also the case that even when observing a single conductor, a potential difference should arise between one end of the conductor and another if the two ends are held at different temperatures [MacDonald[8] (1962)]. If a contactless measurement of this potential difference could be made, then its value would

correspond to that given by using Equation 7.1.9 but with the absolute thermopower of the single metal alone. This is despite the fact that a contactless measurement will not result in the formation of a contact potential because as stated above, the Seebeck potential is (partly) due to the change in the contact potential rather than the value of the contact potential difference itself. In this case the potential arises in a similar manner to the formation of the contact potential. Assuming that the Fermi energy is reduced at the hot end of the metal wire, Fermi equalisation once again takes place between the hot and cold ends of the metal. Electrons from the cold end travel to the hot end and produce an electric potential which opposes any further flow of electrons. At equilibrium, the chemical potential gradient is exactly compensated for by the electric potential gradient as before [Barnard[11] (1972)].

However, as stated earlier, this is only part of the contribution to the Seebeck potential. The other part of the contribution is due to the flow of thermal energy from the hot end to the cold end of the conductor, the so-called *heat current*. This contribution alters the Seebeck potential from that solely due to the changes in contact potential difference with temperature.

Consider the steady state situation described above. Although this situation has been described as being in a steady state, it clearly cannot be in equilibrium while a temperature gradient exists. The steady state only applies to the movement of electronic charge and not to the movement of energy. Thus in any section of this conductor, the number of electrons per second entering the section from the cold end is equivalent to the number of electrons per second entering from the hot end. This means that no net electrical current flows, and so no further accumulation of charge at either end takes place (this being the integral of current with time). However, there has to be a net energy flow from the hot end towards the cold end. This energy flow takes place by several processes, but the one with most bearing on the Seebeck potential is that energy carried by the electrons themselves [MacDonald[8] (1962)].

At the hot end of the metal, conduction electrons tend to occupy higher energy states in comparison to the conduction electrons at the cold end. Electrons travelling towards the cold end absorb energy so that they remain in equilibrium with the temperature of the surrounding metal lattice. They absorb this energy through scattering mechanisms, and gradually become promoted to higher energy states as they approach the hot end of the metal [Ziman[15] (1972)]. Similarly, electrons travelling towards the cold end lose energy, again through scattering mechanisms, gradually becoming demoted to lower energies as they approach the cold end of the metal. This cycle persists until the entire conductor reaches the same temperature, and results in no net charge transfer.

However, it is also the case that the electrons travelling from the hot end to the cold end must experience different scattering and energy loss mechanisms to those travelling in the reverse direction, *i.e.* electrons travelling one way experience a different resistance to

those travelling the other way [Barnard[11] (1972)]. Thus an electric potential must appear across any length of the conductor corresponding to this difference in resistance (Ohm's law) since the electric current flowing in each direction is the same. This resistance can be viewed as being dependent on the rate of change of the mean free path with the energy of the electron state. The sign of this electric potential will depend on whether higher energy electrons have a shorter mean free path or a longer mean free path and could enhance the Seebeck potential or retard it, in some cases this effect is so large that it can cause the sign of the potential to reverse so that the Fermi energy of the cold end is reduced and charge accumulates at the cold end rather than the hot end [see Barnard[11] (1972)]. The mean free path is itself a function of temperature and is not well described by theory for individual metals, let alone its value as a function of electron state and energy [Barnard[11] (1972)]. Thus a detailed understanding of the Seebeck potential is extremely complex and is beyond the scope of this thesis. A complete description must include mechanisms such as the transfer of thermal energy by phonons and the effects of temperature on the rate of phonon-phonon and electron-phonon scattering [Jonson and Mahan[17] (1990); Durczewski and Ausloos[18] (19996)]. [For a more rigorous treatment see, for example, MacDonald[19] (1962); Barnard[20] (1972).]

From the above discussion, it would seem that the Seebeck potential starts to become modified from its bulk value when the dimensions of the metals approach that of the mean free path of conduction electrons in those metals. Therefore as long as this is not the case it can be expected that the bulk value of the Seebeck coefficient is valid. The mean free path of electrons in a metal is not easily calculable from first principles, however a simple model of resistivity combined with the free electron gas model of a metal can be used to estimate its value [Myers[5] (1990)]. In this model the mean free path, l_c, is given by:

$$l_c = v_{th}\tau_c, \tag{7.1.14}$$

where v_{th} is the thermal drift velocity of electrons in the metal and τ_c is the mean time between collisions (sometimes known as the *relaxation time*). The thermal drift velocity at a temperature, T, can be estimated from:

$$\frac{1}{2}m_e v_{th}^{\ 2} = \frac{3}{2}k_B T, \tag{7.1.15}$$

where m_e is the mass of the electron. The mean time between collisions can be calculated from [Myers[5] (1990)]:

$$\sigma = \frac{n q_e^{\ 2}\tau_c}{m_e} \tag{7.1.16}$$

where σ is the electrical conductivity of the metal and n is the number density of conduction electrons in the metal. The number density of conduction electrons in the metal can be easily determined from tabulated data if the number of valence electrons of the metal is well defined and it is assumed that each atom contributes its valence electron to the conduction band. Table 7.1.5 shows the relevant information and the

value of the mean free path as calculated for aluminium. This shows that the mean free path is less than 1nm at room temperature. Thus based on the resolution limits of electron-beam lithography as carried out for this thesis, it is unlikely that any deviation of the Seebeck coefficient from the bulk value will be seen, even when the smallest thermocouple that can be fabricated is used.

Element	Aluminium
Atomic Number	13
Atomic Mass	26.98 amu
Mass Density	2700 kg/m^3
Resistivity at 293K	2.67 x 10$^{-8}\Omega$m
Number of Valence Electrons	3
Conduction Electron Density	1.81 x 10^{29}/m^3
Relaxation Time	7.35fs
Thermal Drift Velocity at 300K	117km/s
Mean Free Path at ~300K	0.86nm

Table 7.1.5: Data used to estimate the mean free path of conduction electrons in aluminium at 300K. Data extracted from Goodfellow[16] (1993).

Additionally it is worth noting that the best choice of thermocouple metal should have a large Seebeck coefficient (large driving voltage), low resistance (so that most of the Seebeck potential is dropped across the 'voltmeter') and low thermal conductivity (so that the thermocouple does not reduce any temperature difference across it).

Thus in summary, theoretical considerations suggest that provided all the dimensions of the fabricated thin-film thermocouple much exceed the mean free path of electrons in each metal, which is of the order of 1nm, the observed thermocouple signal should be as obtained using bulk metals.

7.2 FABRICATION OF THERMOCOUPLE PROBES

Initial development concentrated on evaluating the difficulty of fabricating working sub-micron thermocouples on a flat substrate. The main area of concern was the fact that both components of the thermocouple would have to be lifted off separately with different metal depositions, so that the second metal would have to cross over the first without losing electrical continuity, and while still lifting off successfully at dimensions below 100nm.

The initial design used a 25nm thick first layer and a 50nm thick second layer to reduce the likelihood of electrical continuity being lost. The choice of metals was initially limited to those available in the Plassys MEB 450 evaporator. In the final device these

thermocouples would have to withstand the release wet etch [§4.4.2] which placed further limitations on the choice of thermocouple metals.

Two combinations were chosen: the first was based on noble metals and consisted of gold and palladium; the second was based on non-noble metals and consisted of nickel and nichrome (60% nickel, 40% chromium). Using data in Table 7.1.1 (which is actually for 200°C) the Seebeck coefficient of a gold/palladium thermocouple would be around 15μV/K. For the nickel/nichrome thermocouple, assuming that the nichrome (60% nickel, 40% chromium) behaves like chromel (90% nickel, 10% chromium), the Seebeck coefficient would be around 45μV/K. The latter would produce a bigger signal, however the former has a lower resistance and is likely to be more stable with time (both materials in the latter are prone to being oxidised which is a more significant issue for thin-film thermocouples than ones defined in bulk material).

For the gold/palladium thermocouple, the lowest thermocouple resistance is achieved by using gold (bulk resistivity $2.2 \times 10^{-8}\Omega m$) as the bottom (thinner) layer with palladium (bulk resistivity $10.8 \times 10^{-8}\Omega m$) forming the top (thicker) layer. In the case of the nickel/nichrome thermocouple, nickel (bulk resistivity $6.9 \times 10^{-8}\Omega m$) would form the bottom layer with nichrome (the bulk resistivity of 80% nickel/20% chromium alloy is $108 \times 10^{-8}\Omega m$) forming the top layer. [Note that all data is taken from Goodfellow[16] (1993).]

The first successful attempt (two previous attempts failed due to the use of incorrect exposure doses) consisted of three levels (ignoring an initial alignment marker level). The first level defined the first thermocouple layer, the second level defined the second thermocouple layer while the final level defined ohmic contact pads required for the bonding and probing of the devices. The first level was carried out on a silicon substrate coated with a 50nm thick LPCVD silicon nitride layer. This served to isolate the thermocouple from the silicon substrate. [Normally 150nm to 200nm of PECVD silicon nitride is necessary for adequate electrical isolation, however, the higher quality and density of LPCVD silicon nitride layers meant that even 50nm was enough. During one set of tests, voltages as high as 25V were applied using an HP 4145 parametric analyser without any electrical breakdown occurring.]

After an initial level of lithography to define alignment markers, the first thermocouple level was carried out. This level was split into three matrices to reduce the overall exposure time by making use of larger spot sizes. The samples were processed as follows:

(a) 4% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(b) The resist was dried by baking at 180°C for one hour.

(c) 2.5% Elvacite was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(d) The resist was dried by baking it overnight at 180°C.

(e) The resist was exposed using a resolution of 5nm and a spot size of 15nm at 50kV, with global alignment (100kV was not available due to a problem

with the beamwriter). The pattern consisted of exposing a 1μm long wire, 40nm wide, and tapered regions at either end, tapering to a width of 1μm over a distance of 5μm using a dose of 1000μC/cm^2. This defined the thinnest part of the thermocouple, *i.e.* the part that would be crossed over by the other metal layer.

(f) The resist was also exposed using a resolution of 20nm and a spot size of 40nm at 50kV, with global alignment. The pattern consisted of exposing wires overlapping with the ends of the previous pattern by 0.5μm and tapering out to a width of 25μm before proceeding to 100μm wide square pads using a dose of 400μC/cm^2.

(g) The resist was also exposed using a resolution of 150nm and a spot size of 300nm at 50kV, with global alignment. The pattern consisted of exposing 500μm wide square pads overlapping the pads in the previous pattern using a dose of 350μC/cm^2.

(h) The pattern was developed in MIBK:IPA 2.5:1 at 23°C for 60 seconds.

(i) 5nm of nichrome ('sticky' layer) and 20nm of gold were evaporated onto the resist using the Plassys MEB 450.

(j) Lift-off was then carried out.

The next level was virtually identical to the first, but with the whole pattern rotated 90° about the centre of the narrow single pixel wire. In this case the metal layer evaporated onto the resist pattern consisted of 5nm of nichrome ('sticky' layer) and 45nm of palladium. The use of a very thin sticky layer was so that the layer would not be continuous (thickness of the order of the grain size), although still sufficient to hold the palladium on the substrate, and so that the junction region was as close to isothermal as possible (by having any intermediate layer being as thin as possible) so that the developed Seebeck potential would be that of a gold/palladium thermocouple.

The final level of lithography was carried out as follows:

(a) 15% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(b) The resist was dried by baking at 180°C for one hour.

(c) 4% Elvacite was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].

(d) The resist was dried by baking it overnight at 180°C.

(e) The resist was exposed using a resolution of 150nm and a spot size of 300nm at 50kV, with global alignment. The pattern consisted of exposing 500μm wide square pads, overlapping the pads of step (g) in the previous level, using a dose of 350μC/cm^2.

(f) The pattern was developed in MIBK:IPA 1:1 at 23°C for 35 seconds.

(g) 200nm of aluminium was evaporated onto the resist using a modified bell-jar evaporator.

(h) Lift-off was then carried out.

The devices were then ready for probing with the HP 4145 parameter analyser. For test purposes these first devices actually had two contact pads at each end to enable a four terminal resistance measurement to be made of each thermocouple wire. Figure 7.2.1 shows the overall pattern structure used for these devices.

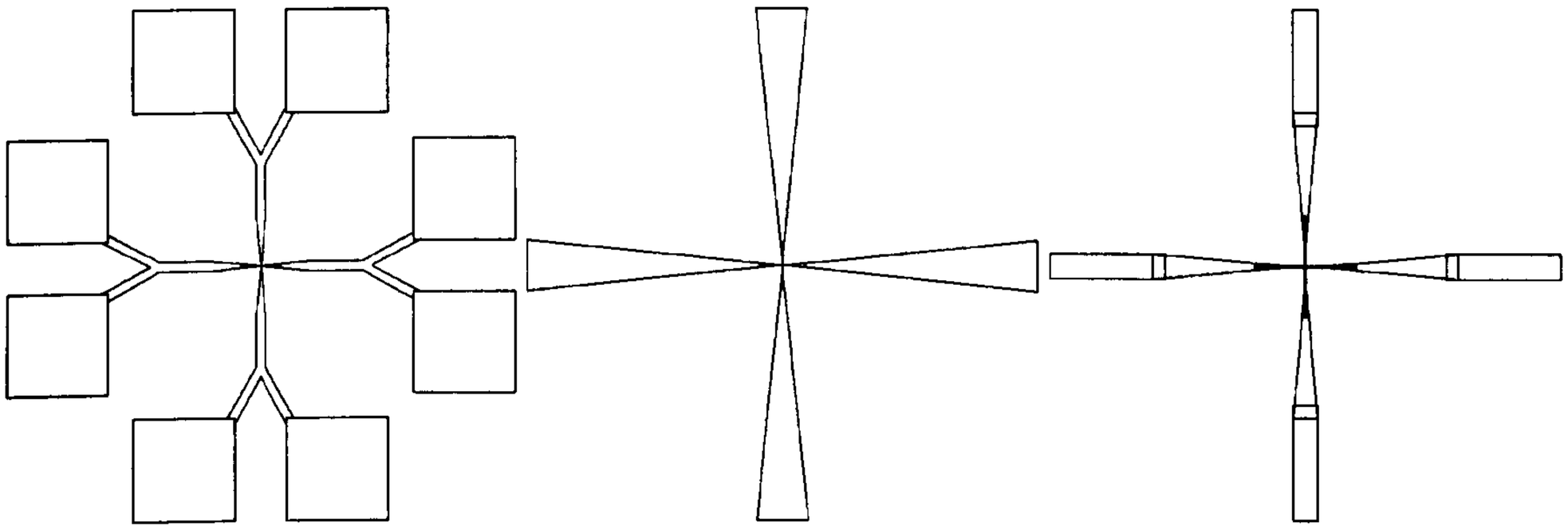


Figure 7.2.1: The design used for initial flat substrate thermocouple tests. From left to right the pictures zoom in on the active central region of the device.

Furthermore, no explicit use of any reference junction was made, instead the thermocouple reference consisted of the contact pads placed more than 1000μm away from the active thermocouple region. As the pads cover a large area, they were expected to be in thermal equilibrium with the device substrate which in turn would be at ambient room temperature. It was also assumed that each pad would be at the same temperature so that any temperature gradient would be solely due to the thermocouple junction being at a different temperature to the two (isothermal) contact pads. However, ambient room temperature would be expected to fluctuate with time and so the reference junction temperature would also change with time. This would not be a major problem since the intended use of the device was not to determine exact absolute temperatures, but to give qualitative and semi-quantitative information about temperature differences across the specimen under examination. Provided that sensible precautions were taken by the user to ensure that the ambient room temperature fluctuated on a timescale much larger than the image acquisition time, such information could be adequately obtained without the use of an additional thermocouple reference.

Measurements carried out on individual wires of each thermocouple showed that the yield was 100% over 16 devices (32 wires), with the typical wire resistance being 1kΩ (measurements made by applying a maximum of 1V across the resistors). However, even though both wires crossed over each other and were electrically continuous, in many cases there appeared to be no electrical contact between the wires. This was assumed to be due to the presence of an organic contamination layer between the two wires. It was found that applying up to 10V across the thermocouple junction while limiting the current to a maximum of 1μA resulted in the breakdown of the organic layer. After such a breakdown, the junction appeared to perform as expected with a typical thermocouple resistance of around 1.5kΩ.

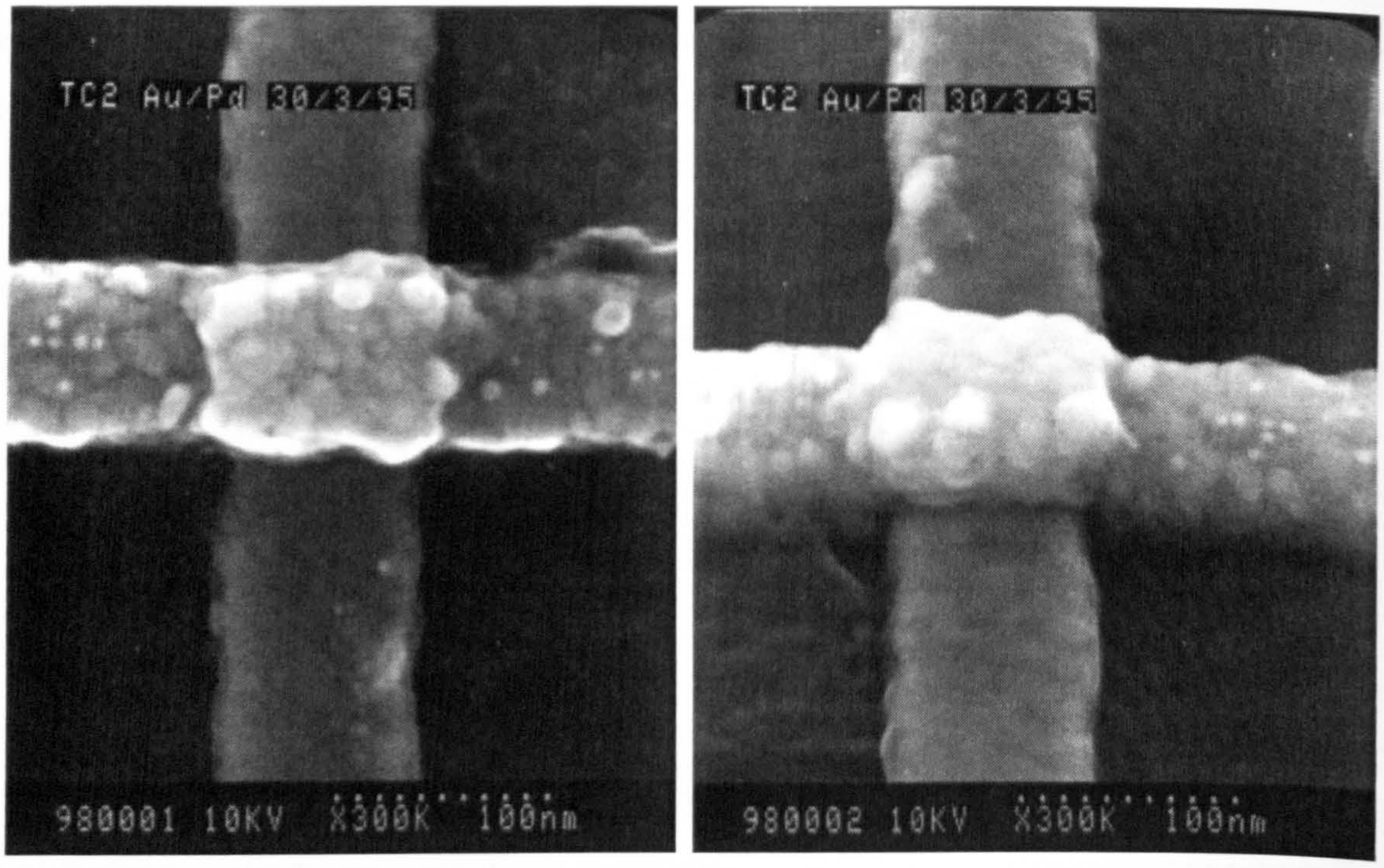


Figure 7.2.2: Micrographs showing a test gold/palladium thermocouple defined on a planar substrate ('amt0005'). Both wires appear to be continuous. (a) Left Image: flat view; (b) Right Image: tilted view.

The need to 'form' the junction was a source of concern as without an automatic probe station, some of the advantages of mass production would be lost. Fortunately, it turned out in subsequent experiments that this was the exception rather than the rule. Figure 7.2.2 shows two micrographs of a thermocouple junction from this sample. It can be seen from the figure that the junction size achieved was about 100nm square, which was extremely encouraging, given that: the exposure had been at 50kV; single pixel lines had not been used, and; a proper exposure test had not been carried out to determine the optimal dose.

Another sample was prepared as above but using the nickel/nichrome metal combination. This was not thoroughly tested, although typical junction resistances were measured to be about 7.5kΩ.

It was now necessary to show that comparable resolutions could be obtained on a pyramidal tip. During the development of the float coating resist technique, as described in §5.2, while the beamwriter was unavailable, Dr. J. M. R. Weaver, attempted to expose crosses on a substrate with pyramidal tips using the Jeol 100 CXII STEM. This was done with manual alignment. He developed the samples and lifted off a layer consisting of 10nm of titanium and 15nm of gold. Figure 7.2.3 shows two such devices. As can be seen from the figure, lift-off was successful and the resolution achieved did not appear to have been compromised by exposure on this non-planar substrate. The cross in Figure 7.2.3a appears to be less than 50nm across. Although this was only a single high resolution level of lithography, the development of the SNOM probes had shown that good alignment

accuracy could be achieved [§6.2] and so it seemed as if the fabrication of thermocouple probes would be relatively straightforward.

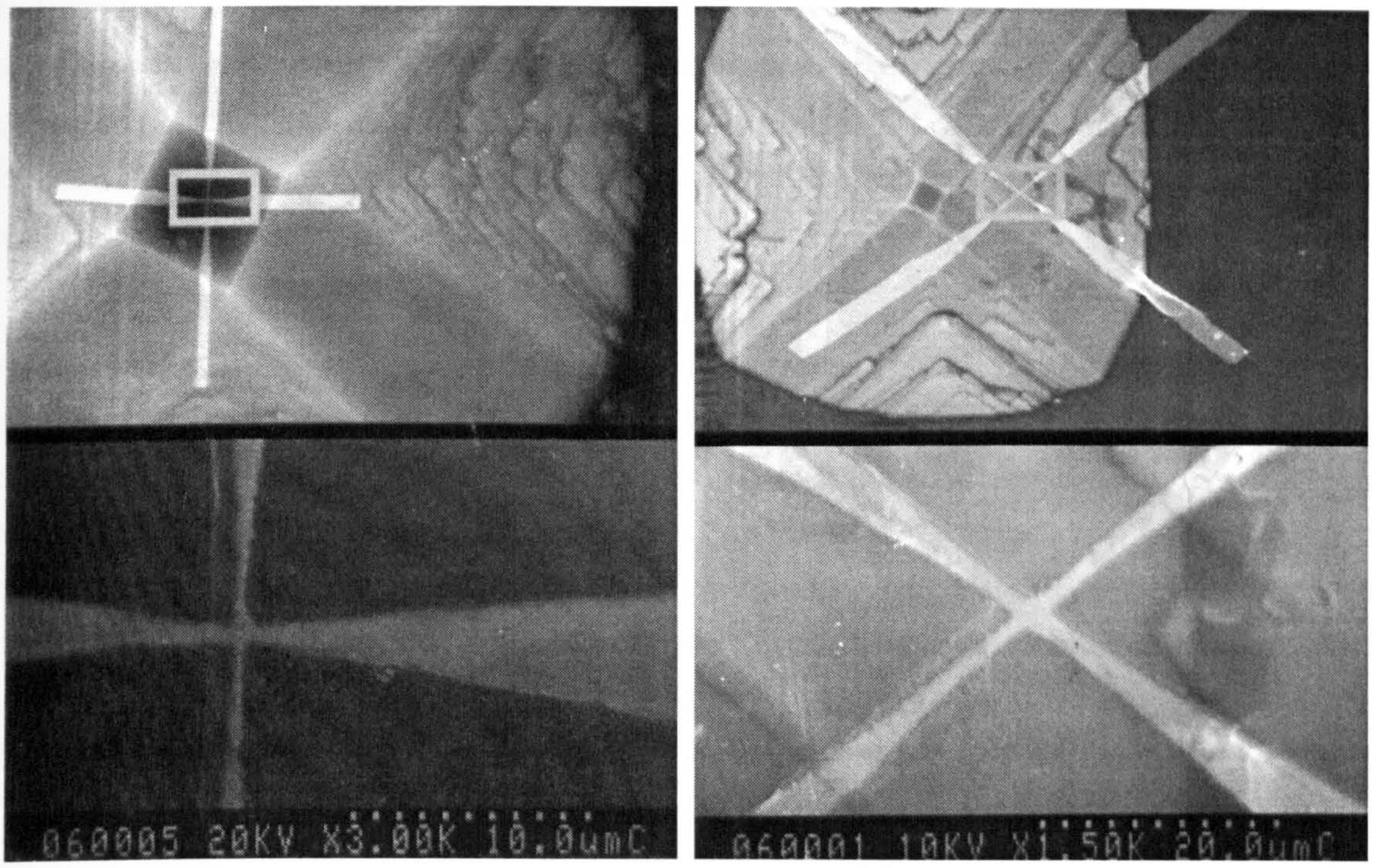


Figure 7.2.3: Micrographs showing the first successful demonstration of high resolution pattern transfer on pyramidal tips. Both images are of cross structures written by manual alignment. (a) Left Image: a properly aligned cross; (b) Right Image: a misaligned cross. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

Some further tests were carried out from which several conclusions were drawn. The first of these was that in order to form a contact pad at each end of each thermocouple wire, it was necessary to bend the wires round so that the pads could be placed off the cantilever and on the main part of the substrate. Lift-off, when it failed, would fail at the inside vertex of any sharp corners. As mentioned in §3.5.3 this was probably because the inside edge was subjected to a greater backscatter dose which reduced the overhanging (undercut) profile in the resist, making it more difficult for the evaporated metal to break over the edge of the resist. The problem could be avoided, and the yield much improved, by replacing sharp vertices with a gradual bend.

The second conclusion was that to reduce the exposure time further, the fine part of the thermocouple should be exposed at 100kV with a 12nm spot, while the tapered region (of step (e), first level pattern, as described above) should be exposed with a 20nm spot. Misalignment would be avoided by making the fine wire longer (typically 5μm long) with the tapered region pattern overlapping this wire but leaving a gap of between 0.5μm and 1.0μm in the middle, for the thermocouple cross-over. This tapered region would take the wire width up to 1μm. In addition a larger spot size of 300nm would be used to expose the next level of tapering to a width of 10μm. However, as noted in §3.5.3 it is not advisable to expose resist with a spot size exceeding or comparable to the resist thickness, thus a 40nm spot would be used to trace the outer edges (2μm typically) of the

larger region (this can be accomplished quite simply using the CATS software by reducing the pattern by 2μm and subtracting it from the original pattern). The larger region would then be exposed after being shrunk by 1μm to ensure that there is a 1μm overlap between the exposure regions. [This is the technique known as 'sleeving' as discussed in §3.5.3.] In addition, no pads would be defined at this level and the 10μm leads only needed to be extended to a safe distance from the junction to allow the contact pad level to overlap and make electrical contact with it. The contact pad level was therefore modified to overlap part of these 10μm wide wires with a 100μm wide wire.

At this point it was decided that further development be handed over to Dr. H. Zhou as time was running short and the author had to concentrate device fabrication on the device that was perceived to be the easiest to complete – SNOM. [However, as it turned out, SNOM would be the hardest of all probes: it had been assumed that because the thermocouple required two levels of alignment and three levels of lithography, while in principle, the SNOM aperture could be defined at the same time as the cantilever with a single level of alignment, the former would be the harder device to fabricate. However, this severely underestimated the development effort required to achieve the correct SNOM aperture profile as discussed in Chapter 6 (§6.2).]

Dr. H. Zhou managed to achieve what appeared to be functioning devices on a cantilever shortly after taking over the development [Zhou, *et al.*[21] (1998)]. Further development and the responsibility for probe characterisation was subsequently handed over to Gordon Mills. [This was to be the focal point of his doctoral research, see Mills[22] (1999).]

7.3 THERMAL RESISTANCE/HEATER PROBES

Another thermal probe was also designed at the request of Dr. H. Pollock of Lancaster University. The attributes of the probe were specified and probe development was carried out by the author in the first instance. This was the first collaborative project undertaken to take advantage of the generic nature of the developed fabrication process and achieve bespoke probes within a shorter development time than would otherwise be possible.

The intended application for this probe was to obtain high resolution differential scanning calorimetry (DSC) measurements [§1.4.2]. In this technique, a constant rate of energy is supplied locally, usually by means of a laser, to a region of a material, typically a polymer, and the temperature of that region is measured as a function of time [*e.g.* Wunderlich and Gaur[23] (1983); Compton, Johnson and Powell[24] (1993); Johnson, Stout, Hill and Krishnan[25] (1994)]. During a phase transition, the temperature of the material does not rise even though energy is still being absorbed. The positions of phase transitions can be located by looking for points of inflection in the variation of temperature with time (*i.e.* energy supplied) and the amount of energy required to pass through the transition can also be measured.

The temperature is usually measured by using a thermocouple located on the specimen, while a focused laser beam is used to supply energy to the specimen. The main drawbacks with this technique are therefore that the temperature is not necessarily measured locally to the region in which energy is being deposited, and the spatial resolution is limited by the non-local temperature measurement as well as the diffraction limited spot size of the laser [see §2.2.2 and §5.1 for the spot size achievable with a diffraction-limited gaussian beam].

A variation of this technique, which combines reflectance Fourier transform infrared (FTIR) microspectroscopy with differential scanning calorimetry, can be used to gain additional information regarding chemical and physical composition simultaneously with the thermodynamic information obtained from differential scanning calorimetry [Compton, *et al.*[24] (1993)]. In this technique the wavelength at which energy is supplied to the specimen is modulated in time and the phase response of the temperature variation is used to determine time variations in the infra-red absorption spectrum of the specimen. This information can then be used to help differentiate between chemical and physical changes in the specimen.

Another variation of this technique, known as modulation differential scanning calorimetry was being practised by Dr. H. Pollock's group. In this case the rate of energy supply is modulated in time and the phase response of the temperature variation is used to obtain information about reversible and irreversible processes in the specimen. The drawbacks of both these variations on DSC are the same as those with the basic DSC technique, however non-local temperature measurement causes more problems for the variations than the basic technique because the time lag between temperature changes and energy deposition is also used to obtain information.

In order to determine the viability of using the probe fabrication methods developed herein to manufacture customised probes for DSC applications, the initial probe specifications were kept relatively modest. The probe as requested by Dr. H. Pollock would allow the localised supply of heat and a localised temperature measurement to be carried out simultaneously. These functions could be fulfilled by what was later called a 'thermal resistor' (due to the fact that the probe could be used both as a thermoresistor for temperature measurement, or as a resistive heater for supplying thermal energy).

The probe design was very simple: a single level resistor would be placed over a pyramidal tip, with its narrowest constriction at the apex of the pyramid. The material used would be of sufficiently high resistivity for it to function as an effective heater. During operation, its resistance could be measured and the temperature calculated from knowledge of its temperature coefficient of resistance, in a similar manner as is commonly used in platinum resistance thermometers [Gardner[7] (1994)].

It was soon realised that this device was essentially a scanning thermal conductivity probe. If constant power was supplied to the resistor, any temperature change would be

as a result of either more or less efficient cooling of the heater. This would be caused by changes in the thermal conductivity between the probe and the sample. If contact mode AFM was used to maintain constant separation between the probe and the sample, then changes in thermal conductivity could be attributed to changes in the rate of energy absorption by the sample. An image with such a probe would therefore reveal areas of high and low thermal conductivity on the specimen. [Alternatively, the current through the heater could be adjusted closed-loop so as to maintain a constant temperature using an appropriately configured feedback loop. This voltage would then vary with the thermal conductivity of the sample.] By modulating the heater current, and measuring the temperature response in time, the phase lag could be used in a similar manner to that in modulation differential scanning calorimetry.

Metal or Alloy	Resistivity at 20°C (Ωm)	Temperature Coefficient of Resistivity (K^{-1})
60% Nickel, 16% Chromium, 24% Iron	109.0×10^{-8}	0.2×10^{-3}
Constantan	49.0×10^{-8}	$\pm 0.02 \times 10^{-3}$
Manganin	43.0×10^{-8}	-0.02×10^{-3}
Palladium	10.8×10^{-8}	3.77×10^{-3}
Platinum	10.6×10^{-8}	3.92×10^{-3}
Iron	9.71×10^{-8}	6.51×10^{-3}
Indium	9.00×10^{-8}	4.70×10^{-3}
Nickel	6.84×10^{-8}	6.81×10^{-3}
Tungsten	5.50×10^{-8}	4.60×10^{-3}
Rhodium	4.70×10^{-8}	4.57×10^{-3}
Aluminium	2.69×10^{-8}	4.20×10^{-3}
Gold	2.30×10^{-8}	3.90×10^{-3}
Copper	1.67×10^{-8}	4.30×10^{-3}
Silver	1.63×10^{-8}	4.10×10^{-3}

Table 7.3.1: Resistivity and temperature coefficient of resistivity for selected metals. Data from Gardner[7] (1994).

The temperature coefficient of resistivity, α_r, is defined by [Gardner[7] (1994)]:

$$\alpha_r = \frac{1}{\rho_o}\frac{d\rho}{dT}, \tag{7.3.1}$$

where ρ_o is the resistivity of the metal at a standard temperature of 0°C and ρ is the resistivity of the metal as a function of temperature, T. For most metals the resistance varies approximately linearly with temperature for temperatures above 50K, so that the resistivity is adequately represented as a quadratic in temperature of the form [Gardner[7] (1994)]:

$$\rho \approx \rho_o\left(1 + \alpha T + \beta T^2\right), \tag{7.3.2}$$

where α and β are material constants. Substituting this into Equation 7.3.1 yields:

$$\alpha_r \approx \alpha + 2\beta T. \tag{7.3.3}$$

For most metals the resistance rises with temperature and $\beta << \alpha$. Table 7.3.1 shows the value of the temperature coefficient of resistivity and resistivity for various metals and alloys. From this it is clear that the best material for heating (most resistive) is an alloy of nickel, chromium and iron. Of the metals available in the Plassys MEB 450 evaporator this is most closely approximated by nichrome with a 60% nickel content and a 40% chromium content. The material offering the greatest temperature sensitivity is nickel, which is also available in this evaporator. Both of these metals are resistant to the release wet etch [§3.7.3] and which is chosen is dependent on which aspect of the probe is deemed more critical. A good compromise is achieved by using palladium.

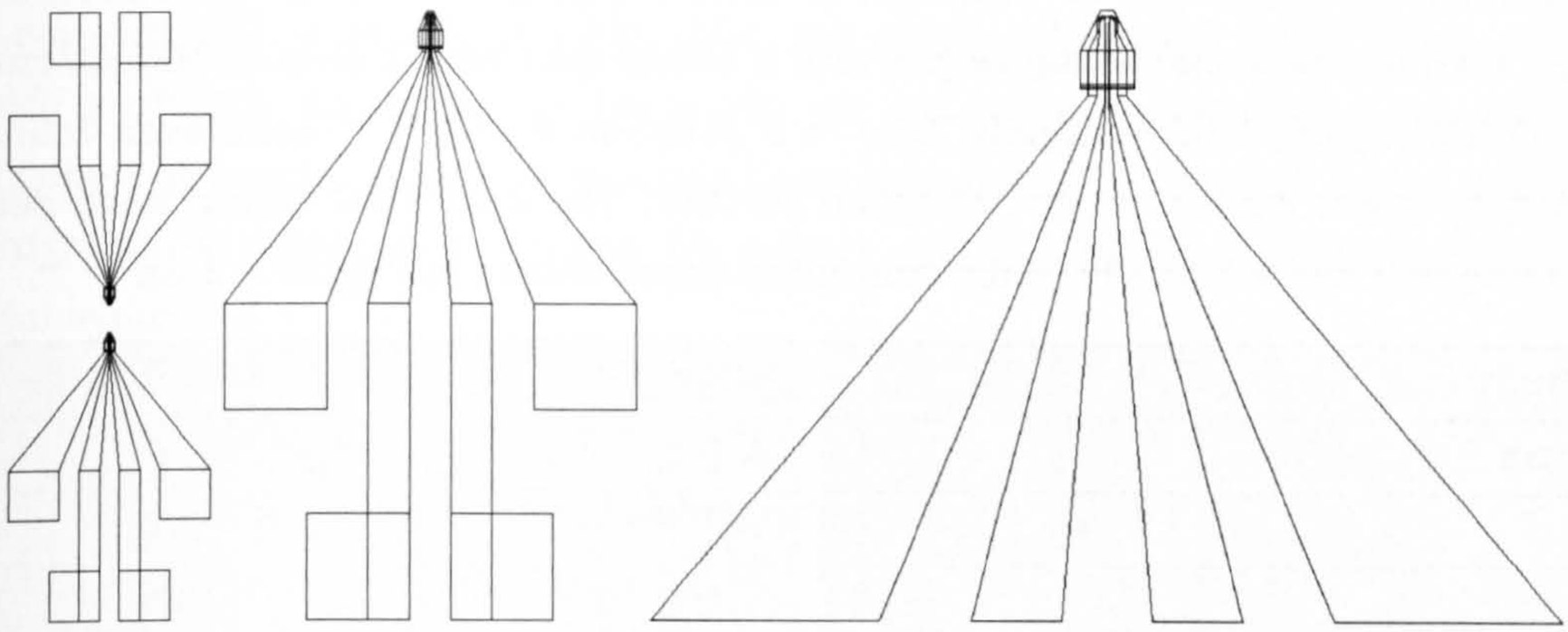


Figure 7.3.1: The design used for the first thermal heater/resistance probes. A pair of facing devices is shown in the picture on the left. The central picture focuses on a single probe. The picture on the right shows a close-up of a single probe in which the outline of the protruding cantilever can be clearly seen.

In order to be able to measure the resistance of the heater accurately it was decided that a four terminal arrangement would be most appropriate. Two heater leads could function as the current (power) supply for both the heater and a four terminal resistance measurement. These would be wide, low resistance leads and would constitute the main part of the device. Two narrower voltage probing leads would come off from the current leads at 90°, some distance away from the heater constriction. This design, as applied to an actual cantilever and pyramidal tip, is illustrated in Figures 7.3.1 and 7.3.2.

The fabrication process required is exceedingly similar to that for the thermocouple probes [discussed in §7.2], but requires one less lithographic level. The results from a test run are shown Figure 7.3.3. Here a layer consisting of 15nm of nichrome and 25nm of palladium was used to form the heater. As can be seen from Figure 7.3.3a, the wider current leads (harder to lift-off on a pyramidal tip) have been made curved while the narrower voltage leads have been positioned at 90°, either side of the constriction. From Figure 7.3.3b it can be seen that the constriction is about 150nm across.

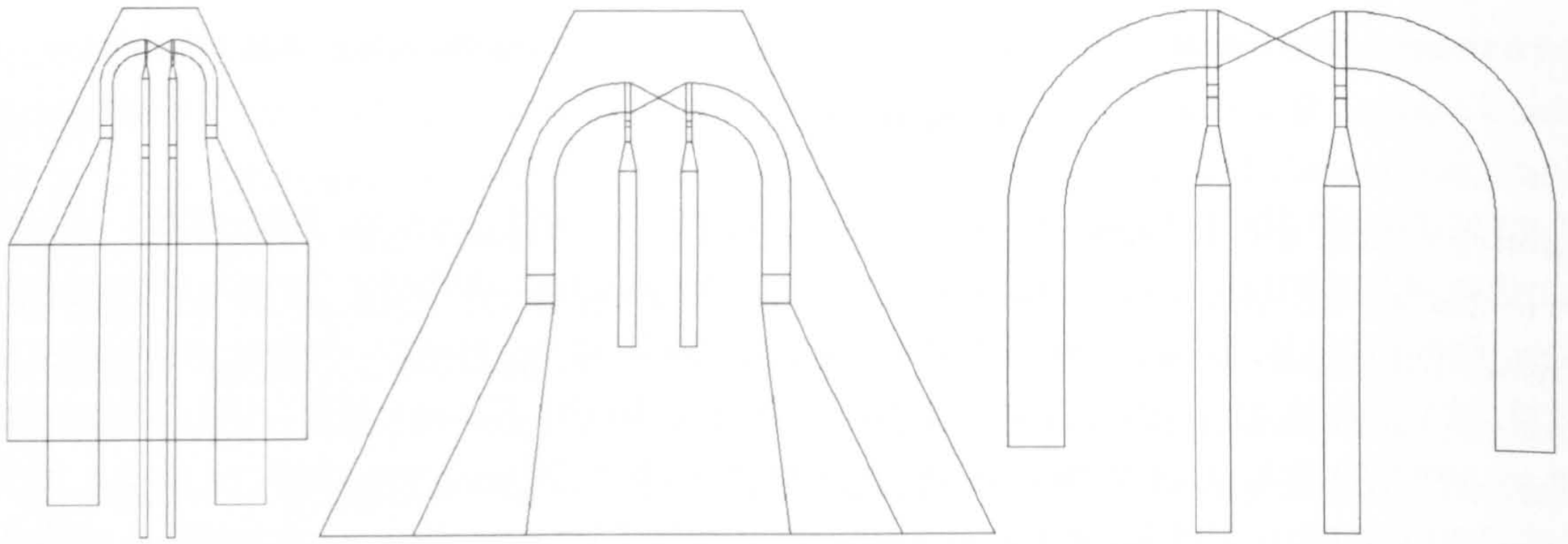


Figure 7.3.2: A closer view of the design used for the first thermal heater/resistance probes. From left to right the pictures zoom in on the active region of the device.

This level was written at 50kV (at this point demonstrating high resolution was not important) with a spot size of 15nm at a dose ranging from 1000 to 1500μC/cm² (exposure test) for the constriction region and a 56nm spot with a dose of 500μC/cm² for the remaining region. Unfortunately, due to a problem during the cantilever definition level [§4.4.2], these devices were not taken further. The next level would have defined contact pads and wider leads to make contact to those shown in Figure 7.3.3a.

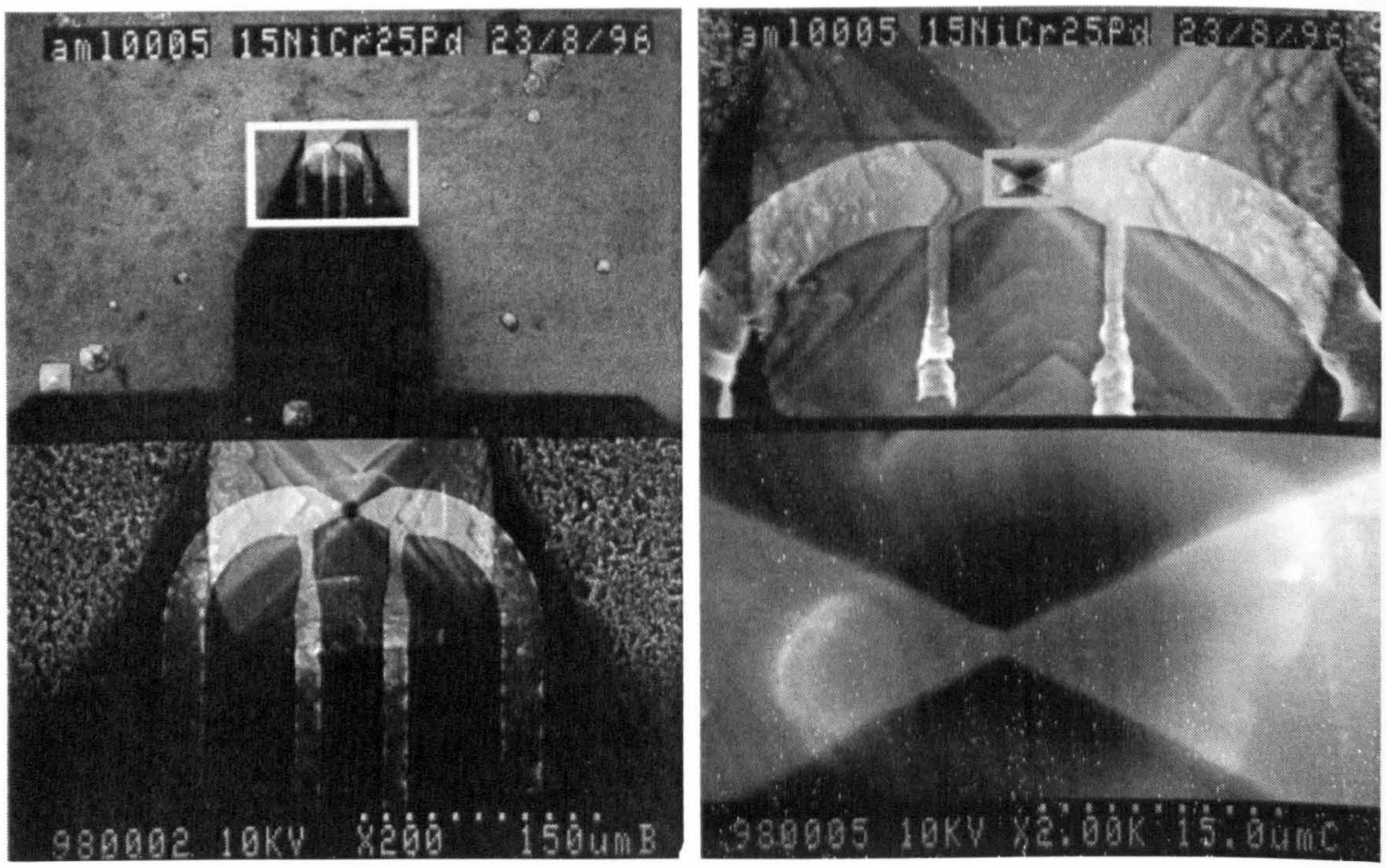


Figure 7.3.3: Micrographs showing the results of a test run for fabricating thermal conductivity probes ('aml0005'). (a) Left Image: large scale view of the device. The bottom half is a 5x magnification over the displayed scale; (b) Right Image: a close-up of the constriction. The bottom half is a 10x magnification over the displayed scale.

Due to time constraints, further development was passed onto Dr. H. Zhou and Gordon Mills. However, process viability had been successfully demonstrated.

7.4 SUMMARY AND CONCLUSIONS

Although a fully functioning and completed thermal probe, either of the thermocouple or thermal resistor variety, was not demonstrated, the basic underlying technology was proven. No major technical hurdles remained to be overcome, although further refinements to the design would be desirable in the future. Thermocouples fabricated on a flat substrate demonstrated electrical continuity, and lift-off tests showed that it was entirely feasible to transfer this flat substrate technology onto the non-planar pyramidal tips. All other aspects of device fabrication were as dealt with in preceding chapters. The choice of material for each probe was not investigated in detail as knowledge gained about process-material compatibility through the development of the previous levels, when coupled with available material data, (as given in Tables 7.1.1 to 7.1.5 and 7.3.1) allowed workable decisions to be made on practical fabrication and theoretical performance grounds. [Had the process been completed, a detailed investigation may have highlighted differences between the actual and theoretical performance of different material systems.]

Thermocouple	Gold/Palladium	Nickel/Nichrome
Seebeck Coefficient at 200°C (μV/K)	15	45†
Resistance at Room Temperature (kΩ)	1.5	7.5
Johnson Noise (nV/√Hz)	4.98	11.15
Temperature Sensitivity (μK/√Hz)	330	250
Minimum Detectable Temperature (mK)	10.5 in 1kHz	7.8 in 1kHz

Table 7.4.1: The temperature sensitivity of fabricated thermocouples calculated from data in the text. †Seebeck coefficient for nichrome is estimated as that given for chromel in Table 7.1.1.

Finally it is instructive to estimate some of the (theoretical) properties of the thermocouple probes designed here. By assuming that the sensitivity of the thermocouple is principally limited by Johnson noise [Gardner[7] (1994)], an estimate of the thermocouple sensitivity can be made. The Johnson noise, V_n, is given by:

$$\frac{V_n}{\sqrt{B}} = \sqrt{4k_B T R_{el}}\,, \tag{7.4.1}$$

in units of V/√Hz, where k_B is the Boltzmann constant, T is the temperature of the device, R_{el} is the electrical resistance of the device and B is the bandwidth in which the measurement is made. Combining this with Equation 7.1.9, the thermocouple sensitivity, in units of K/√Hz, can be defined by:

$$\frac{\Delta T_n}{\sqrt{B}} = \frac{\sqrt{4k_B T R_{el}}}{S_{BA}}, \tag{7.4.2}$$

where ΔT_n is the minimum detectable temperature difference and S_{BA} is the Seebeck coefficient for the thermocouple (as defined in §7.1). Table 7.4.1 shows calculated values of the thermocouple sensitivity for the fabricated thermocouples. [The values used in Equation 7.4.2 are those stated in the text of §7.1 and §7.2.] It is interesting to note that although the nickel/nichrome thermocouple appears to be superior to the gold/palladium thermocouple in terms of signal output, the calculated temperature sensitivity data suggests that it is the gold/palladium thermocouple which is (slightly) more sensitive (due to the increased resistance of the nickel/nichrome system more than offsetting the benefits of a larger Seebeck coefficient).

Another quantity of interest is the energy sensitivity of the thermocouple. This is an analogue of the temperature sensitivity, based on the minimum amount of deposited energy detectable using the thermocouple. This value can be estimated using the heat capacity of the thermocouple junction to calculate the amount of energy required to raise its temperature by the minimum detectable temperature. From the definition of heat capacity [Bolton[26] (1986)], the energy sensitivity, in units of J/√Hz, is given by:

$$\frac{\Delta W_n}{\sqrt{B}} = C_{th}{}^{T} \frac{\Delta T_n}{\sqrt{B}} \tag{7.4.3}$$

where ΔW_n is the minimum detectable energy and $C_{th}{}^{T}$ is the total (as indicated by the superscript) heat capacity of the thermocouple junction. The total heat capacity is simply the sum of the heat capacities of the individual components of the thermocouple junction, each given by:

$$C_{th} = mc_{th} \tag{7.4.4}$$

where m is the mass of material and c_{th} is the specific heat capacity of the material.

Material	Gold	Palladium	Nickel	Nichrome
Density (kg m^{-3})	19300	12000	8900	8400†
Specific Heat Capacity (J kg^{-1} K^{-1})	129	244	444	481‡
Thermal Conductivity (W m^{-1} K^{-1})	318	71.8	90.9	13.4†

Table 7.4.2: Thermocouple material data used for calculating energy sensitivity data. All data from **Goodfellow[16] (1993)**. Density data valid for 20°C, specific heat capacity data valid for 25°C, thermal conductivity data averaged over a range from 0 to 100°C. †Estimated as the value for 80% nickel/20% chromium alloy. ‡Estimated as the (unweighted) average of values for nickel and chromium.

Yet another useful quantity is the time it takes for the thermocouple junction to reach equilibrium after energy is deposited onto it, the so-called *thermalisation* time. This determines the speed of response of the thermocouple to temperature changes. The easiest way to see how to estimate this is by considering the electrical analogy to this problem.

Table 7.4.3 shows the electrical 'equivalents' of various thermal parameters. Since the thermocouple junction stores heat energy, it is analogous to a capacitor which stores

electric charge. The analogous problem is that of how quickly a capacitor reaches its equilibrium voltage if charge is deposited onto one plate. In order for this to make sense it is necessary to consider real capacitors which have an internal leakage resistance associated with them [Duffin[27] (1990)]. The equilibrium voltage (thermal analogue to temperature) will reach some final value with respect to a reference potential, but the voltage difference between the plates will be zero after half the charge deposited has leaked onto the other plate. This is equivalent to saying that when energy is deposited onto one side of the thermocouple junction, the junction will reach some final temperature with respect to a reference (in the case of temperature there is an absolute zero reference) but will have no temperature difference across it.

Thermal Parameter	Symbol	Units	Electrical Parameter	Symbol	Units
Temperature	T	K	Voltage	V	V
Heat Flow, Power	P	W	Current	I	A
Heat	W	J = W s	Charge	Q_{el}	C = A s
Resistance	R_{th}	K/W	Resistance	R_{el}	Ω = V/A
Capacity	C_{th}	J/K	Capacitance	C_{el}	F = A s/V
Thermal Conductivity	κ	W/(m K)	Electrical Conductivity	σ	1/(Ω m)

Table 7.4.3: Electrical analogies of thermal properties. Adapted from van Herwaarden and Meijer[28] (1994).

As is known from electrical theory, the charge will leak in proportion to the difference in charge between the capacitive plates, *i.e.* the final equilibrium situation will be approached exponentially with the 'charge difference' decaying exponentially [Duffin[27] (1990)]. This results in the initial voltage difference between the plates decaying exponentially. From elementary circuit theory [Duffin[27] (1990)]:

$$\frac{V}{R_{el}} + C_{el}\frac{dV}{dt} = 0 \Rightarrow V = V_o \exp\left[-\frac{t}{\tau_{el}}\right], \tau_{el} = R_{el}C_{el}, \tag{7.4.5}$$

where V is the voltage difference between the plates, V_o is the initial voltage difference between the plates, R_{el} is the leakage resistance, C_{el} is the capacitance, t is the time and τ_{el} is the characteristic time constant of the decay (in this case it is the *relaxation time* for the capacitor). The thermal equivalent of this expression is:

$$\Delta T = \Delta T_o \exp\left[-\frac{t}{\tau_{th}}\right], \tau_{th} = R_{th}C_{th}, \tag{7.4.6}$$

where ΔT is the temperature difference across the junction, ΔT_o is the initial temperature difference (due to the deposition of energy on one side of the thermocouple junction), R_{th} is the total thermal resistance of the junction, C_{th} is the total heat capacity of the junction and τ_{th} is the thermalisation time constant.

To simplify this calculation, it will be assumed that the energy is deposited at one end of the thermocouple junction, normal to the plane of contact between the two wires. Then from elementary thermodynamics, for each metal, the rate of energy transfer, P, through it is given by [Bolton[29] (1986)]:

$$P = \frac{dW}{dt} = \frac{\kappa A_{\perp} \Delta T}{l}, \tag{7.4.7}$$

where W is the energy transferred through the material, κ is the thermal conductivity of the material, $A_{\perp}$ is the cross-sectional area in the material normal to the direction of energy transfer, l is the length of the material and ΔT is the temperature difference across the material. By analogy with Ohm's law [see Table 7.4.3], the thermal resistance is given by [Bolton[29] (1986)]:

$$R_{th} = \frac{l}{\kappa A_{\perp}}. \tag{7.4.8}$$

The total thermal resistance of the thermocouple junction is then equal to the sum of the thermal resistances of each of the two thermal components as they are arranged in series.

By combining Equations 7.4.4, 7.4.6, 7.4.7 and 7.4.8, the thermalisation time constant can be estimated. Table 7.4.4 shows estimated values for the thermalisation time constant (for three different junction widths) as well as the energy sensitivity and the minimum detectable energy in a 1kHz bandwidth for each kind of thermocouple calculated using the data in Table 7.4.2. Note that it has been assumed that the temperature sensitivity is as calculated in Table 7.4.1, *i.e.* no account has been taken of the change in junction width (of the contact area between the two materials) or the change in electrical resistance (and hence Johnson noise). This is because the thermocouple resistance is part junction resistance and part lead resistance (the latter is independent of junction width).

From Table 7.4.4 it can be seen that the energy sensitivity, even of the largest thermocouple, is of the order of a few electron Volts per root bandwidth. This means that in a 1kHz bandwidth, using a 20nm junction (which is probably the smallest practical size achievable), an energy transfer of less than 6eV should be detectable for either choice of thermocouple. This is approaching the levels of sensitivity required to detect the progress of chemical reactions.

While all the temperature and energy sensitivity data suggests that there is not much to choose from between either thermocouple, the main factor favouring the use of the gold/palladium thermocouple over the nickel/nichrome one is the response time. For the former the thermalisation time constant is estimated to be about 0.16ns while for the latter it is almost an order of magnitude larger at 1.2ns. [This value is independent of junction width because the scaling of heat capacity with junction area is compensated by that of thermal resistance.]

Thermocouple	Gold/Palladium			Nickel/Nichrome		
Junction Size (nm)	20	50	100	20	50	100
Bottom Layer Material	*Gold*			*Nickel*		
Thickness (nm)	25			25		
Volume ($10^3 nm^3$)	10	62.5	250	10	62.5	250
Mass ($10^{-21} kg$)	193.0	1206	4825	89.00	556	2225
Heat Capacity (eV/K)	155	971	3885	247	1541	6166
Thermal Resistance ($10^3 K/W$)	196.5	31.45	7.862	687.6	110.0	27.50
Bottom Layer Material	*Palladium*			*Nichrome*		
Thickness (nm)	50			50		
Volume ($10^3 nm^3$)	20	125	500	20	125	500
Mass ($10^{-21} kg$)	240.0	1500	6000	168.0	1050	4200
Heat Capacity (eV/K)	366	2284	9138	504	3152	12609
Thermal Resistance ($10^3 K/W$)	1741	278.6	69.64	9328	1493	373.1
Total Heat Capacity (eV/K)	521	3256	13022	751	4694	18775
Total Thermal Resistance ($10^3 K/W$)	1937	310.0	77.50	10016	1603	400.6
Energy Sensitivity ($eV/\sqrt{Hz}$)	0.173	1.08	4.33	0.186	1.16	4.65
Minimum Detectable Energy† (eV)	5.47†	34.2†	137†	5.88†	36.8†	147†
Thermalisation Time Constant (ns)	0.162			1.205		

Table 7.4.4: Estimation of the energy sensitivity and thermalisation time constant for fabricated thermocouples. †Minimum detectable energy in a 1kHz bandwidth.

The difference in the thermalisation time constants between the two thermocouples is principally caused by the low thermal conductivity of nichrome which increases its contribution to the thermal resistance of the junction. [Note that the thermal conductivity used for nichrome is estimated as that of 80% nickel/20% chromium rather than 60% nickel/40% chromium. Also note that the specific heat capacity for nichrome was estimated as the unweighted average of those of nickel and chromium – see Table 7.4.2.]

However it is also expected that the measured thermal resistance of the real thermocouple junction will be higher than that calculated here in both cases because no account has been taken of the fact that any real junction of two metals will not form as good an electrical or thermal contact as if the wire had been continuous. Nevertheless, these calculations do suggest that thermalisation will take place in these small junctions on a timescale of the order of nanoseconds, which opens up the possibility of carrying out experiments to investigate the time evolution of temperature dependent processes.

7.5 REFERENCES

1. B. I. Bleaney and B. Bleaney, *Chapter 11: FREE ELECTRONS IN METALS* in Electricity and Magnetism, 3rd edition, published by Oxford University Press, Oxford, pp. 330-361 (1976).

2. T. R. Sandin, *Chapter 20: MANY-ELECTRON ATOMS* in Essentials of Modern Physics, published by Addison-Wesley, North Carolina, pp. 231-247 (1989).

3. H. P. Myers, *Chapter 14: ASPECTS OF SURFACE PHYSICS* in Introductory Solid State Physics, published by Taylor and Francis, London, pp. 477-500 (1990).

4. S. M. Sze, *Chapter 3: p-n JUNCTION* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 70-108 (1985).

5. H. P. Myers, *Chapter 6: METALLIC BEHAVIOUR AND THE FREE ELECTRON GAS* in Introductory Solid State Physics, published by Taylor and Francis, London, pp. 149-189 (1990).

6. B. I. Bleaney and B. Bleaney, *Chapter 1: ELECTROSTATICS I* in Electricity and Magnetism, 3rd edition, published by Oxford University Press, Oxford, pp. 1-31 (1976).

7. J. W. Gardner, *Chapter 5: THERMAL MICROSENSORS* in Microsensors Principles and Applications, published by John Wiley & Sons, Chichester, pp. 79-116 (1994).

8. D. K. C. MacDonald, *Chapter 1: GENERAL SURVEY* in Thermoelectricity: an introduction to the principles, published by John Wiley & Sons, Ottawa, pp. 1-36 (1962).

9. J. M. R. Weaver, L. M. Walpita and H. K. Wickramasinghe, *OPTICAL-ABSORPTION MICROSCOPY AND SPECTROSCOPY WITH NANOMETER RESOLUTION*, Nature **Vol. 342,** No. 6251, pp. 783-785 (1989).

10. M. Cutler and N. F. Mott, *OBSERVATION OF ANDERSON LOCALIZATION IN AN ELECTRON GAS*, Physical Review **Vol. 181,** No. 3, pp. 1336-1340 (1969).

11. R. D. Barnard, *Chapter 1: BASIC THERMOELECTRICITY* in Thermoelectricty in Metals and Alloys, published by Taylor & Francis Ltd, London, pp. 1-36 (1972).

12. S. M. Sze, *Chapter 2: CARRIER TRANSPORT PHENOMENA* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 30-69 (1985).

13. T. R. Sandin, *Chapter 14: ONE-DIMENSIONAL POTENTIAL WELLS* in Essentials of Modern Physics, published by Addison-Wesley, North Carolina, pp. 155-169 (1989).

14. T. R. Sandin, *Chapter 24: QUANTUM STATISTICS* in Essentials of Modern Physics, published by Addison-Wesley, North Carolina, pp. 301-318 (1989).

15. J. M. Ziman, *Chapter 7: TRANSPORT PROPERTIES* in Principles of the Theory of Solids, published by Cambridge University Press, Bristol, pp. 211-254 (1972).

16. Goodfellow, *METAL DATA TABLE* in The Goodfellow Catalogue, 1993/94 edition, published by Goodfellow, Cambridge, England, pp. 17-31 (1993).

17. M. Jonson and G. D. Mahan, *ELECTRON-PHONON CONTRIBUTION TO THE THERMOPOWER OF METALS*, Physical Review B **Vol. 42,** No. 15, pp. 9350-9356 (1990).

18. K. Durczewski and M. Ausloos, *THEORY OF THE THERMOELECTRIC POWER OR SEEBECK COEFFICIENT: THE CASE OF PHONON SCATTERING FOR A DEGENERATE FREE-ELECTRON GAS*, Physical Review B **Vol. 53,** No. 4, pp. 1762-1772 (19996).

19. D. K. C. MacDonald, *THERMOELECTRICTY: AN INTRODUCTION TO THE PRINCIPLES*, published by John Wiley & Sons, Ottawa (1962).

20. R. D. Barnard, *THERMOELETRICITY IN METALS AND ALLOYS*, published by Taylor & Francis Ltd, London (1972).

21. H. Zhou, A. Midha, G. Mills, S. Thoms, S. K. Murad and J. M. R. Weaver, *GENERIC SCANNED-PROBE MICROSCOPE SENSORS BY COMBINED MICROMACHINING AND ELECTRON-BEAM LITHOGRAPHY*, Journal of Vacuum Science & Technology B **Vol. 16,** No. 1, pp. 54-58 (1998).

22. G. Mills, *SCANNING THERMAL MICROSCOPY USING NANOFABRICATED PROBES*, Ph. D. Thesis, Nanoelectronics Research Centre, Department of Electrical & Electronic Engineering, University of Glasgow, Glasgow, Scotland (1999).

23. B. Wunderlich and U. Gaur, *DIFFERENTIAL SCANNING CALORIMETRY OF FLEXIBLE, LINEAR MACROMOLECULES*, Advances In Chemistry Series No. 203, pp. 195-208 (1983).

24. D. A. C. Compton, D. J. Johnson and J. R. Powell, *A COMBINED DIFFERENTIAL SCANNING CALORIMETRY FOURIER-TRANSFORM INFRARED APPROACH FOR THE STUDY OF POLYMERIC MATERIALS*, Advances In Chemistry Series No. 236, pp. 661-675 (1993).

25. D. J. Johnson, P. J. Stout, S. L. Hill and K. Krishnan, *DIFFERENTIAL SCANNING CALORIMETRY FOURIER-TRANSFORM IR SPECTROSCOPY AND THERMOGRAVIMETRIC ANALYSIS FOURIER-TRANSFORM IR SPECTROSCOPY TO DIFFERENTIATE BETWEEN VERY SIMILAR POLYMER MATERIALS*, Acs Symposium Series **Vol. 581,** pp. 149-161 (1994).

26. W. Bolton, *Chapter 16: SOLIDS, LIQUIDS AND GASES* in Patterns in Physics, 2nd edition, published by McGraw-Hill Book Company, Maidenhead, pp. 325-366 (1986).

27. W. J. Duffin, *Chapter 6: STEADY ELECTRIC CURRENTS* in Electricity and Magnetism, 4th edition, published by McGraw-Hill Book Company, Maidenhead, pp. 128-161 (1990).

28. A. W. van Herwaarden and G. C. M. Meijer, *Chapter 7: THERMAL SENSORS* in Semiconductor Sensors, edited by S. M. Sze, published by John Wiley & Sons, New York, pp. 331-382 (1994).

29. W. Bolton, *Chapter 17: TEMPERATURE AND HEAT* in Patterns in Physics, 2nd edition, published by McGraw-Hill Book Company, Maidenhead, pp. 367-408 (1986).

CHAPTER 8

HALL-PROBE FABRICATION

This chapter describes the attempts made to design and fabricate probes for scanning Hall-probe microscopy. The first section discusses some theoretical and design issues specific to probes for scanning Hall-probe microscopy. The second section then goes on to describe the problems encountered during the limited amount of development work that was carried out to fabricate these probes, and the measures taken in resolving these problems. The final section concludes with a discussion of some material dependant performance issues.

8.1 PRELIMINARY CONSIDERATIONS

As mentioned in §3.2, the probe being designed for use in Hall probe microscopy applications consists of a contact mode AFM cantilever and pyramidal tip [§4.3.3] upon which a sub-micron Hall-effect device has been defined with the active region being coincident with the apex of the tip. Due to the rigidity of a solid silicon cantilever, the prospect of carrying out lithography on these substrates after the release etch [§4.4.3] was not so daunting as on the silicon nitride cantilevers used in contact mode AFM. It was expected that a functioning device would be produced in a relatively short time, as process-material compatibility issues were avoided by having the release etch prior to sensor definition.

Semiconductor	Silicon	Germanium	Gallium Arsenide	Indium Antimonide
Density (g cm^{-3})	2.33†	5.323¢	5.32†	
Lattice Constant (Å)	5.4309†	5.6575¢	5.6533†	6.47
Melting Point (°C)	1412†	937.4¢	1238†	
Boiling Point (°C)	2355*	2830¢		
Energy Gap (eV)	1.12†	0.66	1.424†	0.17
Band Gap Type	Indirect	Indirect	Direct	Direct
Electron Mobility ($cm^2 V^{-1} s^{-1}$)	1450	3900	8500	80000
Hole Mobility ($cm^2 V^{-1} s^{-1}$)	450	1900	400	1250
Electron Effective Mass (m_e)	0.26‡	0.12‡	0.07‡	0.016‡
Hole Effective Mass (m_e)	0.50‡	0.32‡	0.68‡	0.4‡
Intrinsic Carrier Density (cm^{-3})	1.08×10^{10} †	2.4×10^{13} ¢	2.1×10^{6} †	
Intrinsic Resistivity (Ωm)	2300†	0.47¢	1000000†	
Dielectric Constant	11.9	16.0	13.1	17.7

Table 8.1.1: Intrinsic properties of selected semiconductors. Note all data is for 300K. Data collated from sources as marked with unmarked data from Sze[1] (1985). †Data from Sze[2] (1994). ‡Data from Myers[3] (1990). ¢Data from Gleim[4] (1987). *Data from Goodfellow[5] (1993).

As noted in §1.4.4, two dimensional electron gas materials offer very high sensitivity due to their high mobilities and low carrier concentrations. Such materials tend to be very specialised, typically based on III-V semiconductor heterostructures and manufactured by molecular beam epitaxy [Sze[6] (1985)]. As such, these materials are expensive and, in addition, Hall-effect devices based on these structures are usually on the scale of a few microns rather than tens of nanometres, and therefore offer high magnetic field sensitivity at a relatively low lateral resolution [Chang, *et al.*[7] (1992); Oral, Bending and Henini[8] (1996)].

As such materials are so fundamentally different from silicon, it was not possible to make use of these high sensitivity materials within the context of the process being developed here without having to redevelop from the very first fabrication level onwards. Every aspect from wet etching to substrate design would have to be reconsidered and this would not fit in with the 'generic probe' methodology being pursued. Thus it was decided that field sensitivity would be compromised in order to achieve the general aims of a quantitative scanning magnetic field sensing device offering high lateral resolution.

From §1.4.4, Equation 1.4.33 shows how the Hall coefficient is defined in terms of the number density of carriers in the material, based on modelling transport behaviour as in a metal. It is clear from this that the Hall coefficient of metals must be lower than that of semiconductors as the free carrier density is much higher.

Equation 1.4.34 shows how the carrier mobility is related to the average drift velocity of the carriers and applied electric field. By combining Equations 1.4.34 and 1.4.31 (which relates the current density to the number density of carriers and carrier drift velocity) and applying Ohm's law in the form $J = \sigma E$, the following relationship between the electrical conductivity, carrier mobility and number density of carriers may be obtained:

$$\sigma = nq\mu\,, \tag{8.1.1}$$

where all terms are as have been defined in Equations 1.4.31 and 1.4.34. This can be substituted into Equation 1.4.33 to give the Hall coefficient in terms of the electrical conductivity or resistivity and the carrier mobility:

$$R_H = \frac{1}{nq} = \frac{\mu}{\sigma} = \rho\mu\,, \tag{8.1.2}$$

where ρ is the (electrical) resistivity of the material and all other terms are as defined in Equations 8.1.2 and 1.4.33. This equation suggests that materials with high resistivity and high mobility make better Hall-effect devices. Table 8.1.1 lists some properties of intrinsic semiconductors.

However, in an intrinsic semiconductor the situation is not so simple because of the presence of carriers of both signs (electrons and holes). Then the underlying theory must be modified so that the electrical conductivity is given by [Myers[3] (1990)]:

$$\sigma_i = n_i q_e \mu_n + p_i q_e \mu_p\,, \tag{8.1.3}$$

where σ_i is the intrinsic (electrical) conductivity, n_i and μ_n are the intrinsic number density and mobility of the negative carriers (respectively) and, p_i and μ_p are the intrinsic number density and mobility of the positive carriers (respectively). [In a truly intrinsic semiconductor $n_i = p_i$, however in practice, one type of carrier tends to dominate due to the presence of impurities.]

The Hall coefficient in this case is then given by [Gardner[9] (1994)]:

$$R_H = \frac{\mu_p - \mu_n}{\sigma_i} = \frac{\mu_p - \mu_n}{q_e\left(n_i\mu_n + p_i\mu_p\right)}\,. \tag{8.1.4}$$

For extrinsic material the Hall coefficient is given by [Gardner[9] (1994)]:

$$R_H = \begin{cases} -\dfrac{1}{nq_e}, n-\text{type} \\ +\dfrac{1}{pq_e}, p-\text{type} \end{cases}, \tag{8.1.5}$$

where n and p are the respective number densities of carriers in n- and p- type materials. In practice it is found that a correction factor needs to be introduced (for non-degenerate semiconductors) so that the mobilities of the carriers are given by [Baltes and Castagnetti[10] (1994)]:

$$\mu_n^* = r_n\mu_n, \mu_p^* = r_p\mu_p \tag{8.1.6}$$

where μ_n^* and μ_p^* are the effective mobilities of electrons and holes in the material respectively, and r_n and r_p are scattering factors for electrons and holes respectively. This factor takes account of the energy dependence of scattering mechanisms in a semiconductor and is significant due to the presence of the band gap restricting the states into which scattering can take place. [For low doped silicon, $r_n \approx 1.15$ for n-type material and $r_p \approx 0.7$ for p-type material at room temperature, from Gardner[9] (1994).]

In this case the Hall coefficient is given by [Gardner[9] (1994)]:

$$R_H = \begin{cases} -\dfrac{r_n}{nq_e} = -\mu_n^*\rho_n, n-\text{type} \\ +\dfrac{r_p}{pq_e} = +\mu_p^*\rho_p, p-\text{type} \end{cases}, \tag{8.1.7}$$

where ρ_n and ρ_p are the resistivities of electrons and holes respectively [this can be compared to Equation 8.1.2 for metals]. Applying Equation 8.1.6 to materials in which conduction takes place by both types of carrier then means that the Hall coefficient for intrinsic material is given by [Baltes, *et al.*[10] (1994)]:

$$R_H = \frac{r_p p\mu_p^{\,2} - r_n n\mu_n^{\,2}}{q_e\left(n\mu_n + p\mu_p\right)^2}. \tag{8.1.8}$$

From this it can be seen that the mobilities and carrier densities of each type of carrier add in opposition to reduce the overall Hall coefficient. For this reason (strongly) extrinsic material is desirable, for which the Hall coefficient is determined by Equation 8.1.7 rather than Equation 8.1.8. This means that materials which have a large number of intrinsic carriers at room temperature are undesirable.

Most (room temperature) Hall-effect devices are therefore made from crystalline silicon or gallium arsenide as their large bandgaps, upon which the numbers of intrinsic carriers depend exponentially [Sze[11] (1985)], outweigh the disadvantage of lower carrier mobility when compared with more exotic semiconductors, such as indium antimonide and indium arsenide, which have high (negative) carrier mobilities [80000$cm^2V^{-1}s^{-1}$ and 33000$cm^2V^{-1}s^{-1}$ for electrons respectively, from Sze[1] (1985)], but unfortunately also have

small band gaps [0.17eV and 0.36eV respectively, also from Sze[1] (1985)]. In addition silicon and gallium arsenide offer the advantages material availability and low cost as well as a well-developed and characterised processing technology. Small band gaps mean that intrinsic carrier concentrations are significant at room temperature unless the material is extremely highly doped [Baltes, *et al.*[10] (1994)]. However, at high doping concentrations the Hall coefficient decreases as the number of carriers rises and the resistivity falls so that the behaviour becomes more metal-like.

From Table 8.1.1 it is also clear that *n*-type material is preferable to *p*-type material due to the higher mobility of the former. However, whilst gallium arsenide would appear to be a better choice than silicon, it is a different material technology and was ruled out due to the amount of redevelopment that would be required, for similar reasons as given earlier. This left silicon as the only clear choice. In principle, since the cantilever already consisted of single crystal silicon, the Hall junction and wires need only be defined by doping the silicon in the appropriate geometry. Unfortunately, (thermally induced) diffusion doping would have led to the dopants spreading significantly into the silicon substrate and therefore resulted in a larger Hall junction and lower lateral resolution than originally defined by the dopant masking layer [Sze[12] (1985)].

The alternative would have been to use ion implantation which would have led to less lateral spreading of dopant, but the dopant layer (and hence the Hall junction) would have been buried beneath the surface. In addition, the relatively large spread of dopants in the direction of ion implantation would reduce the observed Hall voltage in accordance with Equation 1.4.30 due to the increased thickness of the active layer [Sze[12] (1985)]. Furthermore, due to the unavailability of this facility within the department, the use of external facilities would have substantially increased the length of the development cycle for these probes.

Thus it was decided that, at least for initial development, the material to be used be limited to those which could either be evaporated or sputtered, thereby completely ruling out semiconductors consisting of more than a single atomic species. This would keep the initial development as simple as possible. Expertise in designing and patterning evaporated materials to form devices on these probes would then come from the thermocouple probe development [Chapter 7], while that in designing and etching sputtered materials would come from the SNOM probe development [Chapter 6].

The best choice then became polysilicon. In this case a thin (around 50nm thick) polysilicon layer would be deposited onto the released cantilever, isolated from the silicon of the cantilever by the layer of silicon nitride used to mask the fabricated cantilever from the release etch [§4.4.3]. The junction itself could then be patterned by dry etching this polysilicon layer. This would have meant either developing a dry etch which would etch stop on the silicon nitride layer beneath, or using in-situ reflectance measurements to indicate when dry etching should cease [Hicks, Parkes, Wilkinson and Wilkinson[13]

(1994)]. The masking layer for this etch would probably have been patterned in a single step, unlike that used to define the SNOM aperture [§6.2.3], *e.g.* the area to remain unetched (*i.e.* the Hall junction) could be defined in positive resist and nichrome could be lifted off as a masking layer. Ideally the pattern would define a Hall junction of sub-100nm dimensions so that the dry etch used would have to be able to etch 50nm deep at dimensions of around 50nm, a relatively straightforward proposition if slight overcutting or undercutting is acceptable. Contacts would have to be made to the polysilicon layer (in the same way as for silicon) and would require annealing, however, as the substrate consisted solely of silicon and silicon nitride the temperature required was not likely to cause any problems.

The main issue would be of doping the polysilicon. Ideally this would be done in-situ, using phosphorus to give an *n*-type layer, and a series of experiments would probably be necessary to determine the change in carrier concentration and Hall coefficient of the layer, before and after deposition, as a function of initial doping concentration.

In addition, unless the deposition temperature was above 600°C, the initial layer would be amorphous [Adams[14] (1988)], with a very high resistivity, caused by conduction taking place primarily via 'hopping' [this occurs when the long range order of the crystal material is upset to such a degree that many silicon bonds remain uncompensated, pinning the Fermi level to midgap, independent of dopant concentration – see Myers[3] (1990)]. Annealing at a temperature of around 1000°C is required to 'recrystallise' the silicon [Sze[15] (1985)]. This anneal is also beneficial for films deposited above 600°C, causing crystal growth and a decrease in resistivity [Adams[14] (1988)].

Unfortunately, polysilicon deposition facilities were also not available within the department and with the high development workload required for the other probes, only part of the required process for this sensor could be developed in the time available. Being aware of this, it was decided that the material choice be further limited to one which could be deposited internally. This left germanium, as evaporated using the Plassys MEB 450 evaporator [§3.6.2], as the material to be used for the initial development of these sensors.

8.2 FABRICATION OF HALL-EFFECT DEVICES

As can be seen from Table 8.1.1 germanium has a bandgap just over 41% smaller than that of silicon and nearly 54% lower than that of gallium arsenide. As a result its intrinsic carrier concentration is over 3.3 orders of magnitude larger than that of silicon and over 6.7 orders of magnitude larger than that of gallium arsenide. However its electron mobility is almost 170% higher than that of silicon although still less than half that of gallium arsenide.

One other potentially useful piece of information about germanium is the variation of the product of the number densities of holes and electrons (in units of m^{-3}) with temperature [Gleim[4] (1987)]:

$$n_i^2 = np = 3.1 \times 10^{38} T^3 \exp\left[-\frac{0.785eV}{k_B T}\right], \tag{8.2.1}$$

where k_B is the Boltzmann constant and all other terms are as defined in §8.1.

An obvious problem with using germanium as evaporated in a conventional evaporator is the inability to carry out in-situ doping. At the physical dimensions required, doping after deposition did not appear to be a realistic possibility (for the same reasons as doping the silicon cantilever). Therefore some initial experiments were required to determine the nature of the available germanium. The supplier-stated characteristics of the germanium used in the Plassys MEB 450 are that it is *n*-type with a resistivity of between 5 and 40Ωcm. Undoped single crystal germanium usually has a resistivity of around 47Ωcm as given in Table 8.1.1. Thus the germanium was either, not heavily doped so that minority carriers would play a significant role in determining its behaviour, or amorphous in which case the level of impurities cannot easily be estimated. However, as this germanium remained in the evaporator and was used as a part of various metal layer combinations evaporated in the Plassys MEB 450, it was also likely that the nature and distribution of impurities within the germanium would have altered significantly from that supplied. For example, it was possible that repeated melting of the germanium could have resulted in some segregation of impurities. It was also likely that the evaporation of other metals resulted in some contamination of the germanium. Therefore there was little alternative to direct experimentation for determining how these processes affected the characteristics of the evaporated germanium film.

The first test was to determine the practicality of lifting off germanium at small dimensions. The pattern used consisted of the patterns for both wires of the thermocouple probes [Chapter 7] written simultaneously to form a Hall junction on a flat LPCVD silicon nitride coated substrate. The sample was processed as follows:

(a) 4% BDH was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].
(b) The resist was dried by baking at 180°C for one hour.
(c) 2.5% Elvacite was spin coated at 5000 rpm for 60 seconds [Table 3.5.1].
(d) The resist was dried by baking it overnight at 180°C.
(e) The resist was exposed using a resolution of 5nm and a spot size of 15nm at 50kV, with no alignment. The pattern consisted of exposing a 1μm long wire, 40nm wide and tapered regions at either end, tapering to a width of 1μm over a distance of 5μm using a dose of 1000μC/cm^2. This defined the thinnest part of the Hall junction.
(f) The resist was also exposed using a resolution of 20nm and a spot size of 40nm at 50kV, with no alignment. The pattern consisted of exposing wires

overlapping with the ends of the previous pattern by 0.5μm and tapering out to a width of 25μm before proceeding to 100μm wide square pads using a dose of 400μC/cm^2.

(g) The resist was also exposed using a resolution of 150nm and a spot size of 300nm at 50kV, with no alignment. The pattern consisted of exposing 500μm wide square pads overlapping the pads in the previous pattern using a dose of 350μC/cm^2.

(h) The pattern was developed in MIBK:IPA 2.5:1 at 23°C for 60 seconds.

(i) 50nm of germanium was evaporated onto the resist using the Plassys MEB 450. [A 'sticky layer' was not found to be necessary.]

(j) Lift-off was then carried out.

As a method for establishing contacts had not yet been developed, it was hoped that simply breaking through the germanium oxide layer with the probes of the HP 4145 parametric analyser would be sufficient to allow device continuity and resistance to be estimated. Figure 8.2.1 shows that lithographically defining a germanium Hall junction with a width of around 100nm was relatively straightforward.

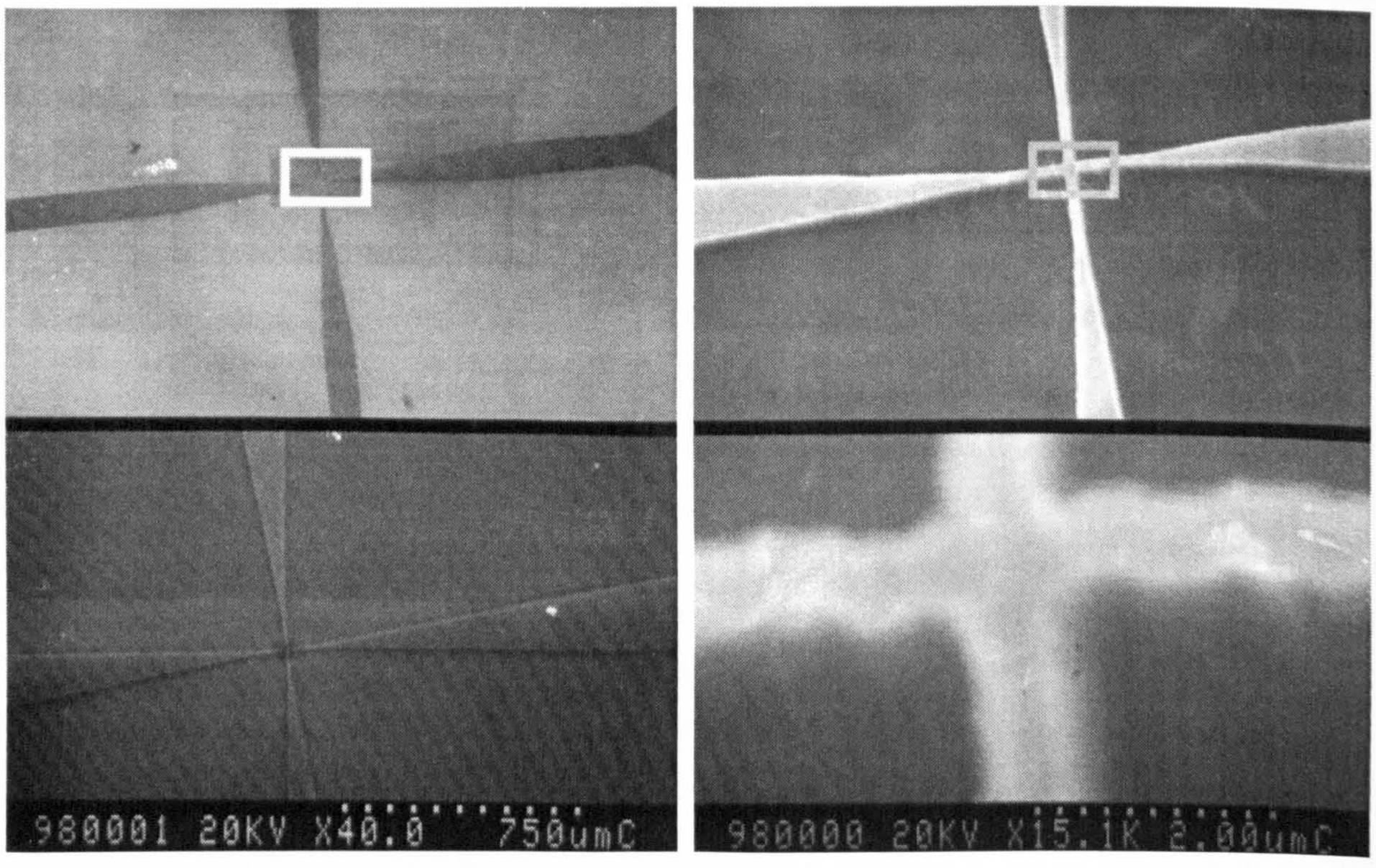


Figure 8.2.1: Micrographs showing the results of a test run for fabricating Hall junctions on a flat substrate. (a) Left Image: large scale view of the device; (b) Right Image: a close-up of the junction. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

However, when electrical measurements were made of this device, it was found that although electrical contact could be made directly by pressing the probes into the layer, the measured resistance was of the order of 500GΩ (the reliability of this measurement is somewhat dependent on the impedance of the HP 4145 analyser). Due to this high resistance, the measurement had to be made very slowly (large 'RC' time constant).

Despite this, the initial charging of the device could not be avoided. [An alternative explanation for the observed I-V characteristic would have been that a Schottky contact had been made between probe and device. However as the voltage sweep extended from negative applied potentials to equally positive applied potentials, such a contact would have been expected to yield an I-V characteristic rotationally symmetric about zero volts, *i.e.* that of two back-to-back diodes. The observed characteristic was a simple continuation of a linear increase in current for linearly increasing applied potential, once the initial charging curve had been passed. It was therefore concluded that ohmic contact had been made between probe and device.] An additional observation was the sensitivity of the measurements to the ambient fluorescent lighting in the room. With the lights off the resistivity went up significantly suggesting that conduction via photocarrier generation was significant.

The high value of this resistance was somewhat surprising initially, since the thermocouple wires with the same geometry but made of gold exhibited a resistance of around 1kΩ. Since the conductivity gold is 2.2×10^{-6} Ωcm [Goodfellow[5] (1993)], and taking the resistivity of germanium to be at the upper limit value of 40Ωcm, a device resistance of around 18GΩ would have been expected. However, as has been noted in §8.1 regarding the deposition of polysilicon, this was probably caused by the deposition of germanium being carried out at room temperature and resulting in an amorphous layer in which conduction principally took place via hopping. Therefore, the solution was to anneal the germanium until it recrystallised to form 'polygermanium' [as already discussed for polysilicon in §8.1].

That an ohmic contact with the deposited film was made so readily was also somewhat surprising, however as can be seen from Table 8.2.1, the energy required to ionise various impurities in germanium is much lower than that in silicon (less even than the 25.8meV of thermal energy available at 300K), and so it was possible that physical contact between probe and device was sufficient to form an ohmic connection at room temperature. However, this would need to be investigated further.

Having such a high resistance was a problem, not only due to the large settling time required for any voltage measurement (which would reduce the rate at which measurements could be taken and hence the maximum scan rate achievable), but also because of the difficulties in impedance matching the sensor with other signal processing circuitry and in finding a voltage measurement device of sufficiently high impedance. In addition, a high resistance would also mean using a higher voltage to obtain sufficient current for any measurements. This, in turn, would mean that the region of the specimen closest to the Hall junction would experience a high electric field gradient, possibly altering its behaviour and resulting in erroneous data. Finally, having a light-sensitive magnetic sensor was also undesirable from a practical perspective.

Impurity Atom	Chemical Symbol	Ionisation Energy in Silicon (meV)	Ionisation Energy in Germanium (meV)
Boron	B	45	10.4
Aluminium	Al	57	10.2
Gallium	Ga	65	10.8
Indium	In	160	11.2
Phosphorus	P	45	12
Arsenic	As	49	12.7
Antimony	Sb	39	1

Table 8.2.1: Ionisation energies for impurity levels in silicon and germanium. Data from Myers[3] (1990).

To investigate the annealing of the germanium film, a pattern was constructed which consisted of defining large scale germanium resistors, 1000μm long and with widths of 10μm, 20μm, 50μm and 100μm, connected to 500μm square contact pads, as illustrated in Figure 8.2.2. These could then be probed before and after annealing. These patterns were written using electron-beam lithography in order to ensure dimensional reproducibility and accuracy.

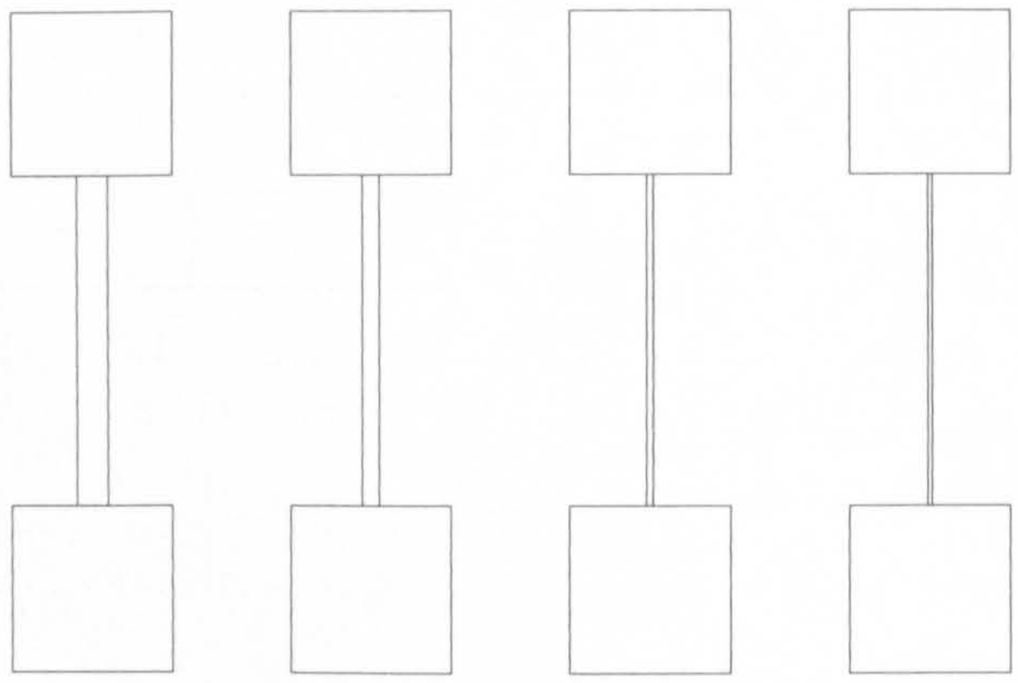


Figure 8.2.2: The pattern used to define resistors for germanium annealing tests. From left to right the resistor widths decrease from 100μm, through 50μm and 20μm, to 10μm.

Figure 8.2.3 shows how the resistance of a 50nm thick germanium film (as deposited) varies with the ratio of the resistor length to the product of resistor width and thickness. This would be expected to yield a straight line of gradient numerically equivalent to the (bulk) resistivity of the material. [Note that all measurements were carried out without illumination with the largest width resistor (100μm) corresponding to the leftmost point and the smallest width resistor (10μm) corresponding to the rightmost point.] The narrowest width resistor appears to have a slightly lower resistance than expected by extrapolating from the other measurements. If this measurement is excluded, a straight line seems to reasonably describe the variation of resistance with the geometric ratio defined above. From fitting the first three data points, the resistivity of the as-deposited film is estimated to be 264Ωcm (2.64Ωm) which is still lower than that for intrinsic silicon [see Table 8.1.1]. This compares with 225Ωcm (2.25Ωm) if all the data points are used.

However when measuring the resistivity of thin films it is more conventional to quote the resistance per square, R_{sq}, defined by:

$$R_{sq} = \frac{\rho}{t} \Rightarrow R_{el} = R_{sq}\frac{l}{w}, \tag{8.2.2}$$

where ρ is the normal bulk resistivity, l, w and t are the length, width and thickness of the resistor (respectively) and R_{el} is the total resistance. As the germanium film was 50nm thick the resistance per square of the as-deposited film is estimated as 52.8MΩ/square (based on the first three points).

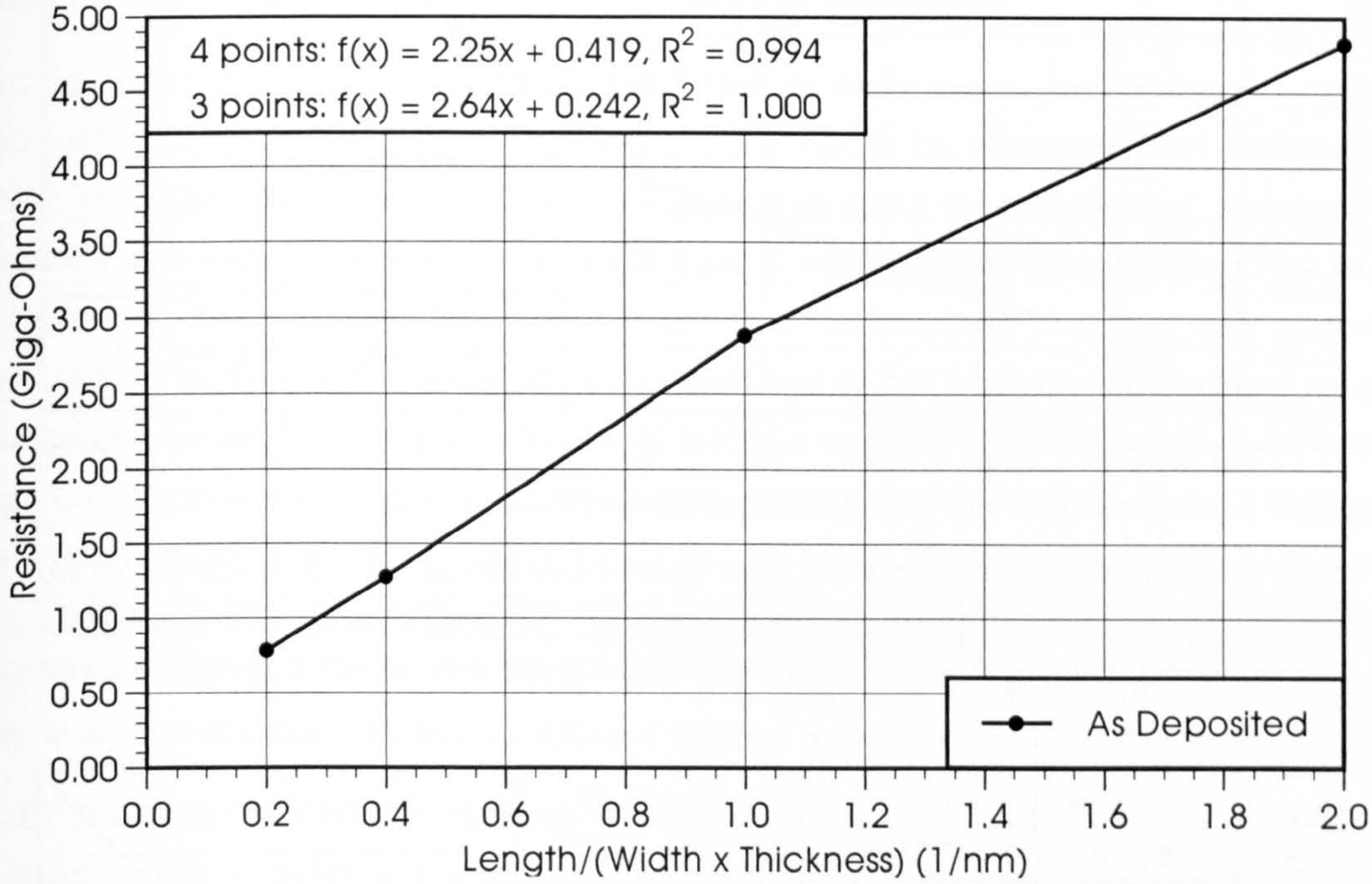


Figure 8.2.3: Resistance, f, of an as-deposited, 50nm thick germanium film as a function of geometry, x (defined as the ratio of length to product of width and thickness). The gradient should be numerically equivalent to the bulk resistivity in units of Ωm. The coefficients of the best fitting line (by least squares) to the first three points and all four points are also given.

Annealing was carried out using the rapid thermal annealing method normally used for annealing semiconductor contacts [§3.6.8]. The process used consisted of radiatively heating the sample in vacuum with the annealing temperature being achieved rapidly. Rapid cooling was accomplished by flushing the chamber with inert gas. Initially a series of tests were conducted at a range of temperatures from 300°C to 600°C and annealing times ranging from 10 to 60 seconds. In all cases the inert gas used to flush the chamber both before (to purge any air and particularly oxygen) and after annealing was argon. [The other choice available was nitrogen, but it was decided that argon, being more inert than nitrogen, would be a safer choice.] The temperature ramping time (both up and down) used was 10 seconds in all cases.

Figures 8.2.4 and 8.2.5 show the results achieved from annealing at 450°C and 475°C for 60 seconds, respectively. Note that temperature measurements were made using an

optical pyrometer rather than a thermocouple. For these tests, the samples were placed on a silicon wafer during heating.

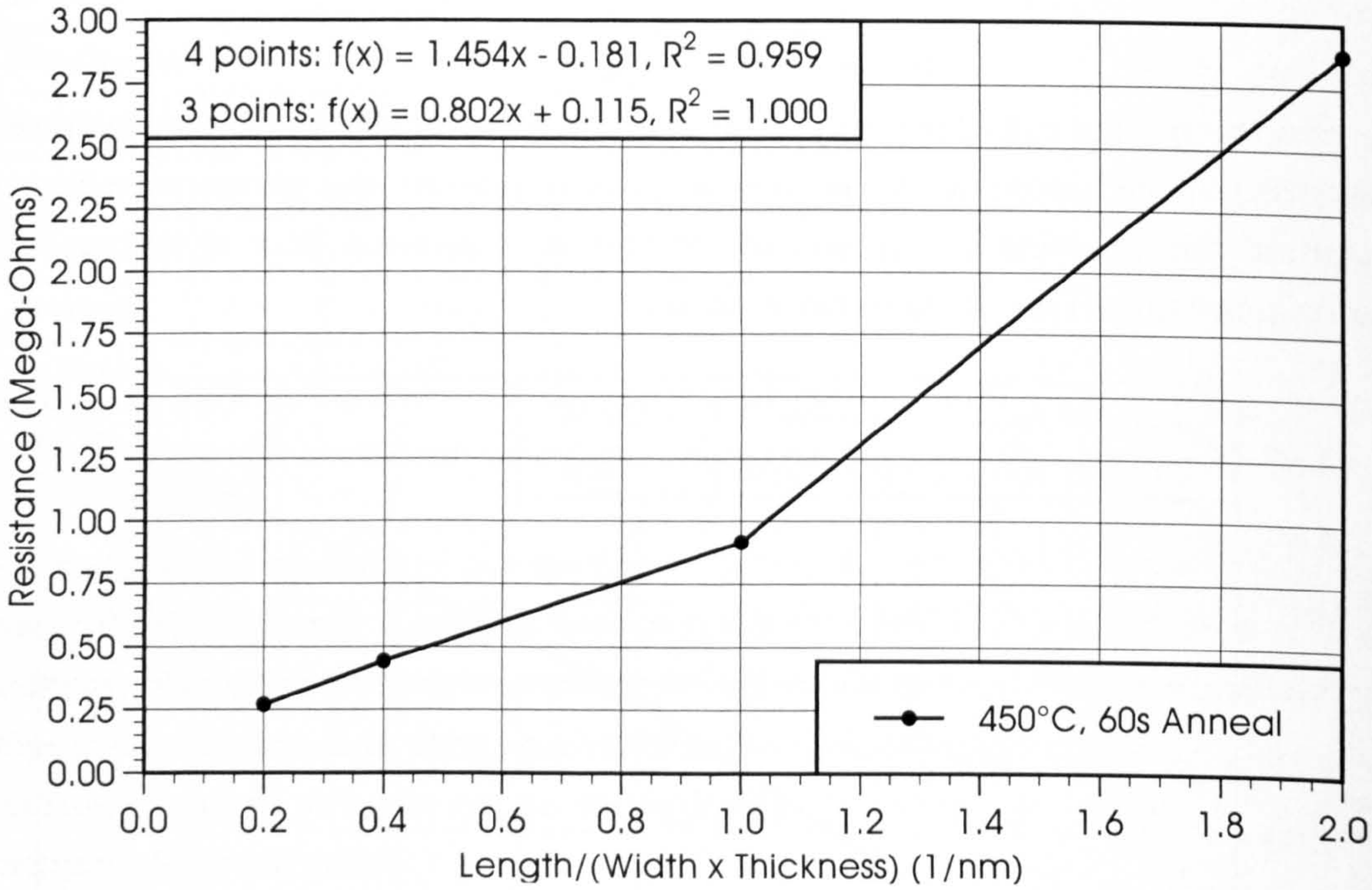


Figure 8.2.4: Resistance, f, of a 50nm thick germanium film annealed at 450°C for one minute as a function of geometry, x (defined as the ratio of length to product of width and thickness). The gradient should be numerically equivalent to the bulk resistivity in units of $10^{-3}\Omega$m. The coefficients of the best fitting line (by least squares) to the first three points and all four points are also given.

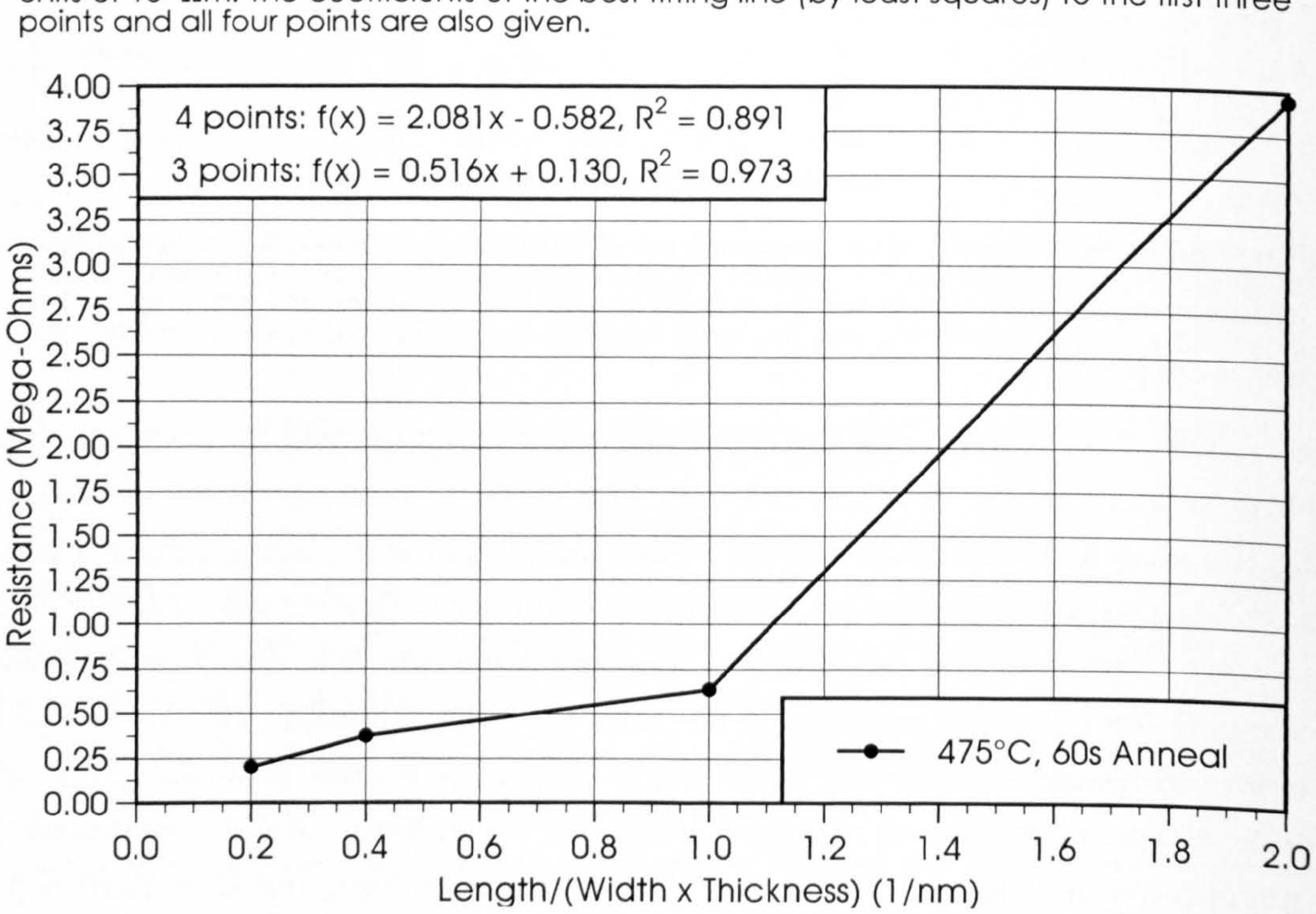


Figure 8.2.5: Resistance, f, of a 50nm thick germanium film annealed at 475°C for one minute as a function of geometry, x (defined as the ratio of length to product of width and thickness). The gradient should be numerically equivalent to the bulk resistivity in units of $10^{-3}\Omega$m. The coefficients of the best fitting line (by least squares) to the first three points and all four points are also given.

It is worth noting that it was found that these resistances were unaffected by whether the measurements were made with the lights on or off. As can be seen from these figures, the resistance has decreased dramatically, with the 475°C anneal showing the biggest fall. It was also noticed that for anneal temperatures much exceeding 500°C, the germanium resistors appeared to 'wither' away. It is not known whether this was due to evaporation of the germanium or diffusion of the germanium into the silicon nitride layer. Evaporation would seem to be an unlikely explanation as the melting point for germanium is 937.4°C [from Table 8.1.1], although the number of impurities in the germanium film may have reduced this temperature somewhat.

It can be seen from Figures 8.2.4 and 8.2.5 that in both cases the 10μm wide resistors (rightmost points) have a higher resistance than would be expected from extrapolating through the other three data points. Excluding this point gives data that appears to be reasonably well characterised by a straight line. It was possible that, because the sample was heated radiatively (so that the amount of heat entering the film was somewhat dependant on surface area), a significant percentage of the total thermal energy reaching the resistors actually arrived via the large 1000μm square pads (these present a larger energy collection area). Thus the temperature reached by the resistor would depend on the thermal resistance of the wires between the pads. Then from Equation 7.4.8 the 10μm wide resistors would have the highest thermal resistance and so could require a longer annealing time. It is noticeable that the effect seems to be more marked in the higher temperature case. However, without further tests it is not possible to verify this.

In both cases the resistance appears to have decreased by more than three orders in magnitude. The results are summarised in Table 8.2.2. In the case of the higher annealing temperature, and looking only at the first three data points, the decrease in resistivity from the as-deposited film is by more than a factor of 5000. However, this would still only reduce the resistance of the previous device from 500GΩ to around 100MΩ. Thus other measures would have to be taken in conjunction with annealing the film.

Data Points	Parameter	As-Deposited	450°C, 60s Anneal	475°C, 60s Anneal
Three	ρ (Ωcm)	264	0.0802	0.0516
Three	R_{sq} (kΩ/sq)	52,800	16.0	10.3
Four	ρ (Ωcm)	225	0.145	0.208
Four	R_{sq} (kΩ/sq)	45,000	29.1	41.6

Table 8.2.2: The effects of annealing on the resistivity of evaporated germanium.

In order to reduce the resistance to more practical levels (less than 1MΩ) it was decided that the device be patterned so as to limit the germanium component to the minimum area possible *i.e.* to the active region of the sensor. Some additional area would be

necessary to allow contact to be made with a low resistance metal layer deposited as part of an additional lithographic level. This additional metal layer would consist of the contact metal, but reduced to a thickness of 75nm [the germanium film would be 50nm thick so that an additional 25nm has been allowed to ensure electrical continuity over the overlapping region – see thermocouple development in Chapter 7 (§7.2)]. A final level of lithography would still be required to define contact pads which need to be between 200nm and 500nm thick. The germanium anneal would be carried out before the deposition of the intermediate contact metal layer.

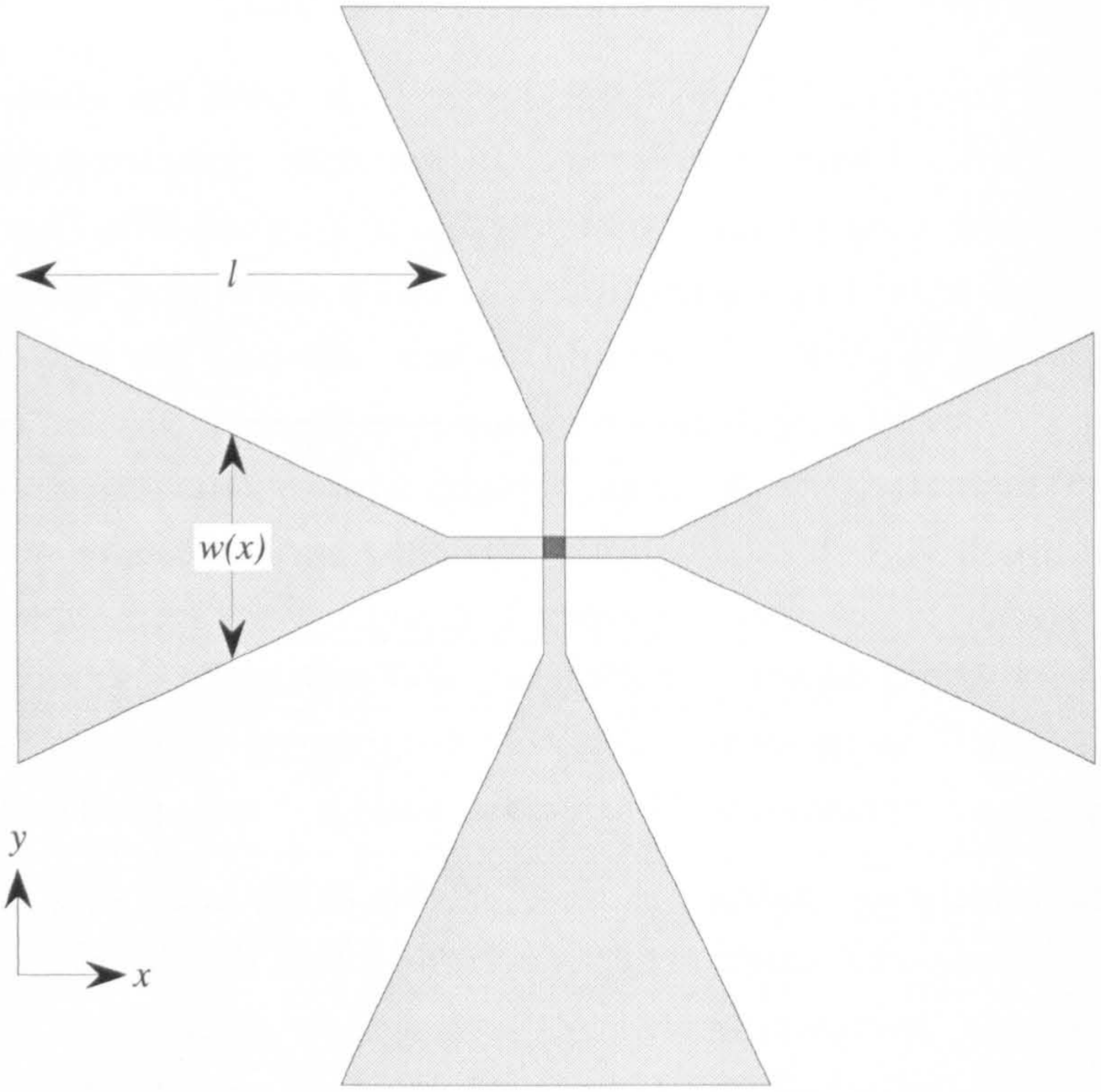


Figure 8.2.6: The geometry proposed for a lower resistance germanium Hall junction.

To ensure alignment between electron-beam lithography levels, it was decided that a maximum 1μm misalignment in any direction would be allowed for. Thus the germanium region would have to extend from the narrow sub-100nm wide wire to a region of around 2μm in both width and length to ensure that contact was made. The minimum practical length of the narrow wire which forms the active junction would be 500nm (from experience with fabricating thermocouples 300nm would also be possible, but exposure dose tolerances then become much tighter) with the wires tapering outwards at a maximum angle of around 22.5° on each side. This gives an effective growth angle (or cone angle in the terminology used in §6.5, of 45°. The geometry is shown schematically in Figure 8.2.6. To calculate the expected resistance of this structure it is necessary to determine the resistance of a linearly tapered wire. The resistance of the intermediate contact level and final contact pad levels can be neglected as the thermocouple experiments indicated that these would be of the order of 1kΩ [Chapter 7 (§7.2)].

By generalising Equation 8.2.2, the resistance of a thin film wire is given by:

$$R_{el} = \int_0^l \frac{R_{sq}}{w(x)} dx, \tag{8.2.3}$$

in terms of the resistance per square and the width of the wire as a function of position [refer to Figure 8.2.6]. Calculating this for a wire that tapers linearly from an initial width, w_a, to a final width, w_b, with $w_a > w_b$, in a distance, l, gives:

$$R_{el}^{\ taper} = R_{sq} \frac{l}{w_a - w_b} \ln \frac{w_a}{w_b}. \tag{8.2.4}$$

Then based on two tapered regions of 4μm initial width and 50nm final width over a distance of 4μm, connected together via a 50nm wide resistor of variable length, the geometric resistance (defined by setting $R_{sq} = 1$ in Equation 8.2.4) can be calculated. Table 8.2.3 shows the geometric resistance and expected resistance after annealing at 450°C and 475°C, for three different lengths of connecting resistor.

Resistivity Data	length = 300nm	length = 500nm	length = 1000nm
Geometric (Unity)	14.9	18.9	28.9
450°C (3 points)	240kΩ	300kΩ	460kΩ
475°C (3 points)	150kΩ	200kΩ	300kΩ
450°C (4 points)	430kΩ	550kΩ	840kΩ
475°C (4 points)	620kΩ	790kΩ	1200kΩ

Table 8.2.3: The expected resistances of Hall-effect devices based on 50nm thick germanium. The resistivity data used are for both fitting three points or all four points in Figures 8.2.4 and 8.2.5.

Using the resistivity calculated from fitting to the first three data points in Figures 8.2.4 and 8.2.5, it can be seen that for annealing at 475°C, the resistance could be brought down to around 200kΩ for a 500nm long resistor. Even annealing at 450°C should yield a resistance of around 300kΩ for the same length of resistor. Such resistances would be perfectly adequate from the point of view of signal processing and so these measures should be sufficient to resolve the high resistance problem. Figure 8.2.7 shows an initial design to implement this scheme, as would be applied to a non-contact mode cantilever and pyramidal tip.

However, a method for making reliable ohmic contact with the germanium film still had to be found. Initial tests based on sintering aluminium contacts revealed that aluminium 'dissolved' germanium at elevated temperatures. However, the possibility that the dopant level in the germanium film was sufficiently high for ohmic contacts to be formed without annealing was also explored. In this case it would be necessary to deoxidise the germanium films immediately before metal was deposited as part of the intermediate contact level. It was speculated that this deoxidation may be carried out as for gallium arsenide, using hydrochloric acid [Cameron[16] (1993-1996)].

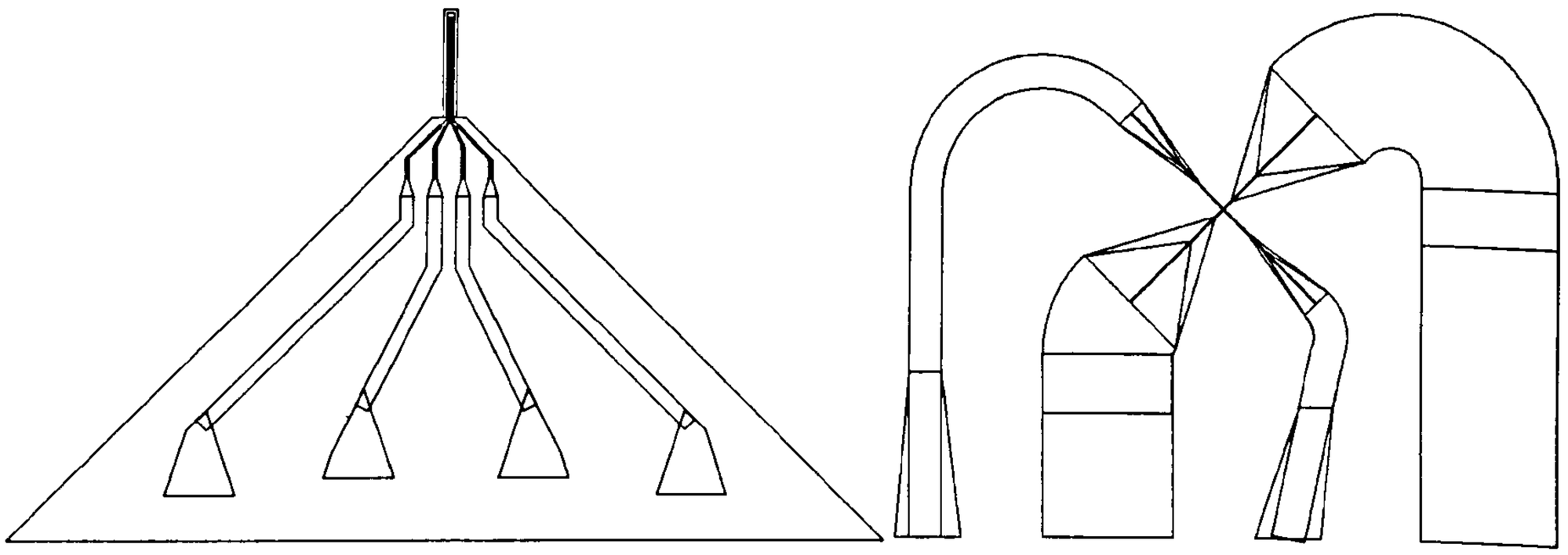


Figure 8.2.7: An initial pattern design meant to reduce the total resistance of Hall-effect probes. The picture on the left shows the way the design would be integrated with a non-contact mode cantilever. The picture on the right shows the active region of the Hall-effect probe with the narrower leads being used for voltage measurement and the wider leads being used as current leads. This helps to minimise the resistance still further by allowing the taper angle of the current leads to be increased at the expense of the voltage measurement leads.

To test this theory, some 50nm thick germanium resistors were first left in concentrated nitric acid for ten minutes. They did not visibly appear to be attacked by this treatment. The same experiment was then repeated using hydrochloric acid:water 1:1 and an identical result was obtained. Finally the resistors were placed in a solution made by mixing both of these together. The germanium appeared to dissolve readily. Due to the similarity of this solution to that which dissolves silicon [and replacing hydrochloric acid with hydrofluoric acid – see §3.4.1] it was likely that the dissolution process was occurring in a similar manner: the nitric acid oxidised the bare semiconductor while the other acid dissolved the oxide. It was therefore concluded that hydrochloric acid did to germanium what hydrofluoric acid did to silicon, *i.e.* deoxidised its surface without attacking it. It was also found that the same result could be obtained using a more dilute mixture, hydrochloric acid:water 1:4. Thus the deoxidation process decided on was a one minute dip in hydrochloric acid:water 1:4 prior to metallisation. Tests carried out on these resistors using a 200nm thick layer of palladium as the contact metal seemed to confirm ohmic behaviour when probed with the HP 4145 parametric analyser.

Unfortunately, no further experiments could be carried out to progress this device further due to lack of time. This was caused by the extended development time expended on the aperture definition procedure for SNOM probes [§6.2]. The next step would have been to test the annealing of Hall junctions using germanium confined to only a small area on a flat substrate, contacts being made by using an additional lithographic level after annealing, as stated earlier. Following this an actual flat substrate device would have been made to check the resistance and to allow initial measurements of the Hall coefficient to be made. Finally the transfer of the resulting process to released non-contact mode AFM cantilevers would have been carried out.

Without carrying out these tests all that can be said at this point is that there appeared to be no obvious reason why these devices would not work in practice.

8.3 SUMMARY AND CONCLUSIONS

Germanium was put forward as a possible material for forming Hall-effect devices. Tests were carried out to see how an evaporated thin film behaved and from resistance measurements it appeared that the deposited film was amorphous, with a prohibitively high resistivity.

Further tests to determine the effects of annealing with temperature and duration were carried out and from this it was concluded that the annealing temperature required to turn the amorphous film to a polycrystalline film was in the region of 500°C. However, there was some slight suggestion that smaller areas required a longer annealing time. From these results and by modifying the design of the Hall-effect devices it was concluded that, at least in theory, it would be possible to fabricate such devices with resistances below 500kΩ.

Tests were also carried out to determine how contacts could be made to the evaporated germanium films, from which it was concluded that no special processing was necessary, probably due to the high doping level of these films, provided that the surface oxide layer could be removed prior to evaporation. It was found that sintered aluminium could not be used as a contact for the germanium devices as it dissolved germanium. From further tests it was concluded that if the germanium films were immersed in hydrochloric acid diluted with four parts water for 60 seconds, the oxide layer could be removed without the germanium being attacked.

Finally it is interesting to speculate on the resistance that might be achieved by using polysilicon to form the Hall junction. Starting from the resistivity measurements of germanium (using 3 data points for 475°C annealing), and assuming that the film is *n*-type with $r_n = 1$, the carrier concentration can be estimated using Equation 8.1.1 to be about $3.1 \times 10^{16}/cm^3$ [taking the electron mobility value from Table 8.1.1]. [The hole concentration should then be around $1.9 \times 10^{10}/cm^3$ from Equation 8.2.1.]

This corresponds to an extremely high impurity level with a ratio of carrier concentration to intrinsic carrier concentration [Table 8.1.1] of almost 1300. If silicon was *n*-doped to the same ratio, *i.e.* to give a negative carrier concentration of about $1.4 \times 10^{13}/cm^3$, this would result in a silicon resistivity of around 310Ωcm or 62MΩ/square (again taking $r_n = 1$). This would then result in a Hall-effect device with a resistance of about 920MΩ (based on a 300nm long resistor). Even repeating this calculation, but using the highest germanium resistivity value measured, the resistance of the polysilicon device only comes down by a factor of four. Thus it would appear that polysilicon would have to be doped far more heavily to achieve the same low resistance value as achieved using germanium.

However, the main problem with heavy doping is that the Hall coefficient decreases rapidly with carrier concentration, as can be seen from Equation 8.1.7. It is known that

the Hall coefficient in silicon is maximal for a carrier concentration to intrinsic carrier concentration ratio of less than 10 [Gardner[9] (1994)]. This is also exceeded by the proposed germanium devices.

Therefore, it can be concluded that potentially the biggest problem with using polysilicon (and also possibly 'polygermanium') for the active region of these devices is the compromise between device resistance and Hall coefficient or device resistance and Hall junction size. Increasing the junction size would result in reducing the lateral resolution. The necessary compromises would appear to be smaller if germanium is used, due to its intrinsically lower resistivity, however, without measurements of the Hall coefficient for germanium, little more can be said.

8.4 REFERENCES

1. S. M. Sze, *SEMICONDUCTOR DEVICES PHYSICS AND TECHNOLOGY*, published by John Wiley & Sons, New Jersey (1985).

2. S. M. Sze, *SEMICONDUCTOR SENSORS*, published by John Wiley & Sons, New York (1994).

3. H. P. Myers, *Chapter 10: SEMICONDUCTORS* in Introductory Solid State Physics, published by Taylor and Francis, London, pp. 289-335 (1990).

4. P. S. Gleim, *McGRAW-HILL ENCYCLOPEDIA OF SCIENCE & TECHNOLOGY*, published by McGraw-Hill Book Company, New York (1987).

5. Goodfellow, *METAL DATA TABLE* in The Goodfellow Catalogue, 1993/94 edition, published by Goodfellow, Cambridge, England, pp. 17-31 (1993).

6. S. M. Sze, *Chapter 8: CRYSTAL GROWTH AND EPITAXY* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 301-340 (1985).

7. A. M. Chang, H. D. Hallen, L. Harriott, H. F. Hess, H. L. Kao, J. Kwo, R. E. Miller, R. Wolfe, J. Vanderziel and T. Y. Chang, *SCANNING HALL PROBE MICROSCOPY*, Applied Physics Letters **Vol. 61,** No. 16, pp. 1974-1976 (1992).

8. A. Oral, S. J. Bending and M. Henini, *SCANNING HALL PROBE MICROSCOPY OF SUPERCONDUCTORS AND MAGNETIC MATERIALS*, Journal Of Vacuum Science & Technology B **Vol. 14,** No. 2, pp. 1202-1205 (1996).

9. J. W. Gardner, *Chapter 8: MAGNETIC MICROSENSORS* in Microsensors Principles and Applications, published by John Wiley & Sons, Chichester, pp. 199-223 (1994).

10. H. Baltes and R. Castagnetti, *Chapter 5: MAGNETIC SENSORS* in Semiconductor Sensors, edited by S. M. Sze, published by John Wiley & Sons, New York, pp. 205-269 (1994).

11. S. M. Sze, *Chapter 1: ENERGY BANDS AND CARRIER CONCENTRATION* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 1-29 (1985).

12. S. M. Sze, *Chapter 10: DIFFUSION AND ION IMPLANTATION* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 381-427 (1985).

13. S. E. Hicks, W. Parkes, J. A. H. Wilkinson and C. D. W. Wilkinson, *REFLECTANCE MODELING FOR IN SITU DRY ETCH MONITORING OF BULK SIO_2 AND III-V MULTILAYER STRUCTURES*, Journal Of Vacuum Science & Technology B **Vol. 12,** No. 6, pp. 3306-3310 (1994).

14. A. C. Adams, *Chapter 6: DIELECTRIC AND POLYSILICON FILM DEPOSITION* in VLSI Technology, edited by S. M. Sze, published by McGraw-Hill Book Company, New Jersey, pp. 233-271 (1988).

15. S. M. Sze, *Chapter 9: OXIDATION AND FILM DEPOSITION* in Semiconductor Devices Physics and Technology, published by John Wiley & Sons, New Jersey, pp. 341-380 (1985).

16. N. Cameron, *Private Communication*, (1993-1996).

CHAPTER 9

THE FINAL ANALYSIS

This chapter summarises what was achieved by the work carried out and describes what remains to be done for all the objectives of this work to be met. It also goes on to describe some ideas for future enhancements to the probes designed here and for some entirely new probes. The wider application of the results of the work carried out here are also briefly discussed.

9.1 SUMMARY

In Chapter 1, the state of scanning probe microscopy at the time the work here was being carried out was overviewed. The precursor of all scanning probe techniques, scanning tunnelling microscopy, was described and discussed. How this then led to the development of the second scanning probe technique, atomic force microscopy, four years later in 1986, was also discussed. This technique then spawned several offshoot techniques which became collectively known as scanning probe microscopy, a concept rather than just a single technique. Some of the larger offshoots were then described and it was pointed out that one of the main problems hindering the growth and widespread adoption of many of these techniques was the availability of suitable probes. The chapter concluded with the fact that many practitioners of these techniques had to manufacture their own probes individually, and often using techniques more akin to handicrafts than sensor production, which resulted in probes of variable resolution and sensitivity. This fact justified the need for a project focused on the particular manufacturing issues relevant to the batch fabrication of nanosensors for scanning probe microscopy applications.

In Chapter 2, some of the many fabrication techniques used to produce both conventional and novel sensors were described. The merits and limitations of each technique were discussed. It was pointed out that because one of the main attributes of scanning probe microscopy was the spatial resolution achievable, and this often translated into the physical size of the active sensor area, a high resolution lithographic technique would be required to define the sensor. It was concluded that the benefits of higher resolution more than offset the additional cost of using such a technique, which would ultimately have to be passed on to any purchasers of such sensors. High volume, high resolution *x*-ray lithography was ruled out as a suitable technique on the grounds of being too inflexible for sensor development, while focused ion-beam lithography was ruled out because the required facilities were not available, leaving electron-beam lithography as the logical choice of high resolution technique.

In Chapter 3, the concept of a flexible, batch fabrication technique for producing scanning probe microscopy sensors was discussed. The idea of using atomic force microscope probes as a substrate upon which nanosensors would be defined was presented. It was explained that in order to avoid using manual alignment techniques for defining such sensors, which would have departed from the idea of batch fabrication, a substrate of atomic force microscope probes would have to be specially designed to include the alignment markers necessary for high resolution lithography. The need to compromise on the topographic resolution of these probes in order to obtain a substrate upon which nanosensors could be fabricated was also discussed. This meant that silicon micromachining levels based on photolithography would also have to be developed.

Micromachining techniques were then discussed, with particular emphasis on techniques for the bulk micromachining of silicon. [Surface micromachining was not considered due to the lack of internal polysilicon deposition facilities.] The fabrication facilities available within the department were also briefly described. Practical issues relating to electron-beam lithography on the beamwriter were discussed in some detail and some of the design rules used in electron-beam lithography were listed. Pattern transfer techniques from electron-beam exposed resist were also briefly discussed. Much of this practical information was collated from various sources within the department and was part of the learning process of the author. This information helped to define the boundaries of what would be achievable in practice and served as a solid foundation from which process development could take place.

Also discussed in Chapter 3 was how the flexibility of the developed process would be demonstrated by the fabrication of probes customised for three specific applications: scanning near-field optical microscopy, scanning thermal microscopy and scanning Hall-probe microscopy. It was envisaged that the hardest to produce would be the thermal probes, due to their fabrication requiring two levels of well-aligned lithography involving lift-off on a non-planar substrate. The scanning Hall-probes were thought to have been the next hardest due to the need for lift-off to be carried out on a non-planar substrate. The scanning near-field probes were expected to be the easiest to produce due to an absence of any obvious requirement for lift-off to be carried out.

Towards the end of Chapter 3, the results of an initial feasibility study carried out as a final year project by an undergraduate student, Shalim Rahman, was presented and the implications of these results on the processes to be developed by the author were discussed. It was concluded that the main hurdles were carrying out lift-off on pyramidal tips, increasing the quality of micromachined surfaces, coating pyramidal tips with resist and defocusing the electron beam. Following this the strategy used to develop the fabrication processes necessary to overcome these problems was discussed. It was suggested that the best approach was to carry out some 'coarse' development to help decide the general direction of development before pursuing a 'divide and conquer' strategy where each part of the process would be tackled independently, in parallel, and evolve towards a final process. The final stages of development would then have to concentrate on integrating these processes together.

The results from some initial 'coarse' development work was presented at the end of Chapter 3. The first such work concentrated on the development of through-wafer alignment markers to enable both-side wafer processing to be carried out. It was found that using silicon oxide as a masking layer for wet etching right through a silicon wafer with caesium hydroxide solution produced variable results. Part of this was caused by the silicon oxide mask being of insufficient thickness due to the elevated etching temperature used (this was done to reduce the etching time to a more practical value). The use of nickel as an etch mask in place of silicon oxide was briefly investigated and

rejected due to the lack of a suitable process for patterning nickel (wet etching resulted in severe undercutting of the photoresist masking layer). The use of nichrome as an etch mask was also investigated but was found to yield inexplicable results during silicon wet etching and was finally rejected in favour of more conventional silicon nitride mask based micromachining.

The next piece of work concentrated on evaluating how difficult it would be to fabricate adequate pyramidal tips and carry out lift-off on these. It was found that the pyramidal tips achieved using potassium hydroxide solution had an 'egg-timer' profile. These 'tips' would snap in an uncontrollable manner when the 'neck' became sufficiently thin during etching. This resulted in pyramidal tips with apices which were not centred properly, which were not of uniform height and in many cases were totally unsuitable for electron-beam lithography. Lift-off was attempted on some samples (without defocus), but failed, and alignment accuracy could not be evaluated because the tips were not centred correctly. Chapter 3 concluded with a list of problems that had to be resolved for all objectives of this thesis to be met.

In Chapter 4, all aspects of the design and engineering of the substrate were considered. The use of LPCVD silicon nitride as a masking layer for wet etching silicon with potassium hydroxide solution was demonstrated to be a reliable and stable combination for the bulk micromachining of the substrates required. The 50nm thick layer of LPCVD silicon nitride used initially was found to be too susceptible to damage by mishandling. Altering this to a 100nm thick layer was found to improve matters. It was also found that the use of a 100nm thick layer of nichrome to protect the front face of the wafer improved the situation still further. Also discussed in Chapter 4 was the design and construction of apparatus for wet etching silicon. This had to allow good control over the etch rate, particularly for the fabrication of pyramidal tips.

The photolithography levels used to pattern the backface of the substrates were described in some detail. This included practical issues relating to the patterning of the silicon nitride masking layer, for which dry etching was found to be the best method. For the first photolithography level, a method involving the use of gratings to align the mask to the crystallographically-defined edges of a whole wafer was described. For the second photolithography level, the design of alignment markers which permitted darkfield optical microscopy to be used to obtain accurate crystallographic alignment, was presented. These markers were found to survive the first silicon wet etch and proved very successful in practice.

The evolution of the mask set designs was also described in Chapter 4 (and Appendix F). This was tied in with the development of many other parts of the fabrication process, most notably that of the pyramidal tips and cantilever. As problems arose with these levels, the solutions were often found to be in the design of the photolithographic levels.

One such issue which was of particular importance was related to the sizes and shapes of through-wafer alignment markers. It was found that these turned out to be of variable size and were often rectangular, rather than square, despite measures such as the use of circular mask openings to preclude crystallographic misorientation during the first photolithography level. The problem was eventually traced to the inability of the photoresist to adequately coat the edges of these markers between the first and second photolithographic levels. This was subsequently found to have a bearing on the size of all features exposed in the first photolithographic level. The solution was found to be the reduction of the size of any such features in the first photolithographic level such that their enlargement in the second level resulted in the removal of the affected photoresist areas.

One problem which was related to the feature growth problem described above was the positioning of the pyramidal tips with respect to the back-etched device substrate. After altering the mask sets to use the method described above, it was found that the experimentally measured positions of the pyramidal tips were typically within 10μm of the theoretical (calculated) positions, with respect to the patterning on the reverse side of the device substrate.

Another problem which was of considerable importance was that of cleaving the substrate and substrate rigidity. After several iterations, the balance between the ease with which devices could be cleaved and the rigidity of the substrate (so that it would survive all the processing steps) was eventually found. This balance could only be achieved through iterative design. The use of designed crack arresters was found to be necessary to improve substrate rigidity and prevent localised damage from propagating across the substrate.

The mask set design transition from one inch substrates to whole three inch wafers was also described in detail. The designs attempted to maximise the number of devices per wafer within the constraints of: the marker layout required by the beamwriter for alignment; the holder markers required for height calibration, and; the positions of clamps required for holding the wafer in place. It was also found that ferric oxide masks made by copying chromium originals (which were patterned directly by electron-beam lithography) using contact photolithography were not sufficiently faithful to the original pattern due to the difficulty in obtaining uniform contact over the larger 4 inch area of the mask plates. As mentioned above, the first photolithography level defined features at a reduced size in comparison with the second photolithography level and so did not affect the final size of etched features. Therefore a method making direct use of a chromium mask plate for the second photolithography level only, was devised, which resulted in dramatic improvements in feature size uniformity across the wafer. The pattern for this had to be carefully constructed to enable alignment to be carried out using an opaque mask. This was greatly helped by the determination of the size and shape of the silicon wafers, iteratively, to an accuracy of a few tens of microns.

The use of convex corner undercutting to taper the edges of the device substrate so as to produce devices affording better access to the specimen by the probes was also described. This made use of the rate of corner undercutting coupled with the duration of the etch and was designed by trial and error, rather than by calculation. Another feature that was added to aid the users of these probes was a method for pseudo-kinematically mounting these devices. This made use of micromachined 'v'-grooves located on the reverse side of the device.

In the final photolithography section of Chapter 4, a set of derived design rules and a summary of the most important aspects of the process was given. The design rules were used to design the final mask sets which proved so successful in practice.

Also described in Chapter 4, was the lithographic level used to form the pyramidal tips on the front face of the devices. The first problem that had to be overcome was the automatic alignment by the beamwriter to the through-wafer 'membrane' markers. It was found that 10μm to 15μm had to be added to the optical size of these markers in order for the beamwriter to find them at 50kV. This was attributed to the sloped nature of the side-walls of these markers coupled with the penetration depth of electrons into the substrate. At 100kV, it was found that 25μm to 30μm had to be added to the optical size of the markers to accommodate the increased penetration depth of electrons at this energy.

HRN resist, whilst satisfactory initially, was eventually rejected as a dry etching mask for patterning the LPCVD silicon nitride layer used to act as a wet etch mask for defining the pyramidal tips. The variability in results achieved with this resist was attributed to its age exceeding its shelf-life. A 50nm thick layer of nichrome defined by lift-off was shown to act as a suitable dry etching mask while avoiding the use of a negative resist.

It was found that the 'egg-timer' profile of the pyramidal tips could not be avoided using potassium hydroxide solution alone. The reason for their formation was assumed to be related to the etch selectivity of different planes of silicon in this solution. Adding isopropanol to the potassium hydroxide solution was found to avoid the formation of this profile, probably by altering the etch selectivities of different planes of silicon. It was found that orienting the square silicon nitride mask approximately along the [310] axes on the silicon (100) surface maximised the mask undercutting rate. From searching the literature, it was suspected that the undercutting planes were the {313} planes.

By controlling the etching time and temperature and making use of a calibration pre-etch it was found that pyramidal tips with flat tops between 1μm and 2μm wide could be reproducibly fabricated. It was found that the calibration samples had to be written by the beamwriter on a whole wafer to obtain the required accuracy of crystallographic alignment. This was due to the specially designed whole wafer holders affording better alignment with the axes of the beamwriter than the manual clamping method used for cleaved pieces, particularly as the edges of such pieces were not as well defined as the flats of a whole wafer.

A problem relating to the number of spurious pyramids and general surface roughness produced by the tip definition etch was eventually traced to the use of buffered hydrofluoric acid to deoxidise the silicon substrate prior to etching. The reason for the roughness was shown to be due to the formation of a polymer on the surface of the silicon which acted so as to mask the silicon etch for the first few minutes of etching. The polymer was recreated by dipping a glass slide in buffered hydrofluoric acid followed by isopropanol. It was found that if unbuffered hydrofluoric acid was used instead, the polymer was not formed and surfaces with a near-mirror finish were obtained.

The design of topological markers for accurate alignment to the pyramidal tips was also discussed. It was found that the beamwriter's measurement of the size of such markers was smaller than the size obtained by measuring the distance between the physical boundaries of the markers. Markers defined by 20μm square openings in the masking layer were measured to be 14μm wide by the beamwriter. This was attributed to the sloped nature of the crystallographically etched markers. The final design used 28μm markers which were measured by the beamwriter to be 20μm wide. It was also found that automatic alignment to these markers was failing due to a combination of these markers being located in isolated mesas above the general height of the substrate, and the beamwriter using less than four backscatter detectors to find the markers. The problem was solved by forcing the beamwriter to use all four available detectors to find these markers.

Other aspects of the pattern design for this lithographic level which were discussed include: the design of front-face crack arresters (to increase substrate rigidity) which would dissolve away during the release etch; the use of 'sleeving' to shorten the writing time of all the critical edges of the alignment markers (this improves the placement accuracy of the markers with respect to the tips so that optimal alignment to the tips can be obtained in subsequent levels), and; the design of a pattern to allow in-situ calibration of the tip definition wet etch. This last idea was not tested due to lack of time.

The design of PECVD silicon nitride cantilevers for contact mode atomic force microscopy was also discussed in Chapter 4. A formula was derived for the approximate compliance expected from cantilevers of different lengths and thicknesses from which it was concluded that 250μm long cantilevers, 500nm thick, would have a compliance of about 0.05N/m and an undamped resonant frequency of around 10kHz. The silicon nitride cantilevers were patterned by dry etching via a sputtered chromium masking layer, which itself was patterned by wet etching from a PMMA mask. Initially, nichrome was used as the masking layer for the dry etch (rather than chromium), however it was found that during patterning (using chrome etch), stubborn areas would sometimes remain unetched. It was suspected that these stubborn areas were nickel rich. Also, for unknown reasons, the residual stress in evaporated nichrome layers seemed to have increased during development, often resulting in the cracking of the film. Such problems were avoided by using a 100nm thick sputtered chromium layer. The use of a sputtered

film also offered the advantages of conformal coverage which was found to be necessary when the substrate surface was particularly rough after the tip definition etch.

The need for some of the silicon nitride tapering away from the base of the cantilever to be removed to prevent a large, fragile membrane forming during the release etch was also discussed. It was also found that adding a 'shoulder' to the cantilever design helped to prevent the silicon being undercut beyond the base of the cantilever, while orienting the edges of the last part of the cantilever with the [310] axes facilitated the release of the cantilever by allowing the rapid undercutting of the silicon beneath the pyramidal tip.

It was also shown that the residual stress in as-deposited PECVD silicon nitride layers resulted in cantilevers which bent upwards upon release. This situation was avoided if 'stress-free' PECVD silicon nitride was used to form the cantilever. A discussion about the ordering of the sensor definition levels within the fabrication process concluded that any nanosensors defined on contact mode AFM cantilevers would have to be defined prior to the release etch, and would therefore have to survive it, as the cantilevers were too fragile for any processing to take place after release.

The design of solid, single crystal silicon cantilevers for non-contact mode atomic force microscopy was also discussed in Chapter 4. After a brief attempt at using dry etching to define the cantilever, an easier solution was found in wet etching the silicon. It was found that if the thickness of the silicon membrane was controlled, the cantilever could be defined by silicon wet etching using a silicon nitride masking layer in an analogous manner to the definition of contact mode cantilevers during the release etch. In this case a rectangular cantilever would be defined and due to the crystallographic nature of the silicon etch, the cantilever would have sloping sides bounded by the {111} planes. A formula was derived for the approximate compliance expected from these cantilevers as a function of length and thickness from which it was calculated that a 350μm long cantilever, 5μm thick, would have a compliance of about 6N/m and an undamped resonant frequency of around 50kHz. Such a cantilever could be defined from a solid silicon membrane twice as thick due to the silicon being etched from both sides during release. This improved thickness controllability by a factor of two over a dry etch based cantilever definition process. It was found that after the second silicon wet etch (second photolithography level), the thickness of the silicon membrane could be estimated by using an optical microscope to measure the etch depth on both sides of the substrate and knowledge of the original thickness of the silicon wafer. Then an additional, slower (lower temperature) silicon wet etch would be carried out to fine tune the thickness of the silicon membrane, prior to patterning the silicon nitride masking layer for cantilever definition. The silicon nitride masking layer was patterned by dry etching using a wet etched, sputtered chromium masking layer, as used for contact mode cantilevers. This meant that a common process could be used for both kinds of cantilevers. 250μm long cantilevers fabricated from a silicon membrane, 5μm thick, and resulting in cantilevers approximately 2.5μm thick were demonstrated using this technique.

At the end of Chapter 4, it was concluded that whilst many of the problems listed at the end of Chapter 3 had been solved, many remained, but that these mostly concerned the fabrication of the nanosensors themselves. The engineering of the substrate upon which these nanosensors would be defined had been completed to the level of a whole wafer process and it was not envisaged that anything more than minor adjustments would be required to the mask sets already designed. Thus the first part of the project had been successfully completed.

The problems of carrying out high resolution electron-beam lithography on pyramidal tips were discussed in Chapter 5. Using some simple electron optics it was shown that without defocus the spot diameter would grow by over 85nm for a 20μm focus error if the 200μm aperture was used in the beamwriter. This would mean that a 12nm spot would effectively become a 100nm spot, which would prevent the fabrication of nanosensors of sub-100nm dimensions. Thus a defocusing mechanism was necessary. It was found that there was a way of adjusting the beamwriter control software such that the focal plane of the beam could be offset by a specified amount. This was found only after several attempts to bypass or work around the misalignment problems induced by switching automatic height calibration off (this was necessary to prevent the focal plane being automatically readjusted to the level of the substrate after an offset had been added to compensate for the height of the pyramidal tips).

The problem of coating such a non-planar substrate with PMMA resist was also discussed. It was found that coating the apices of the tips could not be achieved by spin-casting, even using very low spin speeds and very thick resist. Another method known as dip-coating was attempted but also failed. It was concluded that the problem was related to surface tension (capillary forces) acting on the resist layer during the coating process. This acted so as to minimise surface energy, which meant that the resist dried out so as to avoid the apices of the tips which represented a sharp radius of curvature, and hence a high energy surface. Thus the solution found was to deposit the resist onto the substrate in a pre-dried form. This was done using a novel technique in which a drop of PMMA dissolved in orthoxylene was allowed to fall onto a dish full of water. The spreading coefficient of orthoxylene on water was such that it spread itself over the area of the dish, evaporating as it did so and leaving a dried resist layer on the surface of the water. This resist layer could be draped over the substrate, if the substrate was placed beneath it prior to film formation, by simply draining the water from beneath the film. This technique, which was the idea of Dr. J. M. R. Weaver, was named 'float-coating'.

Several observations about the 'float-coating' technique were noted in Chapter 5, particularly its dependence on the temperature of the water, the concentration and molecular weight of PMMA used and the size of the dish. In addition some undesirable aspects of 'float-coated' resist layers were highlighted, including the fact that the resist was thicker at the centre than at the edges of the dish and the presence of wrinkles in the resist layer. The thickness of the resist in these wrinkles was effectively three times

that of the rest of the resist layer. By noting that the resist wrinkles spanned out radially from the centre of the dish and dried inwards from the outer edge it was postulated that they were caused by the resist shrinking as the solvent evaporated and the resist dried.

To increase control over the formation of the resist layer a method which involved diluting the resist so that a single drop contained insufficient PMMA to result in the formation of a continuous layer was presented. This meant that a macroscopic quantity of resist was required to form the layer, removing the variability in resist thickness caused by the variability in the volume of a single drop. In addition, the resist could be spread out uniformly over the dish before drying, which made the resist formation a more static process by removing the dynamics associated with solvent evaporation rate, resist spreading rate, temperature, viscosity, etc. This also had the effect of resulting in a resist layer that was more uniform in thickness across the whole area of the dish. It was also noticed that in this case the radial nature of the location of the wrinkles disappeared, and instead localised wrinkle domains formed due to the non-uniform drying of the resist layer. It was suggested that these wrinkles could be avoided if the resist was dried from the centre outwards, by the use of a radial heating technique. Unfortunately there was insufficient time to test this idea.

Also noted in Chapter 5 were two points regarding the practical application of 'float-coated' resist layers to the pyramidal tip substrates. The first point was that the postbake had to be reduced in temperature and duration to prevent resist reflow resulting in the resist creeping away from the apices of the tips. The second point was that a layer of resist had to be coated onto the substrate by spin-casting prior to the 'float-coated' layer being deposited, because it was observed that 'float-coated' resist layers often did not make contact with the substrate around the foothills of the pyramidal tips, and that due to the resist layer being stretched over the pyramidal tips, cracks in the resist layer would sometimes appear close to the bottom of these pyramidal tips as a result of the resist layer being highly stressed.

It was finally concluded in Chapter 5 that a reliable defocusing method had been found for the beamwriter and that a method for coating the non-planar substrates with PMMA had been found, although the process was not scaleable to a whole wafer in the form presented. However, suggestions had been made as to how a uniform resist coating over a whole wafer might be achieved by slightly modifying the 'float-coating' technique. Thus all that remained was the definition of the nanosensors themselves.

In Chapters 6 to 8, the definition of the nanosensors for the probes originally specified in Chapter 3 was discussed. The first of these, discussed in Chapter 6, was the probe for scanning near-field optical microscopy. Firstly the choice of metallic shielding layer was discussed. From electromagnetic wave theory and the complex form of Fresnel's equations the reflectivity and transmittance of various metals at several important lasing wavelengths in the visible spectrum were calculated using tabulated complex refractive

index data. It was shown that aluminium was the best choice of metallic shielding layer. However, due to aluminium not being resistant to the release etch this metal could not be used in practice. Due to its availability at the time, the choice of metallic layer was eventually reduced to palladium, for which it was calculated that a 200nm thick layer had an equivalent opacity to a 100nm thick aluminium film.

The problem of defining the aperture was then discussed. It was pointed out that conventional lift-off using a positive resist could not be used to define the aperture due to a combination of the aspect ratio required and the intra-proximity effect, while negative resist did not usually result in a suitable edge profile for lift-off to succeed. Thus an alternative strategy based on defining the aperture in the silicon nitride layer, and subsequently evaporating the metallic shielding layer over it was presented.

The first attempt was based on etching the silicon nitride directly from a PMMA mask. It was shown that this generally resulted in an overcut aperture profile. An overcut profile was not desirable as this meant that the greatest confinement of light in the aperture would take place at the end furthest from the specimen. The cause of the overcut was attributed to the resist mask being attacked by the etching gases. After an attempt at changing the etching conditions this approach was eventually abandoned.

It was found that upon metallising these samples with 100nm of aluminium and 150nm of palladium, the variability of the aperture diameter at smaller aperture diameters was of the order of half the grain size of the metal. The typical size of grains in the aluminium film was about 50nm while that in the palladium film was about 20nm suggesting that more reproducible apertures could be obtained with the latter. Also noted was the increased stress in the palladium film relative to the aluminium film, which forced the silicon nitride beneath to bend upwards, or become warped, particularly at larger aperture sizes.

The design of a substrate consisting of silicon nitride beams which could be released by silicon wet etching, and upon which apertures could be defined for studying aperture reproducibility, aperture profile, aperture diameter and the effects of metallisation was described. An analysis of apertures formed by the above process, using a resist mask exposed with a varying range of doses and patterns, was carried out. It was shown that for patterns consisting of circles with a drawn diameter between 70nm and 240nm, the mean aperture diameter achieved in the silicon nitride increased linearly with drawn circle diameter. It was also found that based on three different exposure doses for these circular patterns, the aperture diameter achieved in the silicon nitride increased approximately cubically with the exposure charge, independent of pattern. For patterns consisting of drawn squares ranging from 5nm to 50nm across, the achieved aperture diameter did not vary linearly with exposure dose and no clear quantitative relationship could be established between the exposing charge and the achieved aperture diameter. The smallest apertures produced using this process were obtained from a 5nm drawn

square pattern, exposed at 100kV with a 12nm spot size and a dose of 20,000μC/cm², resulting in apertures with a mean diameter of about 60nm. It was also found that the standard deviation of the aperture diameter achieved using the drawn square patterns and drawn circular patterns of up to 110nm in diameter, was typically below 10nm except for the smallest apertures where the standard deviation almost reached 20nm. A minimum standard deviation of between 2nm and 5nm was obtained for mean physical aperture diameters exceeding 160nm. For the drawn circular patterns with diameters between 130nm and 240nm it was found that the standard deviation generally peaked at the middle dose (1000μC/cm²) of the range of three doses used.

The development of an alternative aperture etching scheme was also described. This made use of a multi-layer resist scheme based on using titanium as the intermediate 'resist' layer. The titanium was etched directly from PMMA using ECR-RIE with $SiCl_4$ at low pressure and high power. It was found that 100nm of PMMA (Elvacite) could be used to etch through a 70nm thick titanium layer. The aperture was then defined by standard RIE using C_2F_6 with the titanium layer being used as the etch mask. A groove 30nm wide and 150nm deep in silicon nitride was produced in this way with very vertical sidewalls. An analysis of some test apertures (produced using a titanium layer that was subsequently found to be insufficiently thick) showed that apertures as small as 50nm in diameter could be produced with an aperture diameter standard deviation of less than 10nm. The standard deviation of the aperture diameter decreased linearly with increasing aperture diameter to about 2.5nm at a mean physical diameter of 95nm. Unfortunately at this point the ECR machine became unavailable due to technical problems and yet another process had to be developed.

The development of this new process was then described. This process was based on using cross-linked PMMA as a negative resist for lifting-off apertures in a 50nm thick nichrome film. Lift-off was carried out using nanostrip and was successful because it was found that a bi-layer of PMMA (a higher molecular weight layer over a lower molecular weight layer) developed an undercut profile. The nichrome layer was then used as a mask to etch the aperture in the silicon nitride beneath.

Using a pattern consisting of exposing single pixels at 100kV with a 12nm spot size and 5nm resolution it was found that a dose of about 3,310,000μC/cm² was required to reliably lift-off apertures in a 50nm thick nichrome film using a spin-cast bi-layer of 2.5% BDH and 4% Elvacite, and a dose of about 5,550,000μC/cm² was required using a spin-cast bi-layer of 4% BDH and 4% Elvacite. This dose had to be exposed by using patterns consisting of multiple overlaps as the maximum practical exposure dose per overlap was found to be 400,000μC/cm² on the beamwriter. It was also found that three ten minute applications of hot nanostrip were required to ensure lift-off was successful.

Another test using a float-coated PMMA bi-layer of 2.5% Aldrich, 1.5% Elvacite showed that apertures with mean diameters ranging from 45nm to 60nm and measured standard

deviations of less than 6nm could be produced in silicon nitride using this scheme. The smallest aperture produced had a diameter of less than 40nm. It was estimated that the aperture diameter achieved in the silicon nitride was approximately 17.5nm larger than that of the pattern drawn. A test metallisation of 200nm thick palladium on this sample suggested that the aperture diameter was reduced by 70nm which implied an aperture cone angle of about 19°. This meant that an 84nm diameter aperture in the silicon nitride would result in a 50nm diameter aperture after metallisation with a 100nm thick shielding layer, and a 120nm diameter aperture in the silicon nitride would result in a 50nm diameter aperture after metallisation with a 200nm thick shielding layer.

On the same sample the post-metallisation yield of apertures with final diameters greater than 100nm was found to be about 50% and the percentage standard deviation in their diameter was typically between 5% and 10%. It was found that the yield could be improved by using three applications of nanostrip for lift-off rather than two, and a thicker float-coated PMMA bi-layer consisting of 3.25% Aldrich and 2% Elvacite. Using this resist layer it was also found that apertures created from exposing drawn circles 30nm in diameter with a total dose of 1,600,000$\mu C/cm^2$ at 100kV using a 12nm spot size had a diameter of 75nm in the silicon nitride – a good size for apertures based on a 100nm thick shielding layer.

A final observation about this aperture fabrication process was that apertures of the same size formed using smaller patterns with a higher dose were more misshapen than those created using larger diameter patterns and a lower dose. This was attributed to charging of the resist causing the electron beam to be deflected during exposure.

Attempts to make light throughput measurements on apertures produced by the above processes in silicon nitride membranes formed by etching through silicon wafers were also described. It was found that due to the highly stressed nature of the LPCVD silicon nitride layer, the membranes would fracture during the dry etching of the apertures as they penetrated the layer. Unfortunately the cause of this problem was not identified before time had run out for a test based on deposited low-stress PECVD silicon nitride.

The integration of the aperture definition process within the substrate fabrication process was also described. This was done by combining the aperture definition dry etch with the cantilever definition dry etch so that the aperture and cantilever would be etched into the silicon nitride layer using the same etch, prior to the release etch. To achieve this a 5μm square 'window' centred on each pyramidal tip would be opened in the sputtered chromium layer using chrome etch at the same time as the chromium dry etching mask for the cantilever was defined. Then an additional level of electron-beam lithography would be used to define small apertures at the apices of the pyramidal tips in a 50nm thick nichrome layer by combined positive and negative lift-off from a float-coated PMMA bi-layer of 3.25% Aldrich and 2% Elvacite (with a 12% Aldrich spin-coated bottom layer used to coat the planar parts of the substrate as described earlier). The positively exposed

resist pattern consisted of 15μm square 'windows' centred on the pyramidal tips (to limit the nichrome layer to this region), while the negative pattern consisted of drawn circles ranging from 30nm to 100nm in diameter exposed using a total dose of 1,600,000μC/cm^2 (4 overlaps at 400,000μC/cm^2 per overlap). Lift-off would be accomplished by using acetone to lift-off the positive part of the pattern followed by nanostrip to lift-off the negative part of the pattern (the aperture in the masking layer). Then dry etching would be carried out to etch the aperture in the silicon nitride layer and part way into the silicon beneath, so that during the subsequent metallisation (shielding layer) any metal deposited within the aperture would be attached to the silicon and would be removed during the release etch. Due to the problem of stress in the palladium layer causing the cantilevers to bend, the shielding layer metallisation would be limited to the tapered end of the cantilever (rather than covering the entire cantilever) using a positive lift-off step.

Due to a lack of time only a few attempts were made at fabricating fully functional SNOM probes using precursors to the process just described. The final attempt appeared to result in an aperture between 75nm and 100nm in diameter (using a palladium shielding layer 200nm thick) centred on a pyramidal tip in a cantilever consisting of 500nm thick low-stress PECVD silicon nitride. Upon examining this aperture by STEM after the release etch it was not clear whether the aperture had cleared out all the way through.

A simplistic theory for the light throughput expected from the micromachined SNOM probes described above was also developed. This was based on applying waveguide theory to a SNOM probe using a shielding layer consisting of an ideal, infinitely conductive, lossless metal. Based on these assumptions and a linearly decreasing aperture profile, it was shown that an analytical solution could be obtained. Using this formula a table was produced to compare the attenuation of the intensity of light of the least cut-off mode at a wavelength of 488.0nm as a function of output aperture diameter and metal shielding layer thickness for a cone angle of 20°. A table was also produced for the attenuation of light in a fibre probe by assuming that the same model could be applied if the refractive index of the medium was changed to 1.5 and the thickness of the metal shielding layer was made equivalent to the length of the probe over which the mode was cut-off (this is a function of cone angle). This last table was calculated as a function of output aperture diameter and cone angle. In both cases the aperture diameter was assumed to decrease approximately linearly from the point at which the mode became cut-off to the end of the probe.

From this it could be seen that for an output aperture diameter of 50nm, at a wavelength of 488.0nm, a micromachined probe with a 100nm thick shielding layer would give an intensity attenuation of 47dBs, compared with 76dBs of attenuation using a 200nm thick shielding layer. For a fibre probe with the same output aperture diameter, a cone angle of 25° would be required to match a micromachined probe with a 200nm thick shielding layer, and a cone angle of about 39° would be required to match a micromachined probe with a 100nm thick shielding layer. Several contour plots were also presented showing

the effects of various probe parameters on the intensity throughput for both micromachined and fibre probes. As expected the light throughput from these probes decreases strongly with increasing shielding layer thickness, decreasing output aperture size, decreasing cone angle and increasing wavelength. These plots can be easily used to choose and compromise between these design and operational parameters.

In Chapter 7 probes for scanning thermal microscopy were discussed. First the origin of the thermoelectric properties of metals, in particular the Seebeck coefficient (thermopower), was discussed. Some of the factors which complicate the prediction of the size and sign of the thermoelectric coefficients were briefly touched on. A plausible argument was given supporting the idea that the Seebeck coefficient of thermocouples remained relatively independent of the size of the thermocouple junction until its dimensions approached that of the mean free path of electrons in the metal, which is typically a few nanometres.

Then the fabrication of thermocouple probes was discussed. From published data two thermocouples were chosen: nickel/nichrome with a higher Seebeck coefficient of approximately 45μV/K, and gold/palladium with a lower Seebeck coefficient of about 15μV/K. These thermocouples were first demonstrated on a flat substrate from which it was concluded that if the bottom thermocouple wire was 25nm thick, and the top thermocouple wire was 50nm thick, both defined by lift-off, electrical continuity of each wire would be maintained. The resistance of the gold/palladium thermocouples was found to be about 1.5kΩ and that of the nickel/nichrome thermocouples was found to be about 7.5kΩ, at a sub-100nm junction width. Due to problems with the beamwriter, a lift-off test was carried out using manual alignment on substrates with pyramidal tips from which it was concluded that lift-off on these non-planar substrates would not be as difficult as originally thought. As the development of the SNOM probes had shown that good alignment to the apices of the tips was possible using the beamwriter, it then seemed as if the fabrication of fully functioning thermocouple probes would be relatively straightforward. Further development was passed onto others, and in a relatively short space of time, working scanning thermocouple probes were successfully fabricated.

From some simple calculations, it was shown that while the Johnson noise limited minimum detectable temperature of the gold/palladium and nickel/nichrome thermocouples were comparable (10.5mK and 7.8mK in a 1kHz bandwidth respectively), and their sensitivity to energy deposited at the junction was very similar (5.5eV and 5.9eV respectively, in a 1kHz bandwidth with a 20nm wide thermocouple junction), the thermalisation time constant of the gold/palladium thermocouple was more than seven times smaller than that of the nickel/nichrome thermocouple (0.162ns and 1.205ns in a 1kHz bandwidth respectively). This was principally caused by the high electrical resistivity of nichrome, and suggested that the gold/palladium thermocouples were a better choice for experiments in which the time dependence of temperature changes was of interest.

The design and development of another type of thermal probe for modulation differential scanning calorimetry was also discussed in Chapter 7. The probe for this technique was similar to that of a thermal conductance probe in which the sensor functioned simultaneously as a heater and, through the variation of its electrical resistance with temperature, as a temperature sensor. Palladium was chosen as a good compromise between the better heating abilities of nichrome (high resistivity) and the higher temperature sensitivity of nickel (higher temperature coefficient of resistivity). The partial fabrication of these probes was demonstrated, with the sensor/heater defined by a single lift-off level on the pyramidal tips. However, a fully functional probe was not demonstrated due to a lack of time.

Finally the design and development of probes for scanning Hall-probe microscopy was described in Chapter 8. It was pointed out that while specialised, high-mobility materials had very large Hall coefficients, they typically offered poor spatial resolution due to the large mean free path of electrons in these materials. It was decided that magnetic field sensitivity should be compromised to obtain higher spatial resolution. The use of polysilicon as the sensor material was briefly discussed, but due to the lack of internal facilities for its deposition it was not used for fabrication.

The investigation of evaporated germanium (which was readily available), as the sensor material was then described. It was found that such evaporated films were highly resistive (264Ωcm or 53MΩ/square) from which it was concluded that they were in an amorphous state, probably due to the substrate being at room temperature during evaporation. It was found that annealing germanium resistors at 475°C resulted in a greater than three order of magnitude decrease in its resistivity (0.05Ωcm or 10kΩ/square), which was presumed to be due to the recrystallisation of the germanium film. By comparing this with the resistance of the thermocouple probes, it was calculated that this would still result in a device resistance of around 100MΩ, making the accurate measurement of Hall voltages somewhat difficult and slow.

A method for reducing the device resistance by limiting the length of germanium used was devised. This scheme consisted of using contact material for as much of the device as possible and using germanium only for the sub-micron apex of the tips – the sensor region. From some calculations it was found that this would result in a theoretical device resistance of less than 300kΩ.

From some experiments carried out to evaluate the practical aspects of using germanium, it was found that hydrochloric acid and water could be used to deoxidise annealed germanium films prior to the evaporation of contacts. Aluminium was found to be unsuitable as a contact material as upon sintering it dissolved the germanium film. However, palladium was found to be a suitable contact material for the annealed germanium films.

Unfortunately, another possible problem that was uncovered was that the final resistivity of the annealed germanium films may be a function of the feature size defined in the germanium. It was suspected that this was due to the radiative annealing method used which meant that larger areas reached higher temperatures than smaller areas. The obvious solution would have been to use a vacuum oven to anneal the film. However at this point time ran out and no further development could be carried out.

Finally, from the measurement of the resistivity of the annealed germanium films, and assuming the germanium films were *n*-type, the doping level was estimated. It was then noted that if polysilicon had been doped to the same level, its resistivity would have been around 310Ωcm or 62MΩ/square and resulted in devices of much higher resistance. This suggested that germanium may be better suited than polysilicon for Hall-effect probes at the small dimensions being considered here.

9.2 CONCLUSIONS

From the last section it can be seen that while no fully functional probes were produced by the author, all major technical hurdles to the fabrication of such probes had been removed. In particular a substrate consisting of atomic force probes for either contact mode AFM or non-contact mode AFM had been successfully engineered with the required alignment markers and tips for the definition of nanosensors by electron-beam lithography. The definition of the nanosensors were the last issue to be tackled and it was quite likely that an additional six months of development work would have seen the fabrication of fully functional versions of all the probes attempted.

The development of the SNOM probes showed that the required alignment accuracy could be achieved. If the apertures could not be cleared out by dry etching through the 500nm thick PECVD silicon nitride cantilever, the silicon nitride around the tips could be made thinner by an additional level of lithography and dry etching prior to aperture definition. For the thermocouple probes no issues remained, and indeed, fully functional probes were demonstrated by others using the processes described here. The same applies to the thermal conductance probes for modulation differential scanning calorimetry. For the Hall-probes, the main issues that remained were that of carrying out lift-off on released non-contact mode cantilevers and measuring the Hall coefficient of annealed, evaporated germanium films to ensure that the magnetic field sensitivity of these devices would be adequate. The annealing problem could easily be resolved by using a vacuum thermal conductive oven rather than radiative annealing which is actually designed for annealing (large area) contacts without damaging (small area) devices on the substrate.

The development of the thermal conductance probes highlighted the flexible nature of the fabrication process designed here. In particular the design of the probes took place in a very short time and no time was spent on considering how to integrate the fabrication of these sensors with that of the rest of the substrate. This was due to the similarity

between the fabrication methods required for this probe and the thermocouple probes. It is quite probable that many sensors can be fabricated and integrated within the process developed here without much additional work as the selection of sensors chosen for initial development made use of a reasonably wide range of lithographic levels including sputtering, wet etching, chemical vapour deposition, dry etching, evaporation, lift-off and annealing.

In terms of general fabrication issues, the main outstanding one is that of extending the 'float-coating' technique to the level of whole three inch wafers. However, a plausible method for achieving this was presented and some initial test work suggested that there was a clear way forward for its development into a more reliable, controlled and reproducible technique.

It can therefore be concluded that the general objective of this work – the development of an automated, reproducible, batch-fabrication process for nanosensors customised for scanning probe microscopy – has been achieved, although the specific objectives in terms of the demonstration of customised probes for three different SPM techniques was not carried through to completion.

Finally it is quite interesting to note that the anticipated order of difficulty of fabrication of the sensors was the exact opposite of what was found in practice. This was due to the anticipated difficulty of the nanosensors being solely associated with the number of levels of non-planar lift-off required. This somewhat naive grading method failed to take account of the proper application of an engineering approach to each sensor on an individual basis: each sensor had very specific requirements which required very different kinds of development work to be carried out. For the SNOM probes, a vertical or undercut aperture profile was paramount and required significant dry etching development. For the thermocouple probes, lift-off, defocus and alignment accuracy were the key issues and the main focus of development was lift-off on a non-planar substrate. For the Hall-probes, the material system and device resistance were the main problems and required the development of an alternative semiconductor technology to that of silicon. The difficulty of fabricating any sensor should not therefore have been based on the number of novel levels required, but the amount of development time required to optimise each sensor within real-world constraints. It is rather ironic that what was perceived to be the most novel aspect of the fabrication process, non-planar lift-off, was one of the easiest problems to solve in terms of development time.

9.3 FUTURE WORK

It is clear that many possibilities for taking this work further exist. First and foremost would be the completion of the sensors attempted here. Some refinements for these sensors would be the tuning of the silicon nitride thickness at the apex of the tip of SNOM probes to maximise reflectance from the non-aperture part of the tip, and the dry

etching of the silicon nitride apices of thermocouple probes prior to release, to reduce thermal conduction between the sensor and the rest of the device.

Another enhancement which would be of great use would be the development of piezoresistive sensing technology for detecting the cantilever deflection. This would be of particular benefit to the SNOM and thermal probes. For SNOM the use of a piezoresistive sensing technique would dispense with the need for a separate laser beam for deflection detection and therefore remove a potential source of near-field signal corruption while reducing the background light level above which the SNOM signal must be detected. For thermal probes this would remove a source of heating which might affect the temperature measurements obtained. For these last probes there would then not be a need for any optics whatsoever which would mean that a very compact and inexpensive high resolution scanning thermal microscope could be designed.

The piezoresistive sensors might be fabricated using polysilicon or germanium resistors, although a method for protecting these from the release etch would have to be devised. In order to enhance the piezoresistive signal, the resistors would be made to taper as they progress along the length of the cantilever so that the greatest resistance is concentrated where the largest bending moment is applied.

Many other sensors could also be attempted. One such sensor would be for scanning capacitance microscopy. This would be a shielded (coaxial) capacitance probe, which would reduce the effects of changes in the capacitance between sample and cantilever on the measured change in the capacitance between tip and sample.

One idea for achieving this might be based on the fabrication process used for SNOM probes. The starting point would be a non-contact mode AFM cantilever with a silicon nitride insulating layer deposited upon it. A first level of metallisation would define a wire crossing over the apex of the tip. The metal at the tip would actually take the form of a large 2μm pad, upon which a sharp metal tip would be defined later. This metal layer would consist of 5nm of titanium, 20nm of gold and 10nm of nichrome. A contact pad area for this metal layer would also be defined although the contacts would be made during a later level.

100nm of silicon nitride would be deposited upon this metal layer for insulation. Then using combined positive and negative PMMA lift-off, a nichrome masking layer would be patterned on top of the silicon nitride layer. This pattern would open up windows over the silicon nitride immediately above the contact pad area for the first metal layer by positive lift-off and open up a small, 100nm diameter aperture on the apex of the tip by negative lift-off. The silicon nitride would then be dry etched as for SNOM probes, exposing the contact pad area of the first metal layer and a small circular part of the first metal layer on the apex of the tip. The nichrome top layer of the first metal layer would act as an effective etch stop for this process. Then all the nichrome would be stripped off using

chrome etch, leaving a gold contact pad region and gold circular region at the apex of the tip as the only exposed metal.

Finally the ground plane of the device would be deposited on top, in the same manner as the shielding layer for the SNOM probes. Due to the undercut profile in the silicon nitride, a sharp metal tip would form at the apex of the pyramidal tip, recessed into the silicon nitride layer. On top, the metal layer will taper inwards leaving only a small aperture through which electric field lines can pass between the metal tip and the sample. Thus the aperture in the metallic shielding layer as used in the SNOM probe would function as a focusing mechanism for electric field lines between the sample and a partially buried metal tip for a shielded capacitance probe.

Another idea would be a scanning permeability probe. This would consist of a small nanocoil defined at the apex of the tip of a non-contact mode AFM cantilever. This nanocoil would act as a nanomagnet and would be buried beneath a thin insulating layer of silicon nitride. Then a Hall-probe sensor would be fabricate on top of this coil. This Hall-probe would be capable of measuring the amount of magnetic field passing through the Hall junction between the magnetic nanocoil and sample and so give a measure of the degree of flux linkage between tip and sample. This is an example of an active probe which provokes a response from the specimen and measures it.

Other ideas for sensors include the use of spatially displaced thermocouple sensors on a single tip for measuring thermal gradients and the addition of a heater to this configuration to allow lateral heat conduction to be measured.

Finally, it is also clear that the 'float-coating' technique, once developed as explained earlier, would have applications in the fabrication of other devices unrelated to scanning probe microscopy. For example, on one occasion where a substrate containing a large 5mm hole was used, it was observed that a 200nm thick 'float-coated' PMMA layer spanned the gap. This means that suspended resist masks are made possible by this method. Such a technique could be used to fabricate air bridges or lateral metallic wave guides. If the technique could be extended to photoresist then the whole field of micromachining would have access to a non-planar resist coating technique with which even more complex sensors and devices may be fabricated.

APPENDIX A

SILICON WET ETCHING IMPLEMENTATION

A.1 DESIGN OF THE ETCHING KIT

Using 7 molar (+5%) potassium hydroxide solution, the etch rate for the {100} planes at 80°C and 105°C were established as 75±1μm/hour and 255±1μm/hour respectively (measurements taken for etching using bi-directional stirring at 50 rpm). [Note that this etching solution has a boiling point of around 110°C.] This gives an Arrhenius equation (Equation 3.4.1) fit for the etch of:

$$R\{100\} = 8.212 \times 10^9 \exp\left(-\frac{e0.5634}{kT}\right) \mu\text{m / hour}, \tag{A.1}$$

where $R\{100\}$ is the {100} etch rate in microns per hour, e is the electronic charge, k is the Boltzmann constant and T is the etching temperature. From this it can be seen that the change in etch rate, $\Delta R\{100\}$, for a small temperature change, ΔT, is given by:

$$\frac{\Delta R\{100\}}{R\{100\}} = \frac{e0.5634}{kT}\frac{\Delta T}{T}. \tag{A.2}$$

Therefore to control the etch to within 1% requires control of the temperature to within about 0.2°C at 80°C. At this temperature, there is a high rate of water evaporation so that a refluxing system is required to maintain the concentration of the etching solution. Thus an etching kit was designed on the basis that 3 inch wafers would eventually be processed. As this apparatus was going to be installed in a class 100 cleanroom in a fume cupboard alongside other equipment and facilities, stringent safety criteria had to be met by the control system.

The eventual solution utilised a 200W Electrothermal EMA500 heating mantle with built in magnetic stirring facility. As this did not provide closed loop temperature control, a Eurotherm 808 temperature controller with a triac output, proportional, integral and derivative (PID) control and auto-tuning facility, was used to control the heating mantle. In order to do this the stirring circuit within the EMA500 had to be separated from the heating circuit. The temperature sensor chosen for the Eurotherm was a 4-wire 100Ω platinum resistor because of its stability in comparison with most thermocouples [Gardner[1] (1994)]. This sensor was constructed from parts obtained from RS (catalogue number 164-794 and 150-054). Using this sensor in a 3-wire arrangement the Eurotherm controller allows the temperature to be set to 0.1°C with a specified temperature stability of 0.03°C. The heating mantle and temperature sensor is shown in Figure A.1.

The glassware consisted of several Quickfit components: a 500ml borosilicate glass reaction flask (Quickfit number FR500F); a 75mm flask cover (Quickfit number MAF1/75); a 300mm long condenser (19/26 fitting, Quickfit number C3/22/SC); a condenser clamp; jubilee clips and various other accessories. This would allow the etching of substrates of up to about 24mm across (the largest that would fit through the flask cover). The assembled glassware is shown in Figure A.1.

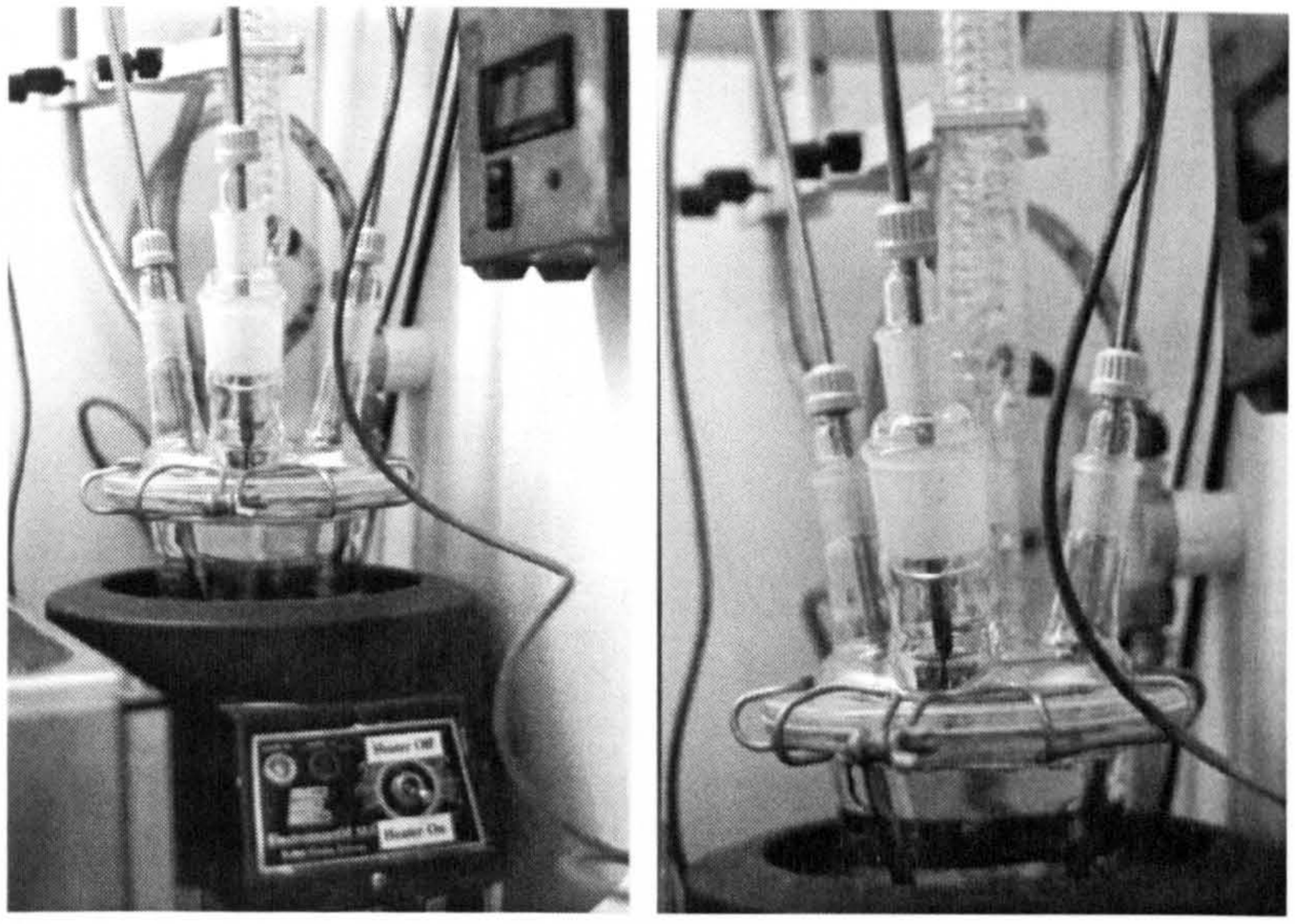


Figure A.1: Photographs showing the glassware, heating mantle and temperature probes assembled together for wet etching silicon.

For 3 inch wafer etching the spherical 500ml reaction flask was replaced with a cylindrical 700ml borosilicate glass reaction flask (Quickfit number FR700F) with the same size of footprint as the heating mantle, so that there was enough etching solution to submerge a 3 inch wafer vertically, and a 100mm flask cover (Quickfit number MAF3/52). However, this flask cover would not allow the passage of a 3 inch wafer through it and the glass-blower of the chemistry department could not produce a suitable flask cover of the geometry required in the space available, so that a custom flask cover made from a corrosion resistant grade of stainless steel (grade 316 S31 which contains molybdenum for extra corrosion resistance) was designed and manufactured by the departmental mechanical workshop (refer to Appendix B).

For holding the wafer in the solution, a PTFE wafer holder was designed. Originally this could hold a one inch wafer, but several generations later, a wafer holder capable of holding up to eight one inch substrates (by passing through the slot used for 3 inch wafers) was also designed (refer to Appendix C). For 3 inch wafer etching, a commercially available two inch wafer dipper was slightly modified so that it could be suspended in the etching solution (Fluoroware PFA Dipper, catalogue number D11-0215). Both the PTFE wafer holders and the stainless flask cover are shown in Figure A.2. The resulting assembled apparatus is shown in Figure A.3.

For safety reasons a flow sensor and a solenoid valve (RS catalogue numbers 257-149 and 342-023 respectively) were used to provide automatic water flow control and shutdown in the event of a leak. To protect against temperature sensor failure a second independently monitored platinum resistance thermometer was used to detect and supply an over-temperature signal. [The temperature monitoring circuit was based on a Kelvin double bridge measurement of the platinum resistor and was designed by Dr. J. M. R. Weaver.] An diagram of the safety system, which was implemented by the departmental electronic workshop, can be found in Figure A.4.

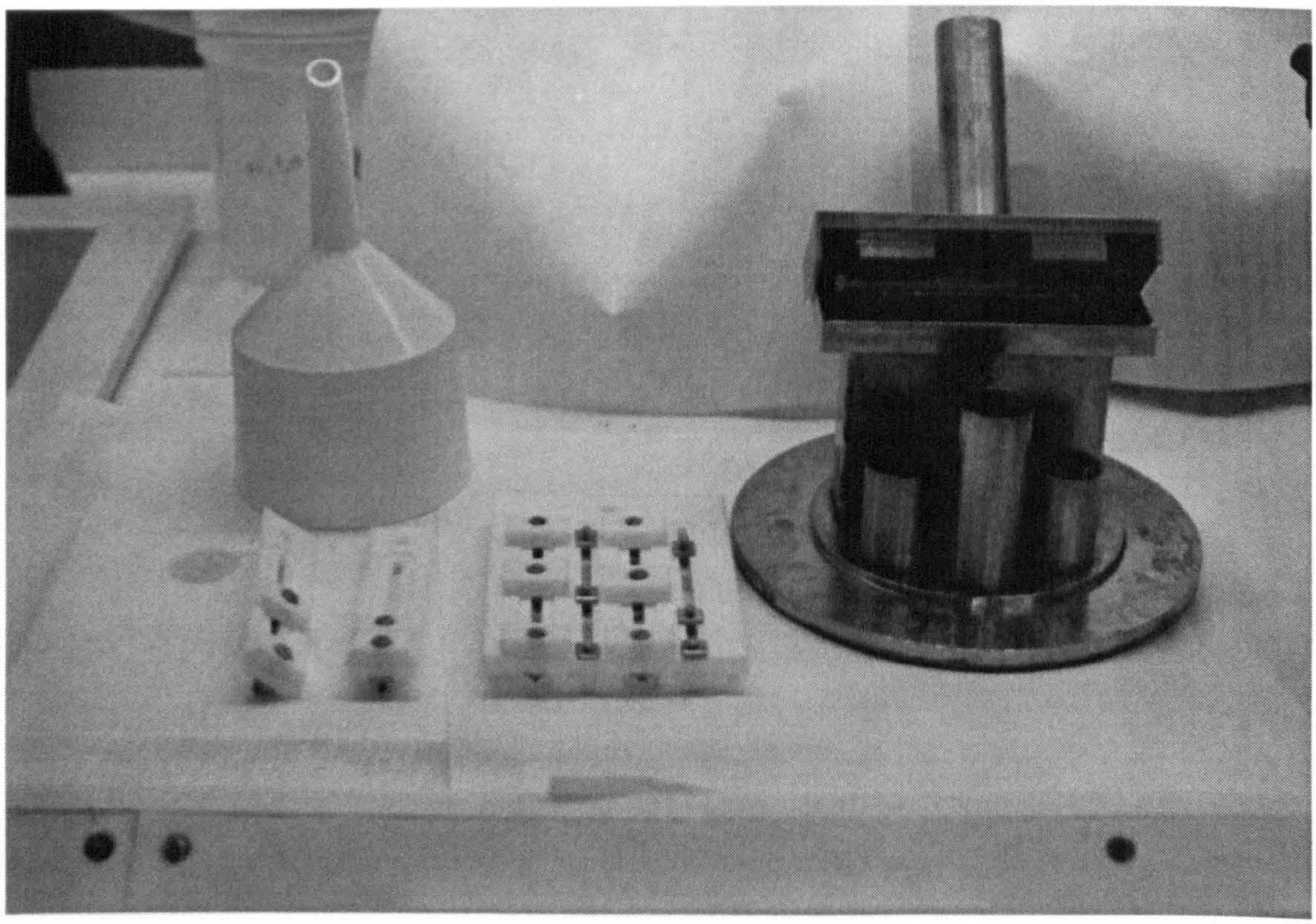


Figure A.2: Photograph illustrating the PTFE wafer holders (left) and stainless flask cover (right) manufactured for the wet etching of silicon. Two single substrate PTFE holders are shown (extreme left) as well as a multi-substrate holder (middle) capable of holding up to eight substrates. This last holder can only enter the solution via the stainless flask cover (right) which is also used to pass in whole 3 inch wafers for etching.

It should be mentioned here that an alternative strategy followed by some groups for micromachining silicon is to etch at the boiling point of the solution so that the temperature self-regulates (provided water loss from the system is minimised). This is ideal if an etch stop layer is used as the etch rate tends to be very high at such temperatures. However, for controlled etch depths without an etch stop layer, *i.e.* timed etches, it is sometimes necessary to slow the etch rate down by reducing the temperature during critical etching steps.

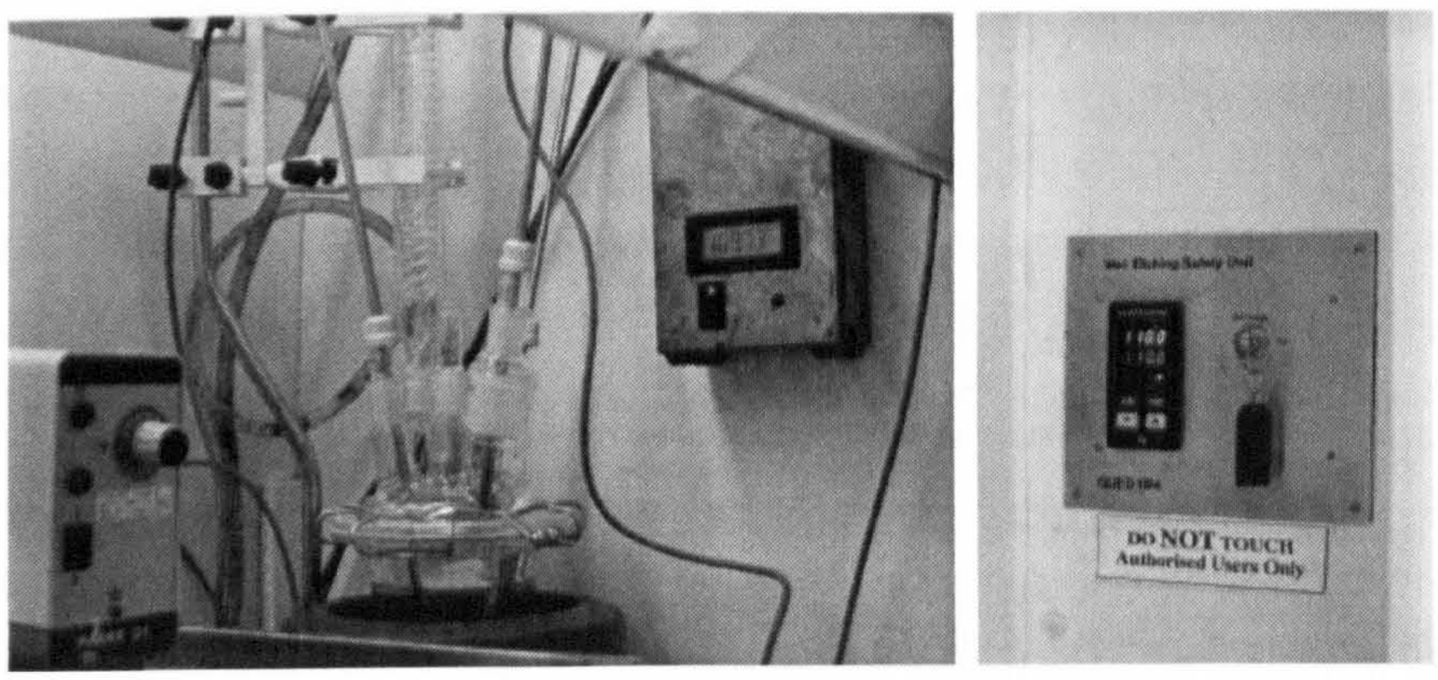


Figure A.3: Photographs illustrating the independent temperature readout (left) and temperature controller unit (right) used to wet etch silicon. The temperature controller unit was integrated into a general safety design unit monitoring water flow through the reflux system and a separate over-temperature signal from a second platinum resistance temperature probe (as can be seen on the left).

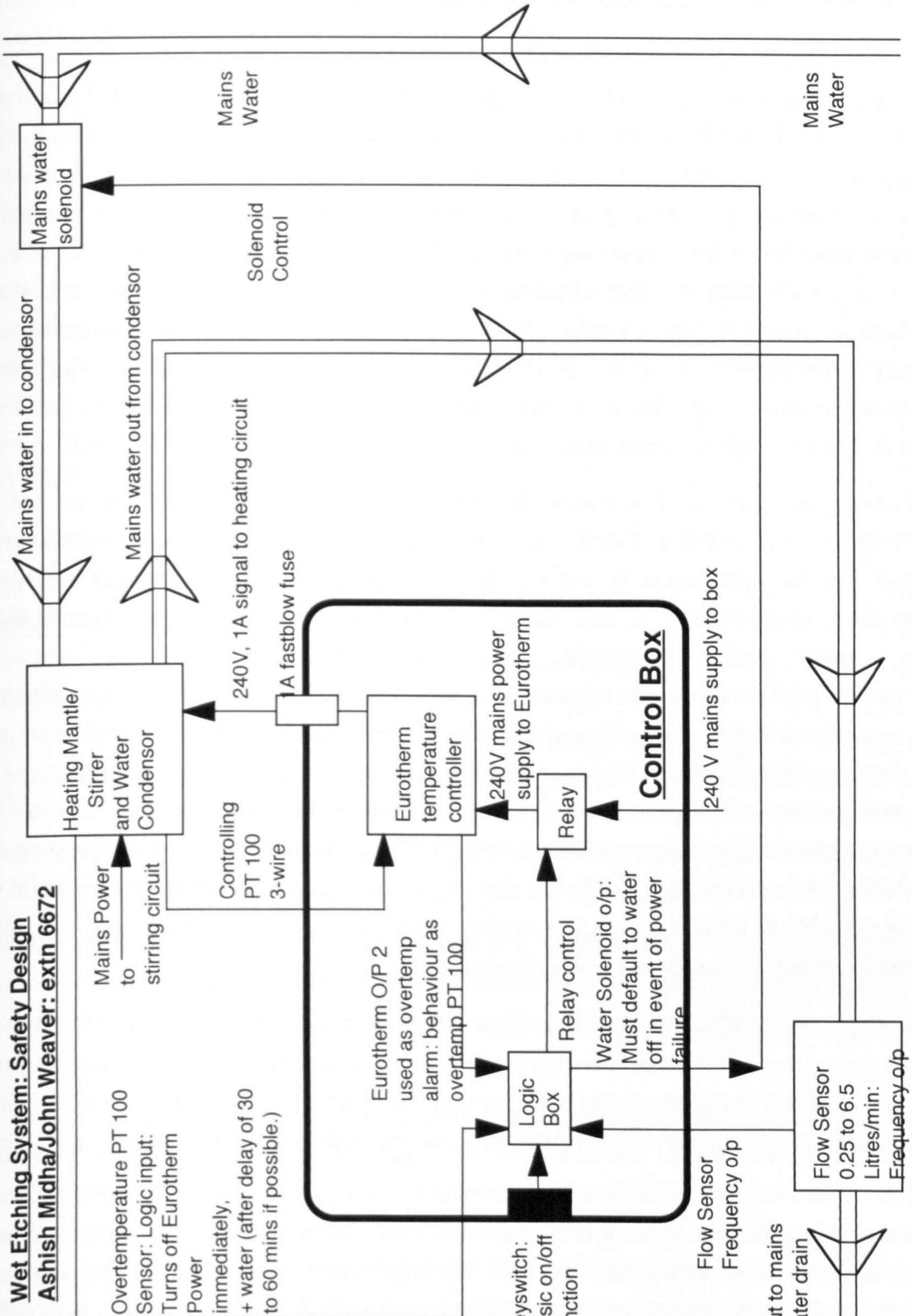


Figure A.4: Block diagram of the wet etching safety system.

A.2 PRACTICAL SILICON ETCHING

The main cause of failure of the wet etch is contamination of the etching solution. Thus from a practical standpoint all the etching apparatus must be kept free from sources of contamination. This always raises the problem of finding a cleaning agent which itself does not alter the etch. It has been found that acetone works well in this role. Therefore, prior to etching, the glassware (and metalware) is thoroughly rinsed in acetone and blown dry. Acetone is very volatile and evaporates readily so that the likelihood of contamination is much reduced. After etching all parts are rinsed thoroughly in ultra-pure water and the reaction flask is left standing in the heating mantle, full of water. This helps to keep the flask clean whilst diluting any residual etching solution. The rest of the larger parts, excluding the condenser, are then stored in a heated cabinet, while smaller parts are stored within the reaction flask, submerged in water.

The etching solution itself is stored in fluorinated high density polyethylene bottles (HDPE) obtained from BDH (1 litre, 2 litre and 4 litre U. S. D. O. T. approved fluorinated Nalgene bottles, catalogue numbers 215/0438/06, 215/0438/08, 215/0438/10 respectively). When such a bottle is used for the first time, the etching solution tends to change from being clear to taking on a brownish tint. This may be related to a problem discussed in §4.3.1. The exact reason for this is not known (the bottles are certified as being capable of holding very strong alkalis), but this does not seem to affect the etch rate, although the etch selectivity may be affected. To avoid this, each new bottle is filled with dilute ammonium hydroxide solution for two weeks before being used to store the etching solution. This has the added advantage that any leeching of the plasticiser occurs in the ammonia solution and not in the etching solution [Glydle[2] (1993-1996)]. Before the etching solution is removed from the reaction flask, it is allowed to cool with the reflux system operating (to reduce water loss by evaporation).

The use of a range of bottle capacities from 1 litre to 4 litres means that the bottles can be kept sealed from the atmosphere (to reduce the absorption of carbon dioxide) over long periods. Solutions being currently used are kept in 1 litre bottles and are replaced with solution stored in 2 litre bottles which in turn are replaced with solution from 4 litre bottles.

By taking all these precautions it is found that the same solution can be reused over several months and tens of etches with no significant change in the etching rate. However, as a precaution the etching solution is replaced every three months, to preempt any concentration change effects due to the loss of water by evaporation or presence of silicon etching reaction products. The etch rate has remained constant even between batches of etching solution.

Occasionally a catastrophic failure means that the solution is replaced prematurely. Usually this is indicated by a change in colour, a sudden and inexplicable change in etch

rate, or another visual indicator (*e.g.* when pieces of wafer or metal are found floating in the etchant after a mishap).

When old etching solution is disposed off, the bottles are rinsed thoroughly in ultra-pure water, before being left full of ultra-pure water for several days. After another thorough rinse, the bottles are allowed to dry in a heated cabinet after which they are ready for storing new etching solution. Manufacture of the etching solution is carried out using accurate weighing scales, although the most reliable method is to use all the potassium hydroxide pellets to make a whole batch of etching solution at one time. The pellets must be added reasonably slowly to the water to avoid excessive heating. To help deal with this problem, the solution is made up in a large conical flask which is partially submerged in a sink full of cold mains water. This acts as a coolant for the flask during the dissolution process. To speed dissolution, the pellets are added at a rate which heats the flask up to about 50°C, aided by vigorous stirring.

For the best results, the wafers must be deoxidised immediately before being placed in the etching solution (which should already be at the right temperature). This removes the native silicon oxide layer, which is a few nanometres thick and can act as a mask to the etch for a short time, resulting in increased surface roughness [Bassous[3] (1978)]. This is particularly important in short etches, such as the tip definition etch, where convex corner undercutting may not be given sufficient time to flatten out any roughness after the silicon oxide has been etched away. Originally, all wafers were deoxidised in silicon dioxide etch 4:1 for one minute, however, for reasons discussed in §4.3.1, this was eventually changed to one minute in unbuffered hydrofluoric acid diluted to one part in four of water without wetting agent.

After this step the wafers are placed in a beaker of warm water at 50°C to 60°C for a few minutes prior to being placed in the etching solution. This helps to reduce the thermal shock to both the etching solution and the wafer so that the etch depth can be more accurately controlled.

The wafers are mounted vertically in the solution to avoid the problems of trapped gas bubbles. This gas is hydrogen [Gardner[4] (1994)] and is one of the reaction products of silicon etching. As expected, faster etching leads to a greater number of larger bubbles. To help maintain temperature uniformity without introducing diffusion gradients, the solution is normally stirred at a very slow rate, 50 rpm, using a magnetic bead (coated in PTFE). This stirring is in a bi-directional mode which changes the stirring direction every 30 seconds by slowing the rate down gradually and eventually spinning up in the opposite direction. These temperature gradients would be more of a problem with larger substrates at higher temperatures, however no significant non-uniformity has been observed with whole 3 inch wafers at temperatures up to 105°C. This may be due to the hemispherical coiling of the heating element which reduces the area of the sides of the reaction flask exposed to the environment.

To avoid the problems of crystallisation occurring when the wafers are removed from the solution and dried (as mentioned in §3.7.3), the wafers are directly transferred to a neutralising solution, either dilute hydrochloric or sulphuric acid, for one minute before being rinsed with water and dried. [If a nichrome protection layer is used, as mentioned in §4.2.1, the neutralisation acid should be dilute sulphuric acid as dilute hydrochloric acid attacks nichrome much more rapidly by comparison.] An additional benefit of this neutralisation step is that the number and size of darkened areas (the dark colour of these areas is due to a high degree of surface roughness), which seem to appear when small isolated drops of alkali are allowed to remain on the substrate for some time while cooling, are reduced. In effect, this step stops any further etching immediately as the neutralisation reaction is fast. It should be noted that transferring the wafer directly to either concentrated acid results in the acid splashing out due to the heat generated, so that only diluted acids should be used. A concentration of 1 part acid to 4 parts water seems to work well in practice.

A.3 REFERENCES

1. J. W. Gardner, *Chapter 5: THERMAL MICROSENSORS* in Microsensors Principles and Applications, published by John Wiley & Sons, Chichester, pp. 79-116 (1994).

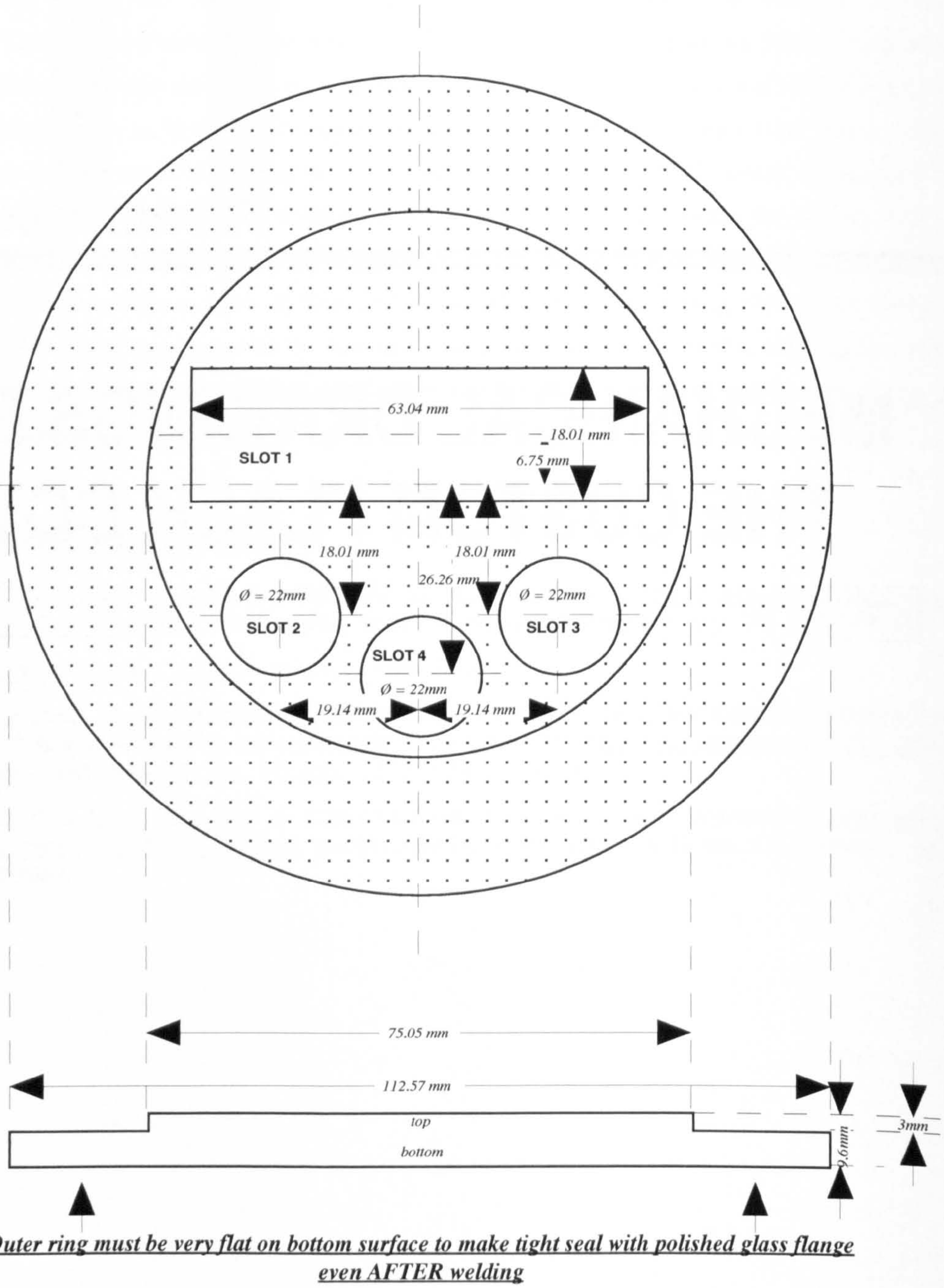
2. A. Glydle, *Private Communication*, (1993-1996).

3. E. Bassous, *FABRICATION OF NOVEL THREE-DIMENSIONAL MICROSTRUCTURES BY THE ANISOTROPIC ETCHING OF (100) AND (110) SILICON*, IEEE Transactions On Electron Devices **Vol. ED-25,** No. 10, pp. 1178-1184 (1978).

4. J. W. Gardner, *Chapter 3: CONVENTIONAL SILICON PROCESSING* in Microsensors Principles and Applications, published by John Wiley & Sons, Chichester, pp. 36-57 (1994).

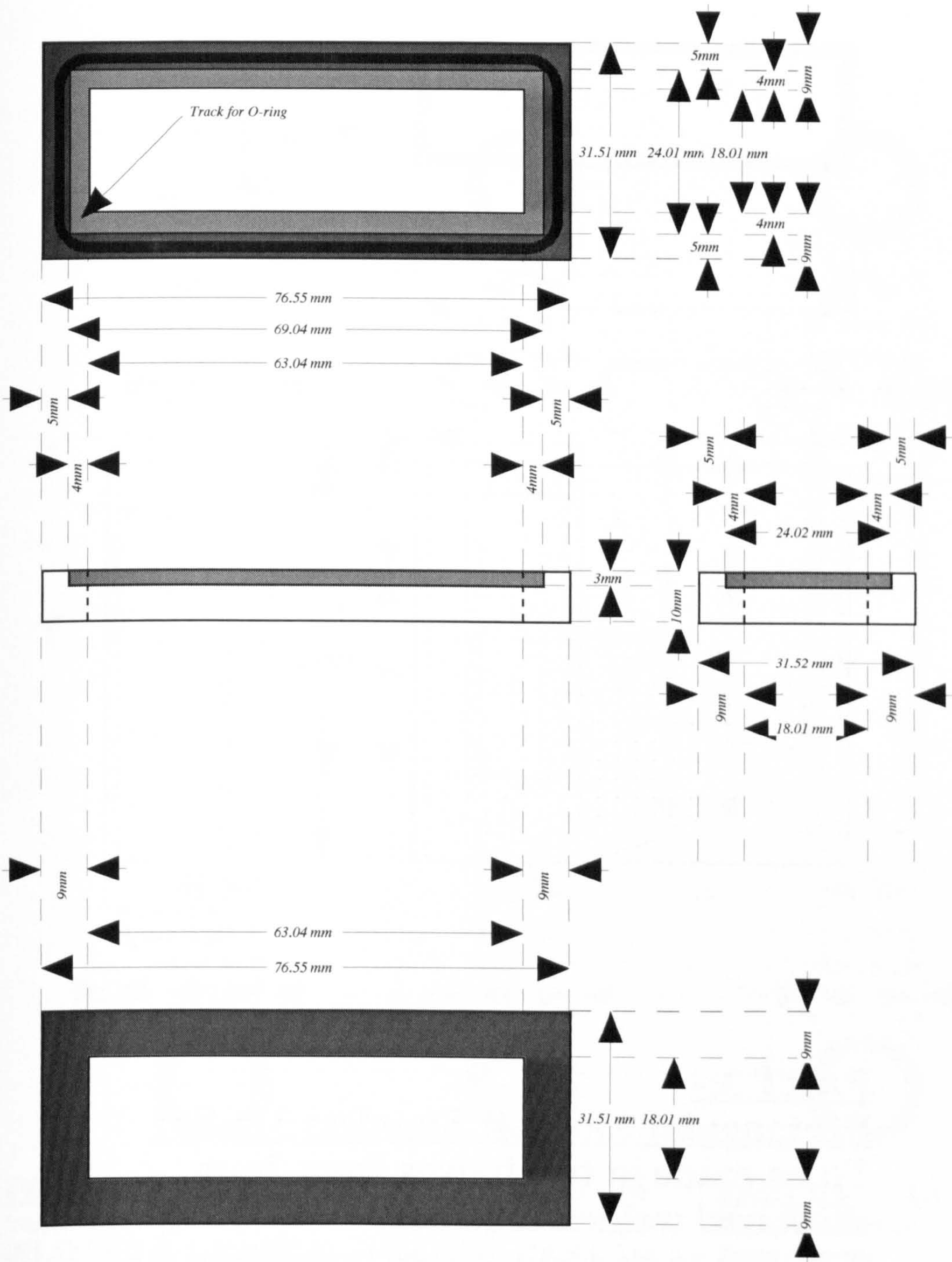
APPENDIX B

MECHANICAL DRAWINGS FOR THE CUSTOM FLASK COVER



Outer ring must be very flat on bottom surface to make tight seal with polished glass flange even AFTER welding

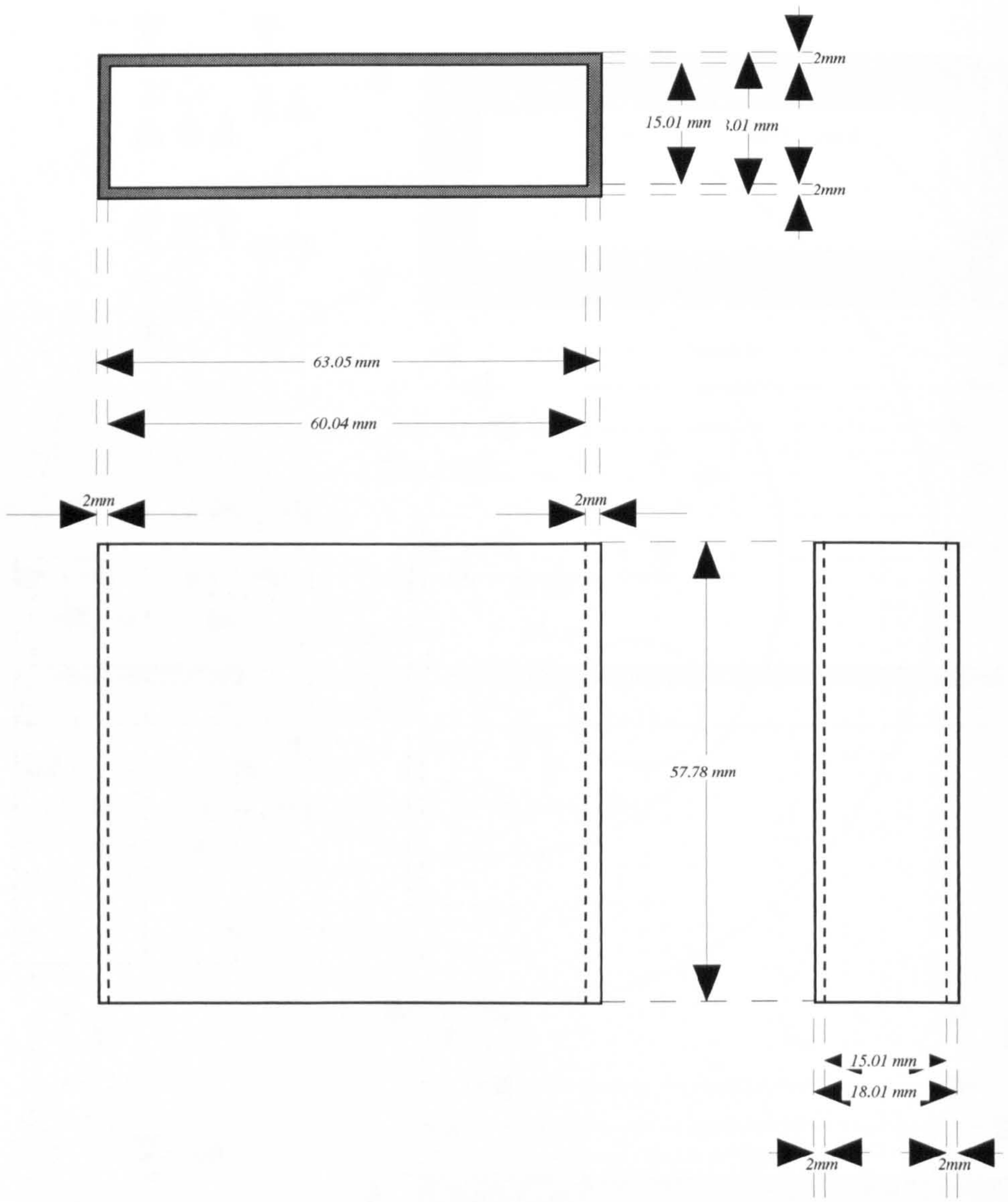
PART 1:-
Base Plate x 1: Stainless 316 S31
Holes to be made in 9.6mm plate



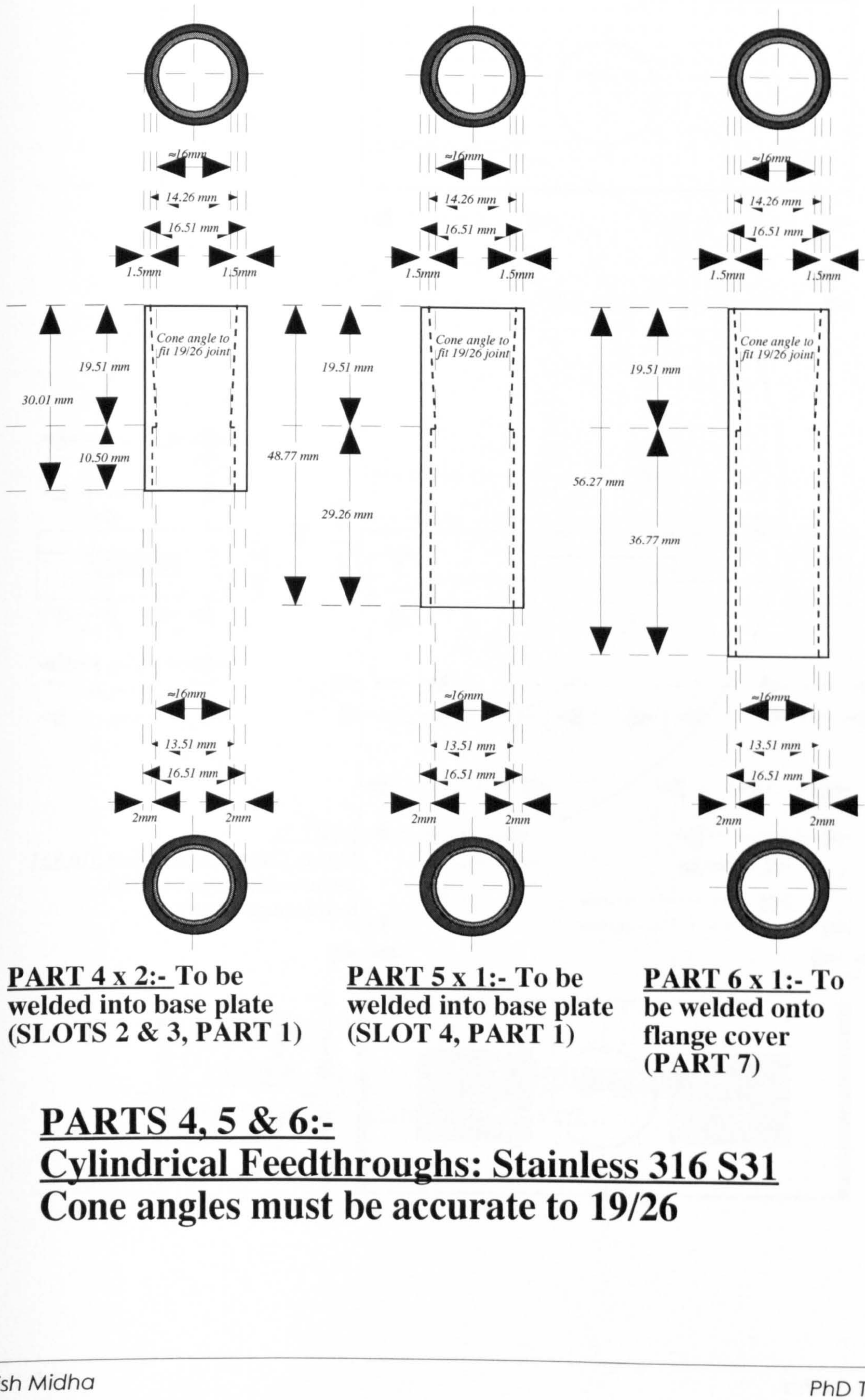
PART 2:-

Rectangular Flange x 1: Stainless 316 S31

To be welded onto rectangular slot (PART 3)



<u>PART 3:-</u>
<u>Rectangular Slot x 1: Stainless 316 S31</u>
To be made in two halves from 2mm sheet and welded together.
One end to be welded into base plate (SLOT 1, PART 1), other end to be welded together with flange (PART 2)

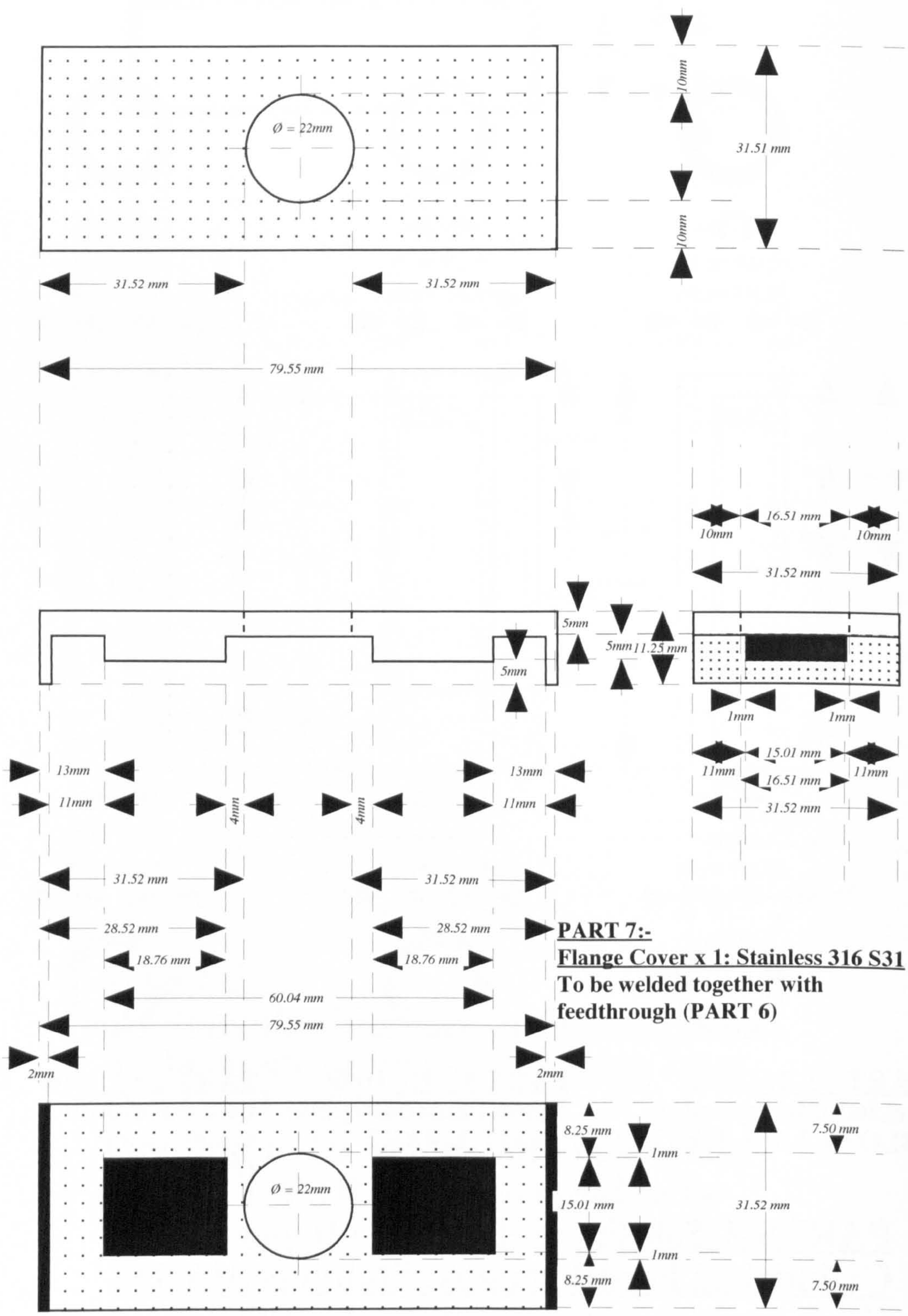


PART 4 x 2:- To be welded into base plate (SLOTS 2 & 3, PART 1)

PART 5 x 1:- To be welded into base plate (SLOT 4, PART 1)

PART 6 x 1:- To be welded onto flange cover (PART 7)

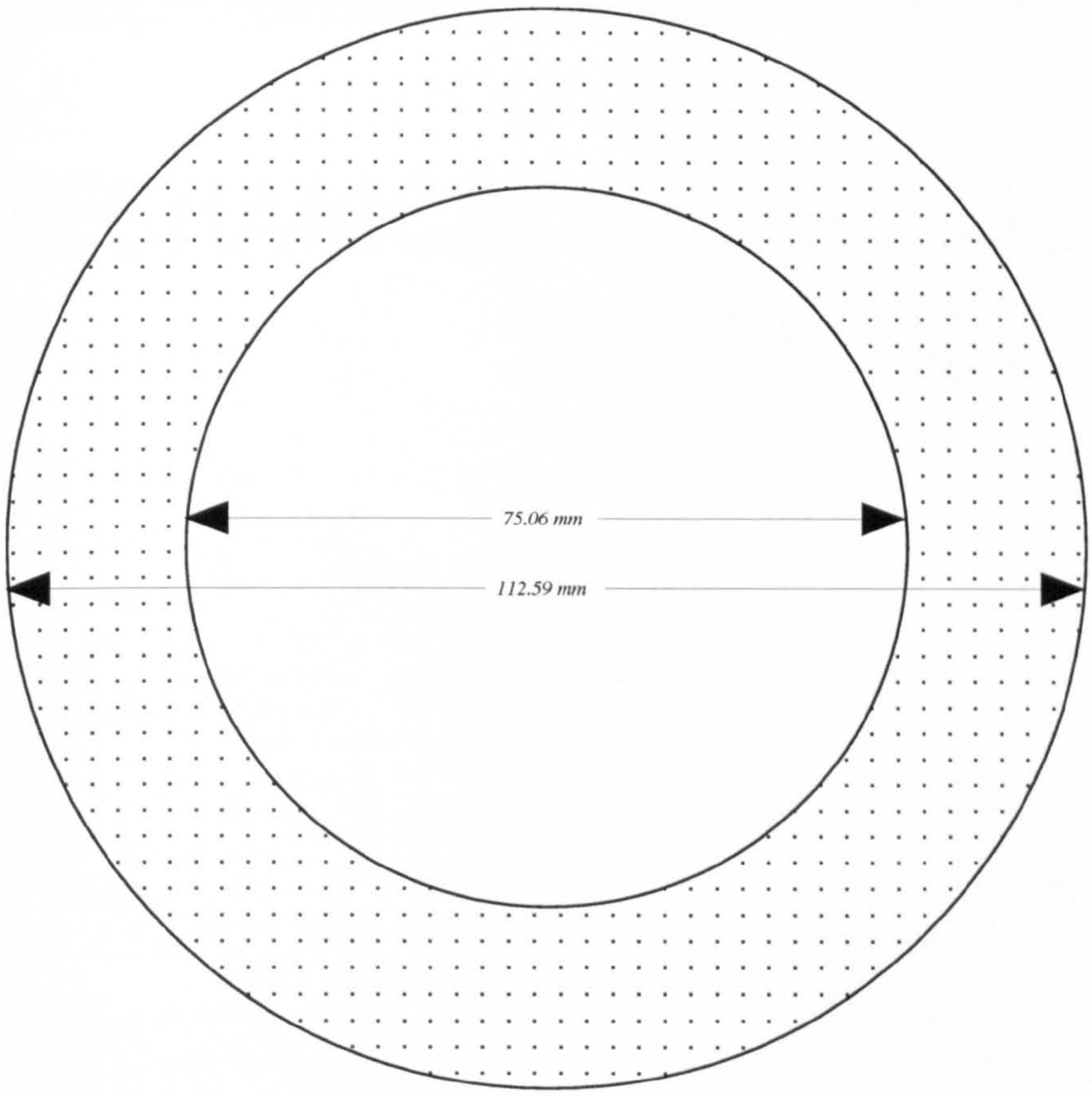
PARTS 4, 5 & 6:-
Cylindrical Feedthroughs: Stainless 316 S31
Cone angles must be accurate to 19/26



PART 7:-

Flange Cover x 1: Stainless 316 S31

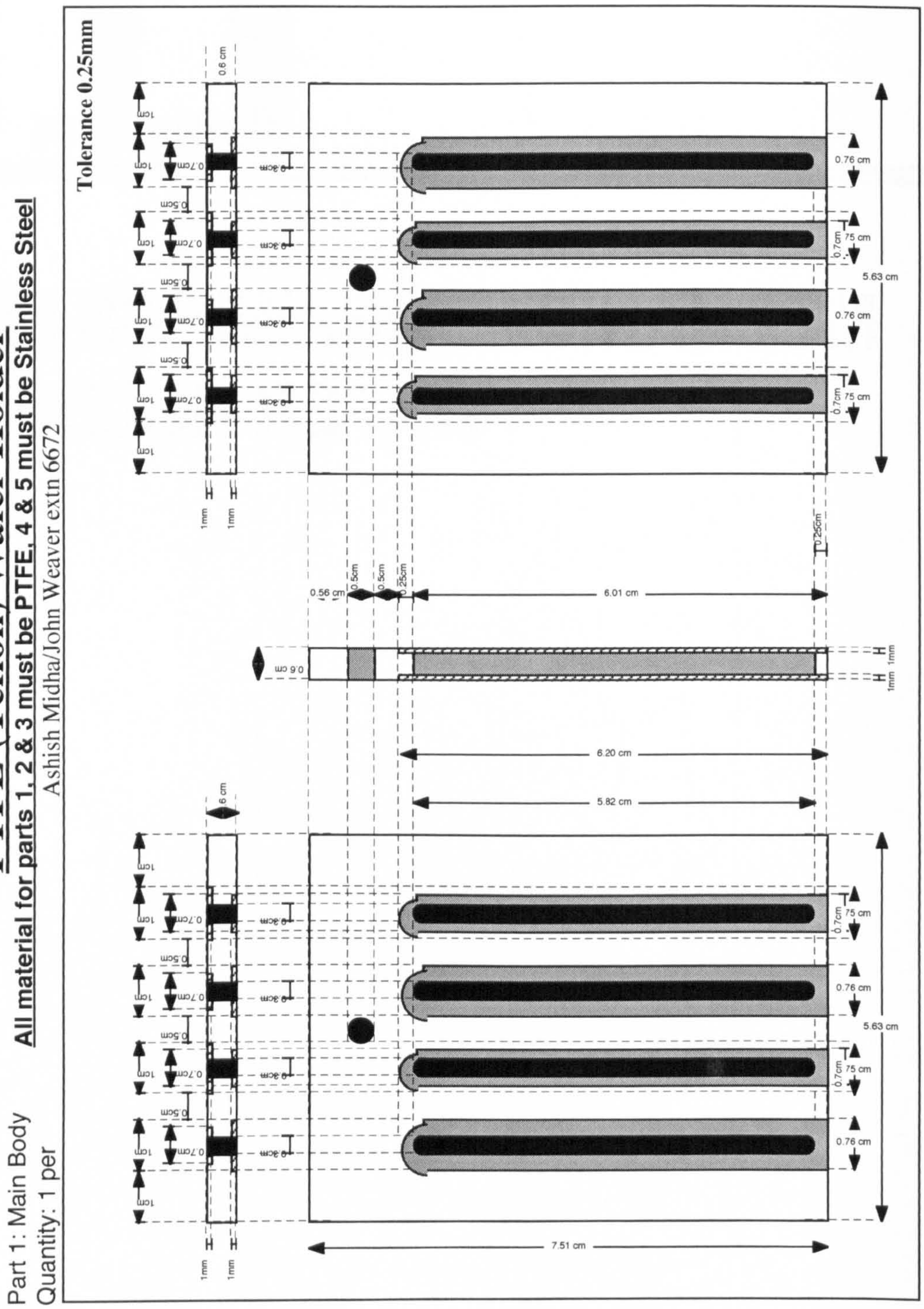
To be welded together with feedthrough (PART 6)



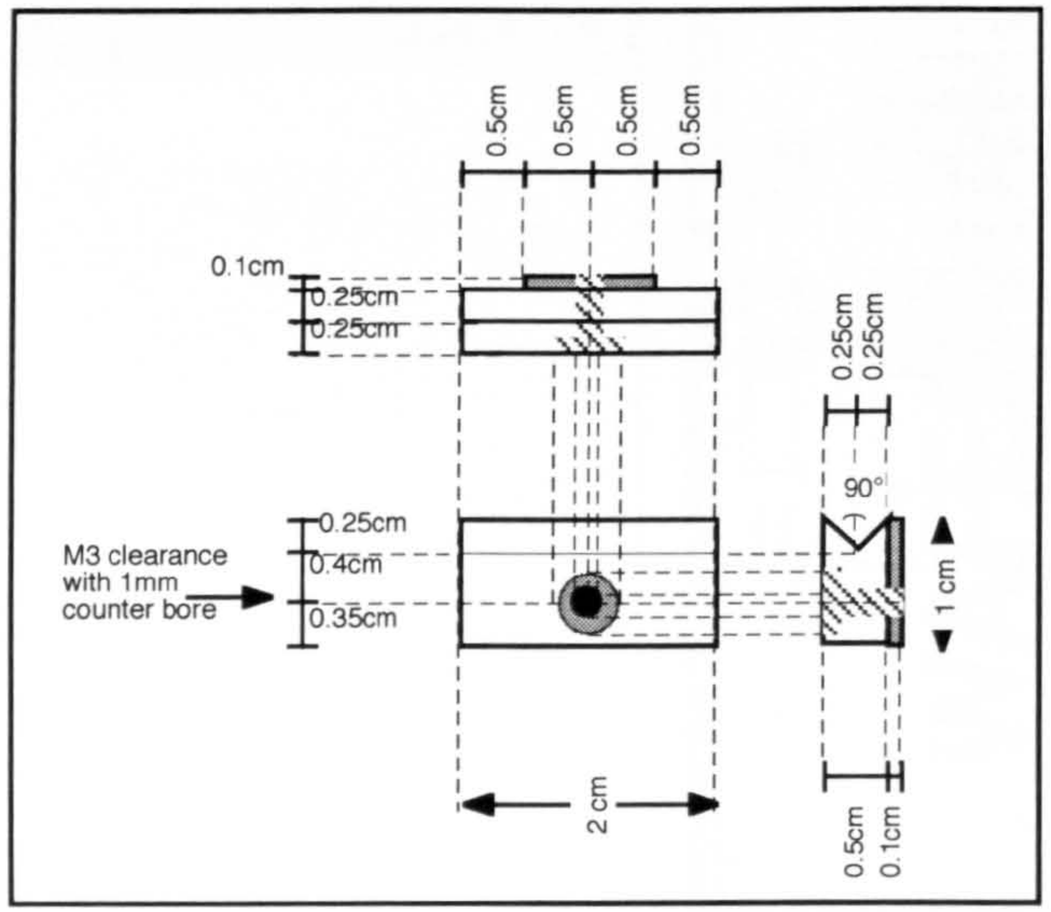
PART 8:-
O'ring x 1: Polypropylene ring
Holes to be made in sheet

APPENDIX C

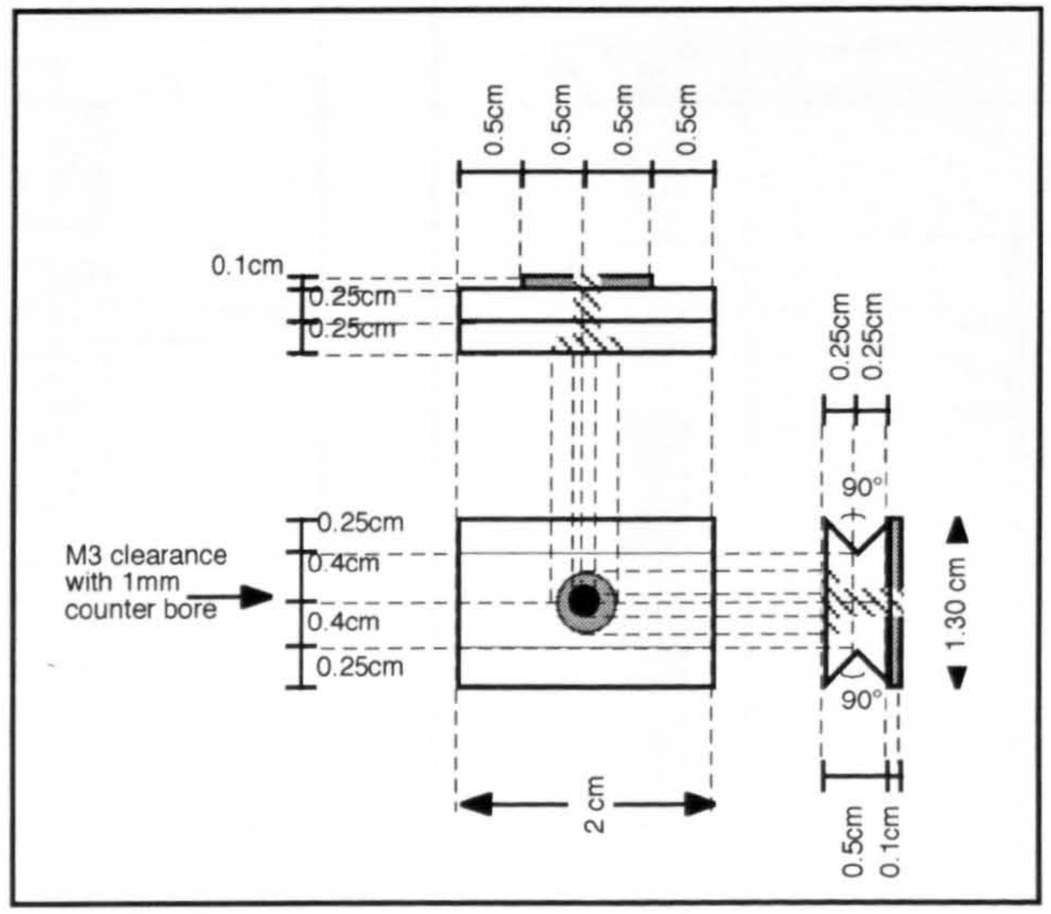
MECHANICAL DRAWINGS FOR WAFER HOLDERS


PTFE (Teflon) Wafer Holder
All material for parts 1, 2 & 3 must be PTFE, 4 & 5 must be Stainless Steel
Ashish Midha/John Weaver extn 6672
Part 1: Main Body
Quantity: 1 per
Tolerance 0.25mm
0.6 cm
1cm
0.7cm
0.3cm
0.5cm
1mm
0.56 cm
0.5cm
0.25cm
6.01 cm
0.76 cm
0.75 cm
0.7cm
5.63 cm
6.20 cm
5.82 cm
6 cm
7.51 cm

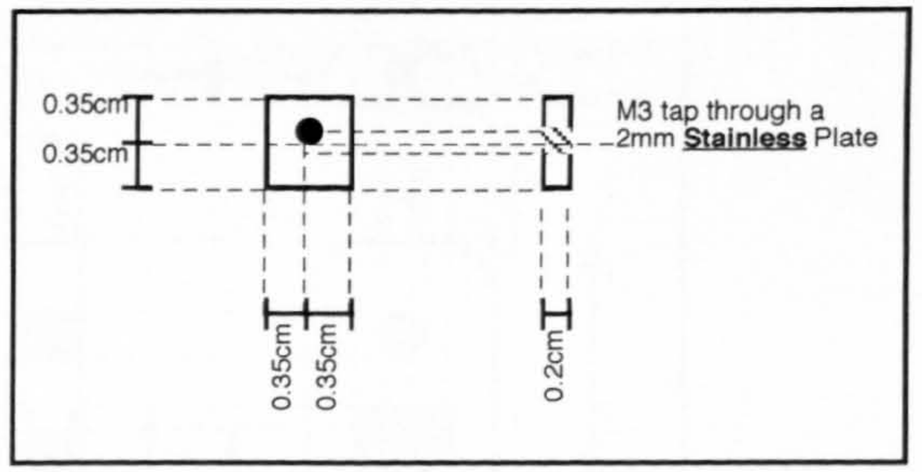
Part 2: End Clamps
Quantity: 8 per



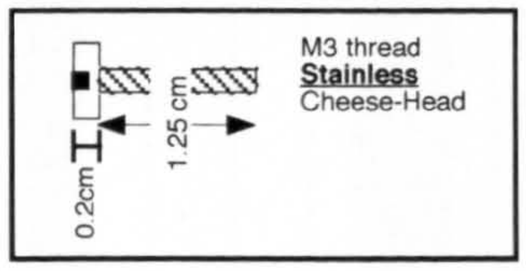
Part 3: Middle Clamps
Quantity: 4 per



Part 4: Backplate
Quantity: 12 per



Part 5: Screws
Quantity: 12 per



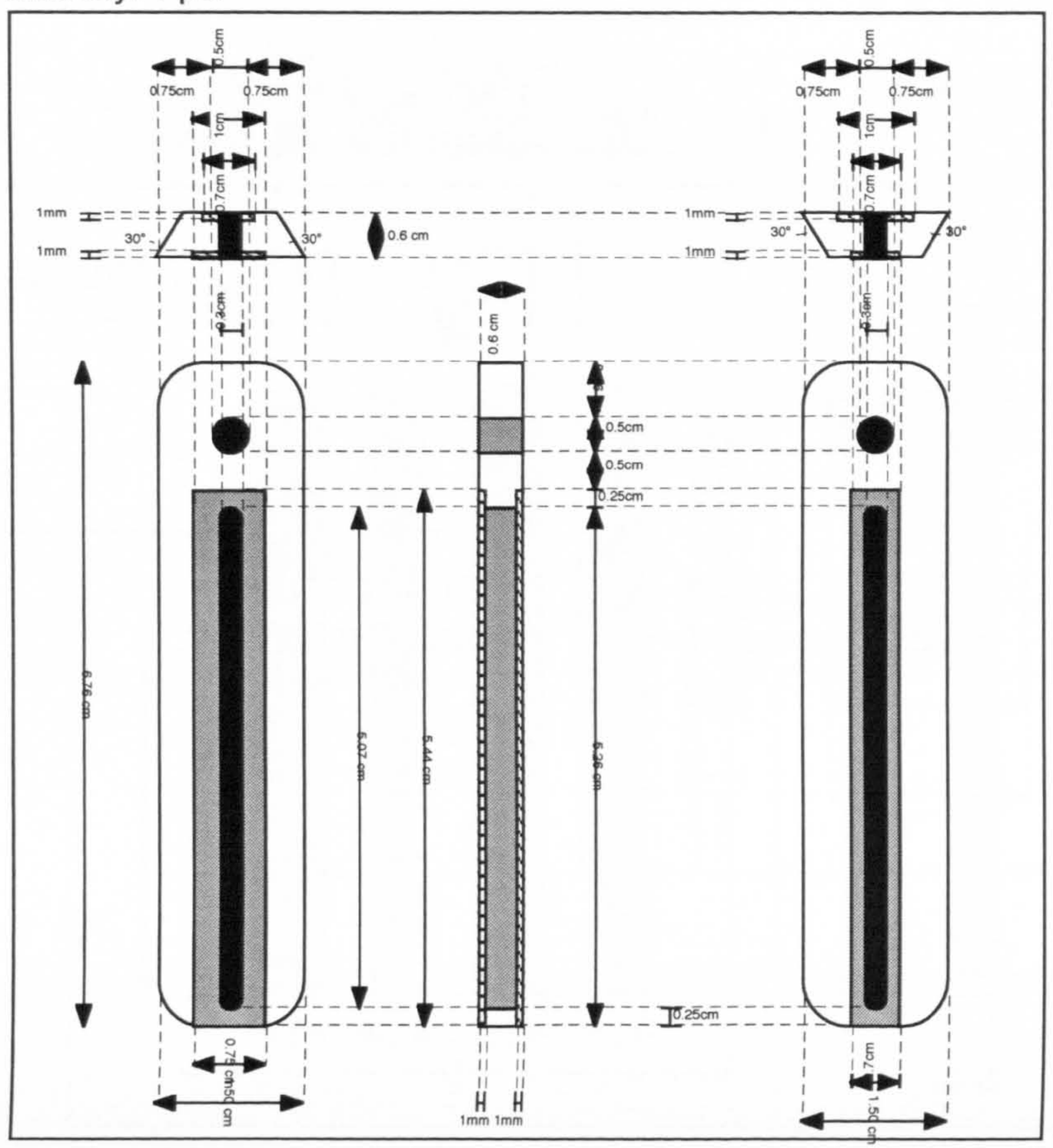
PTFE (Teflon) Wafer Holder

Tolerance 0.25mm

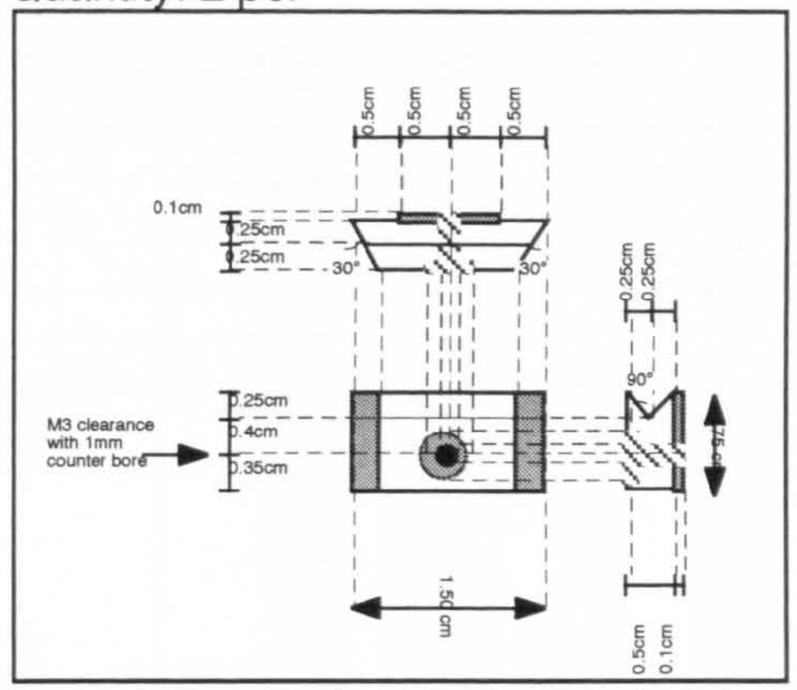
Ashish Midha/John Weaver extn 6672

All material for parts 1& 2 must be PTFE, 3 & 4 must be Stainless Steel

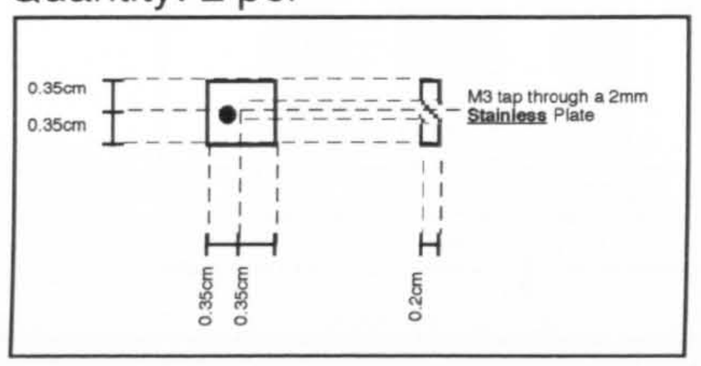
Part 1: Main Body
Quantity: 1 per



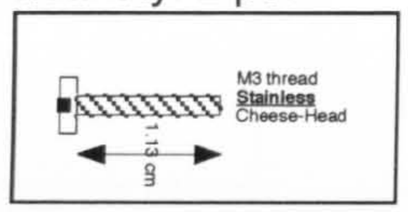
Part 2: Clamps
Quantity: 2 per



Part 3: Backplate
Quantity: 2 per



Part 4: Screws
Quantity: 2 per



APPENDIX D

WAFER HANDLING AND CLEANING PROCESSES

Two general-purpose cleaning processes were used for silicon, one organic, the other inorganic. The organic clean is as follows:

(a) Five minute dip in acetone with ultrasonic agitation, if possible (heating to 50°C may be used instead when ultrasonic agitation is not possible).
(b) Five minute dip in methanol as in (a). The wafers are transferred from acetone to methanol under a jet of acetone to keep them from drying.
(c) Five minute dip in IPA as in (a). The wafers are transferred from methanol to IPA under a jet of methanol to keep them from drying.

The inorganic clean is as follows:

(a) Twenty minute dip in nanostrip diluted with water in the ratio 1:1 (this solution heats up to 120°C) with ultrasonic agitation, if possible (when the wafers are fragile, the ultrasound treatment is omitted, and when the wafers have patterned metal features which are incompatible with this hot concentrated sulphuric acid and hydrogen peroxide mixture, this step is omitted completely).
(b) Five minute dip in ultra-pure water with ultrasonic agitation, if possible (when the wafers are fragile or have patterned metal features, heating to 50°C may be used instead)
(c) Step (b) is repeated with fresh ultra-pure water if step (a) has been carried out to remove all traces of acid.

It is normal to use both cleans sequentially, however for some processes it is better to end with one rather than the other. It has been found that photoresist adheres better to wafers which end with an inorganic clean. The procedure used then is as follows:

(a) Organic clean.
(b) Inorganic clean with wafers being transferred from the IPA in (a) under a jet of IPA to keep them from drying.
(c) Thorough blow dry directly from water followed by a five minute bake.

Prior to the silicon etch, it is also better to end with an inorganic clean. Then the procedure is modified by keeping the wafers in water until the deoxidation step.

Prior to evaporating metal onto a bare substrate or coating it with electron-beam resist, it is found that ending with an organic clean produces better results. Here the procedure is as follows:

(a) Inorganic clean.
(b) Five minute dip in IPA with ultrasonic agitation, if possible (when the wafers are fragile or have patterned metal features, heating to 50°C may be used in place of the ultrasound treatment).
(c) Organic clean with wafers being transferred from the IPA to methanol under a jet of IPA to keep them from drying.
(d) Blow dry directly from IPA, optionally followed by a 5 minute bake.

When metal is to be evaporated onto resist, no cleaning is carried out.

APPENDIX E

PATTERNING LPCVD SILICON NITRIDE LAYERS

E.1 WET ETCHING SILICON NITRIDE

The etch used is a mixture of orthophosphoric acid and water in a particular concentration that boils at 146°C. As water is added to the orthophosphoric acid it goes through different phases which boil at progressively lower temperatures, so that the solution is made up by slowly adding water to concentrated orthophosphoric acid until it boils at 146°C, as measured by a thermometer immersed in the solution.

For this etch, photoresist can be used as a mask. The following procedure is then followed [details regarding the pre-cleaning of substrates can be found in Appendix D]:

(a) S1818 photoresist is spin coated at 4000 rpm for 30 seconds on the polished side of the substrate (to protect this side of the wafer from the etch) [nominally 1.8μm thick]. Care is taken to remove any photoresist on the rear of the substrate using a cotton swab and acetone.

(b) The resist is baked at 120°C for 30 minutes.

(c) A second layer of S1818 is spun on the polished side of the substrate as in (a) [nominally 1.8μm thick]. This second layer dramatically reduces the chances of resist pinholes being replicated in the silicon nitride. As before any resist on the other side of the substrate is removed.

(d) A second bake at 120°C for 30 minutes is used to dry this second resist layer.

(e) S1818 is now spin coated onto the unpolished side of the substrate at 4000 rpm for 30 seconds [nominally 1.8μm thick].

(f) The substrate is then prebaked at 90°C for 30 minutes.

(g) The resist (unpolished side) is then exposed in the mask aligner for 13 seconds.

(h) The pattern is then developed for 75 seconds in AZ 1400 developer:water 1:1.

(i) The sample is then postbaked at the elevated temperature of 180°C for 60 minutes.

(j) The wafer is then etched in boiling orthophosphoric acid at 146°C.

The main feature of this process is the postbaking of the patterned resist at 180°C instead of 120°C. This was found to be necessary as the 120°C baked resist suffered from a significant amount of resist 're-flow' during etching at the higher temperature of 146°C and resulted in resist undercutting and feature rounding. It should be noted that once the photoresist has been baked at 180°C it can no longer be stripped off using acetone. Instead a proprietary organic cleaner, nanostrip, which consists of a mixture of concentrated sulphuric acid and hydrogen peroxide, heated to 120°C by diluting it with water in a 1:1 ratio, is used to strip off the photoresist. The samples are left to soak in

this solution for 20 minutes with ultrasonic agitation. This is then repeated once more to make sure all traces of photoresist have been removed.

The etching was carried out in a separate reaction flask and heating mantle, in a reflux system, with manual control of the heater power and a glass mercury thermometer, graduated in 1°C increments, for measuring the temperature. This thermometer was coated in PTFE as the etch will attack silicon oxide (the basic constituent in most clear glasses). However, as the temperature will not exceed 146°C (the boiling point) temperature control is not difficult. [Note that wafer holders utilising PTFE, stainless steel or tungsten are used as they are unaffected by this etch.]

The etch rate of LPCVD silicon nitride was found to be about 3nm/minute, so that the etch takes about 20 minutes. Overetching is not a problem as the silicon is not etched (noticeably) by this solution. However, adhesion of the resist to the substrate is critical to avoid the resist being undercut and features becoming larger and rounded.

E.2 DRY ETCHING SILICON NITRIDE I

Two dry etching schemes are popular within the department for the dry etching of PECVD silicon nitride. These schemes can also be used to pattern silicon oxide and are carried out on the same machine, the Plasma Technology BP80 reactive ion etcher [see §3.6.4]. One scheme is based on SF_6 while the other uses C_2F_6. However, it is generally found, as is discussed in §6.2, that resist is attacked more in the SF_6 system than the C_2F_6 system, probably due to the generation of more free fluorine in the former. Thus the C_2F_6 scheme was chosen, even though this was not a problem that would specifically affect this process as 1.8μm thick photoresist can easily survive the etching of 500nm or more of silicon nitride.

The etching process is carried out with a flow rate of C_2F_6 of 20sccm at a pressure of 15mTorr and with a radio frequency power of 100 Watts. This generates a self-bias of between -300V and -400V. The following procedure is then used to pattern the LPCVD silicon nitride using dry etching [details regarding substrate pre-cleaning can be found in Appendix D]:

(a) S1818 photoresist is spin coated onto the polished side of the substrate at 4000 rpm for 30 seconds as protection [nominally 1.8μm thick]. Resist on the other side of the wafer is removed using a cotton swab and acetone.
(b) The substrate is baked at 120°C for 30 minutes to dry this resist layer.
(c) S1818 is spin coated at 4000 rpm for 30 seconds on the unpolished side of the substrate [nominally 1.8μm thick].
(d) The substrate is then prebaked at 90°C for 30 minutes.
(e) The resist is then exposed in the mask aligner for 13 seconds.
(f) The resist is developed for 75 seconds in AZ 1400 developer:water 1:1.
(g) The resist is then postbaked at 120°C for 30 minutes.

(h) The substrate is then dry etched in C_2F_6 in the Plasma Technology BP80 using the conditions stated above [§3.6.4].

It has been found that no appreciable difference in etch rate exists between LPCVD silicon nitride and PECVD silicon nitride using this gas (unlike in the case of wet etching where the two behave completely differently). Silicon nitride (also silicon oxide) etches at between 40nm/minute and 50nm/minute. The photoresist etch rate is between 30nm/minute and 40nm/minute under the above etching conditions. The substrate is usually etched for 5 to 10 minutes (for 100nm of LPCVD silicon nitride) to make sure that the silicon nitride has been removed, since, as mentioned elsewhere, even a small trace of silicon nitride is an effective mask for the potassium hydroxide etching solution. This process also etches silicon at about 30nm/minute [see §6.2], however, this is not significant as the silicon is to be etched to a depth of several hundred microns. Overetching also means that if the resist has not cleared out fully, due to underexposure or underdevelopment, there is a high probability that the etch will remove both any thin traces of photoresist and the silicon nitride beneath. This is very important for preventing an excessive amount of surface roughness being generated during the silicon wet etch.

E.3 DRY ETCHING SILICON NITRIDE II

Early attempts to use high resolution negative resist (HRN) to pattern the LPCVD silicon nitride using the orthophosphoric acid wet etching technique of §E.1 were not very promising as the resist layer would be severely undercut, by as much as 5μm in some instances. Thus dimensional control of the pyramidal tips to a few microns would be difficult to achieve with this process, so a dry etching strategy, similar to that in §E.2, was pursued instead. The process used was as follows:

(a) The substrate was cleaned using an inorganic clean followed by an organic clean [App. D].
(b) 8% HRN was spin coated at 5000 rpm for 60 seconds on the polished side of the substrate [nominally 120nm thick].
(c) The substrate was then prebaked at 120°C for 30 minutes.
(d) The pattern was exposed with a dose of 300μC/cm^2.
(e) The pattern was then developed in MIBK for 10 seconds, followed by IPA for 10 seconds, followed once again by MIBK for 10 seconds followed by IPA for 10 seconds.
(f) The resist was then postbaked at 180°C for one hour.
(g) The substrate was then dry etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 3 to 4 minutes (50nm LPCVD silicon nitride) in the Plasma Technology BP80 [§3.6.4].
(h) The resist was stripped during the inorganic clean that followed [App. D].

This worked reasonably well for a while, although the amount of roughness generated during the tip definition wet etch (as much as 5μm in a 15μm deep etch in some cases) was partly attributed to residue left behind by the resist during its removal (even ashing the substrate for an hour did not alleviate this problem). However, a change in the resist character took place that resulted in pinholes in the HRN resist layer which were then replicated in the LPCVD silicon nitride layer during the dry etching stage. These would then be transferred to the apices of the pyramidal tips [see Figure E.1a], resulting in pyramidal tips populated by a multitude of small inverted pyramidal etch pits.

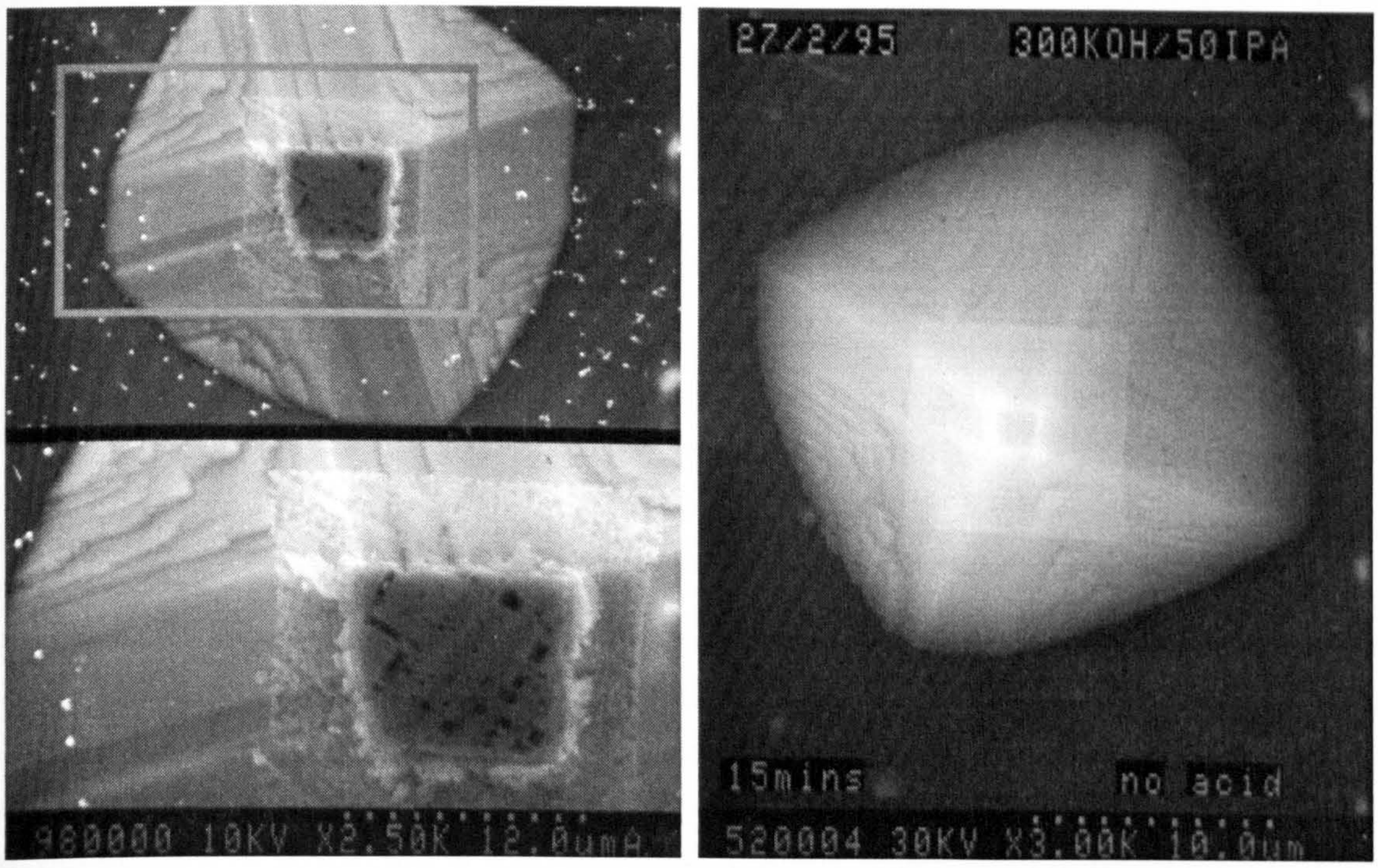


Figure E.1: Micrographs comparing a HRN mask with a nichrome mask for dry etching silicon nitride as a mask for wet etching silicon. (a) Left Image: problems with pinholes in the HRN masking layer being transferred to the silicon nitride and ultimately the silicon. The bottom image is a 2x magnification over the displayed scale; (b) Right Image: the improvement achieved by using a nichrome masking layer formed by a positive resist process for the silicon nitride dry etch.

Since the area to be patterned was very small, any alternative technique based on a positive resist would have had to write an unacceptably large area (reverse image). The solution to this was to write the same pattern in positive resist and use lift-off to negate the pattern.

The metal first used was titanium because of its adhesion properties. However, this metal layer was occasionally too thin for the etch, particularly when the switch to 100nm thick LPCVD silicon nitride coated wafers took place. The process was then altered to 100nm of nichrome, which turns out to be extremely resistant to most dry etches [Murad[1] (1993-1996)]. The process used for 100nm LPCVD silicon nitride coated wafers is then as follows:

(a) The substrate is cleaned using an inorganic clean followed by an organic clean [App. D].

(b) 8% BDH (or 8% equivalent thickness Aldrich) is then spin coated at 5000 rpm for 60 seconds onto the polished side of the substrate [Table 3.5.1].

(c) The resist is dried by baking at 180°C for one hour.

(d) 4% Elvacite is then spin coated at 5000 rpm for 60 seconds onto the previous layer of resist [Table 3.5.1].

(e) The substrate is then prebaked at 180°C for at least six hours (usually overnight).

(f) The pattern is then exposed using a dose of 300 to 350$\mu C/cm^2$.

(g) The pattern is developed for 35 seconds in MIBK:IPA 1:1.

(h) 50nm of nichrome is then evaporated onto the resist using the Plassys MEB 450 [§3.6.2].

(i) Lift-off is then performed using acetone, followed by a rinse in IPA and blow dry.

(j) The substrate is then dry etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 5 minutes or more in the Plasma Technology BP80 [§3.6.4].

(k) The nichrome layer is then stripped off using a ten minute chrome etch [§3.7.3] and followed by an inorganic clean [App. D].

This technique also has the advantage that there is no residue, due to the vigorous nature of the acid step in the inorganic clean [see Appendix D], although care must be taken during lift-off to make sure that no pieces of nichrome are allowed to resettle onto the substrate. The difference between this process and the last process (when it began failing) is illustrated in Figure E.1. The failure of the HRN is attributed to it being too far beyond its shelf life. [The resist was discontinued so that the department had the last remaining stocks of this resist.]

E.4 REFERENCES

1. S. K. Murad, *Private Communication*, (1993-1996).

APPENDIX F

EVOLUTION OF THE MASK SET

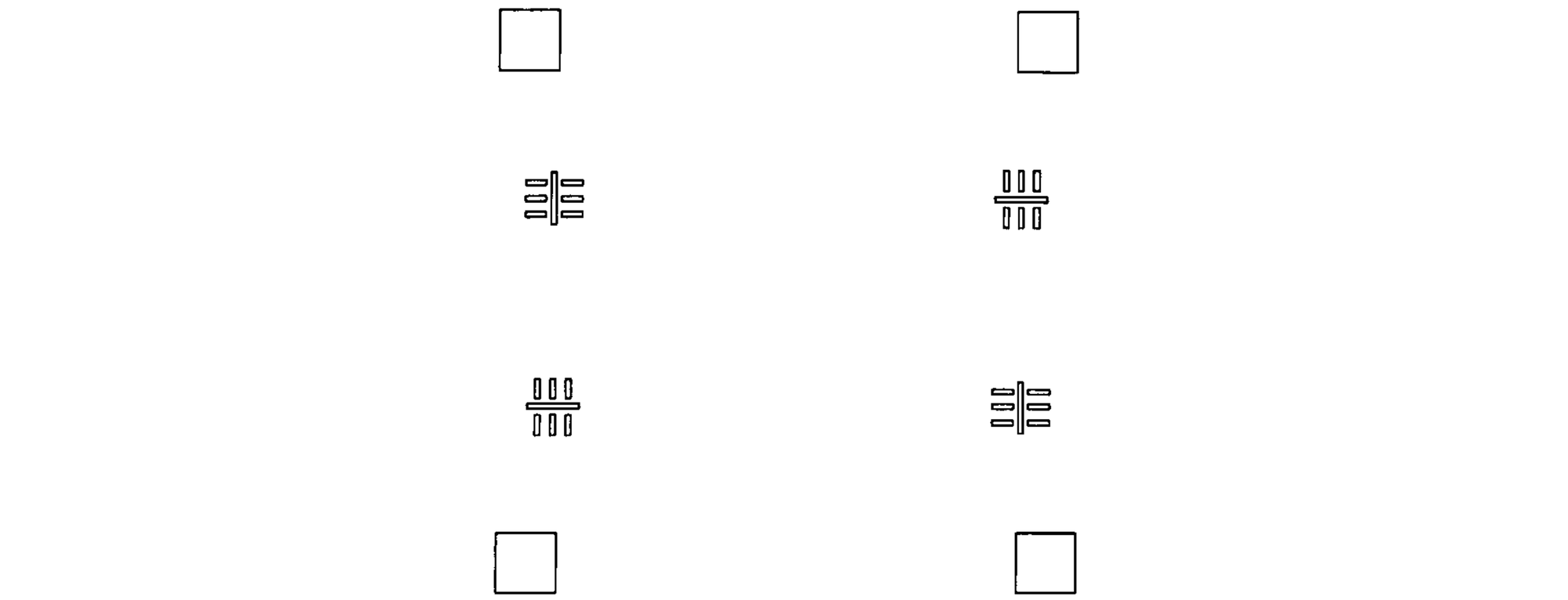


Figure F.1: An individual device pattern for the first level in one of the first mask sets devised.

Initially, while silicon etching techniques were being evaluated, the pattern design consisted of placing devices on a 5mm pitch, with the tip at the centre of a 5mm square region and through-wafer alignment markers at the corner of each individual device. However, having a tip around 20μm tall protruding from the centre of a 5mm device means that the tip would have to be held at an angle of less than 0.5° with respect to the sample in order that the tip, rather than the edge of the device, interacts with the sample. To rectify this, the next mask set placed the tips towards one end of the device and introduced the use of etched 'v-grooves' to make cleaving easier. The mask pattern for such a device is shown in Figure F.1. These cells were arranged in a 4 x 4 manner to give sixteen devices on a one inch square substrate, as shown in Figure F.2.

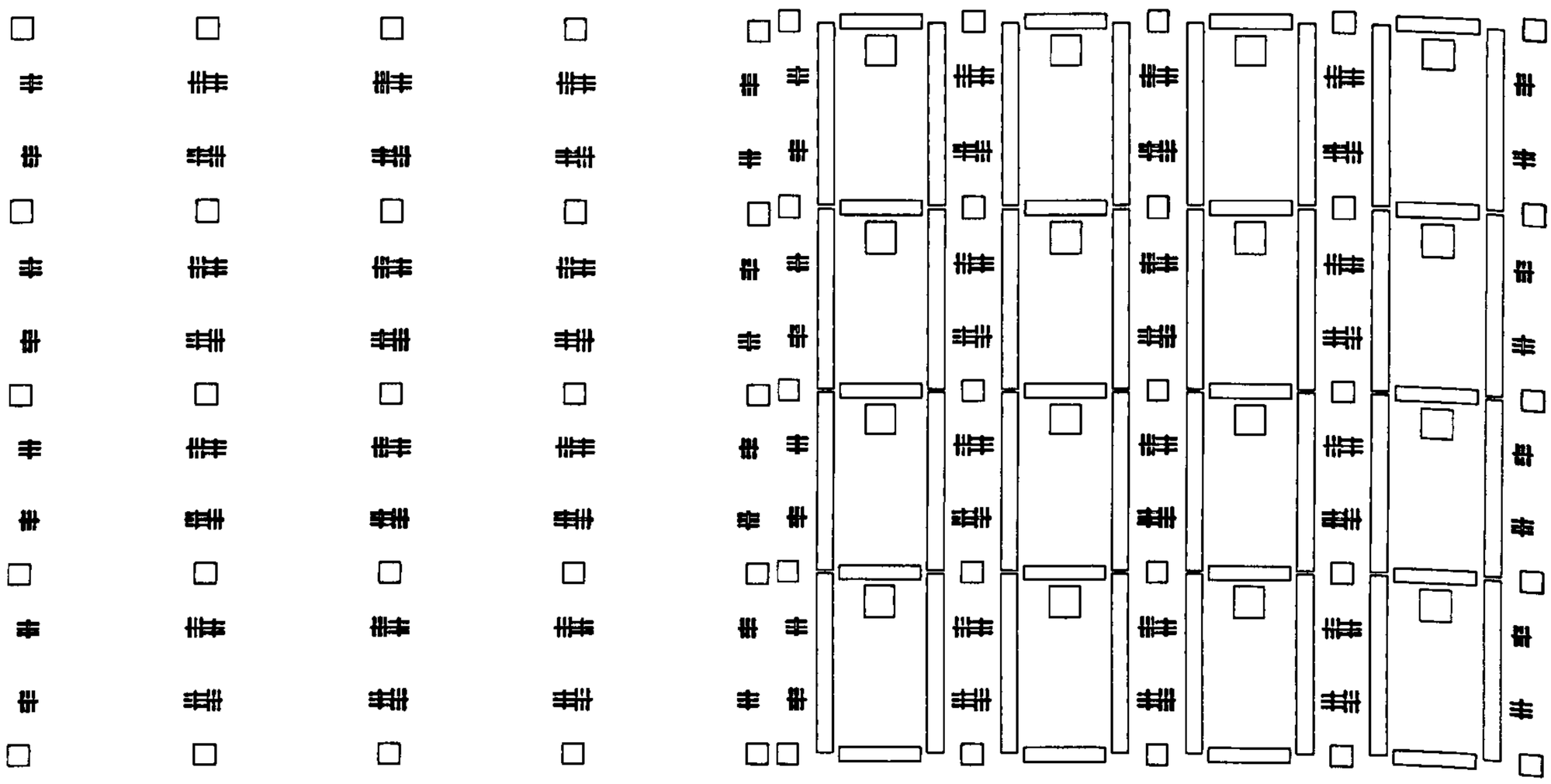


Figure F.2: A 4 x 4 array of devices for the first (left) and second (right) level in one of the first mask sets devised.

Since, at this point substrates were being obtained by cleaving a 3 inch wafer the 2.5 inch mask pattern consisted of two levels based on a wafer thickness of 375μm and 400μm as shown in Figure F.3. The set corresponding most closely to the actual wafer

thickness could then be used. A ferric oxide plate was used so that the mask would be transparent to permit alignment using the mask aligner's integrated optical microscope. This mask, amm0001, was produced in July of 1994. As can be seen from the figure, thin lines were placed at the edge of the first level pattern to allow alignment to the edges of the substrate, the [110] axes for a 25mm square substrate.

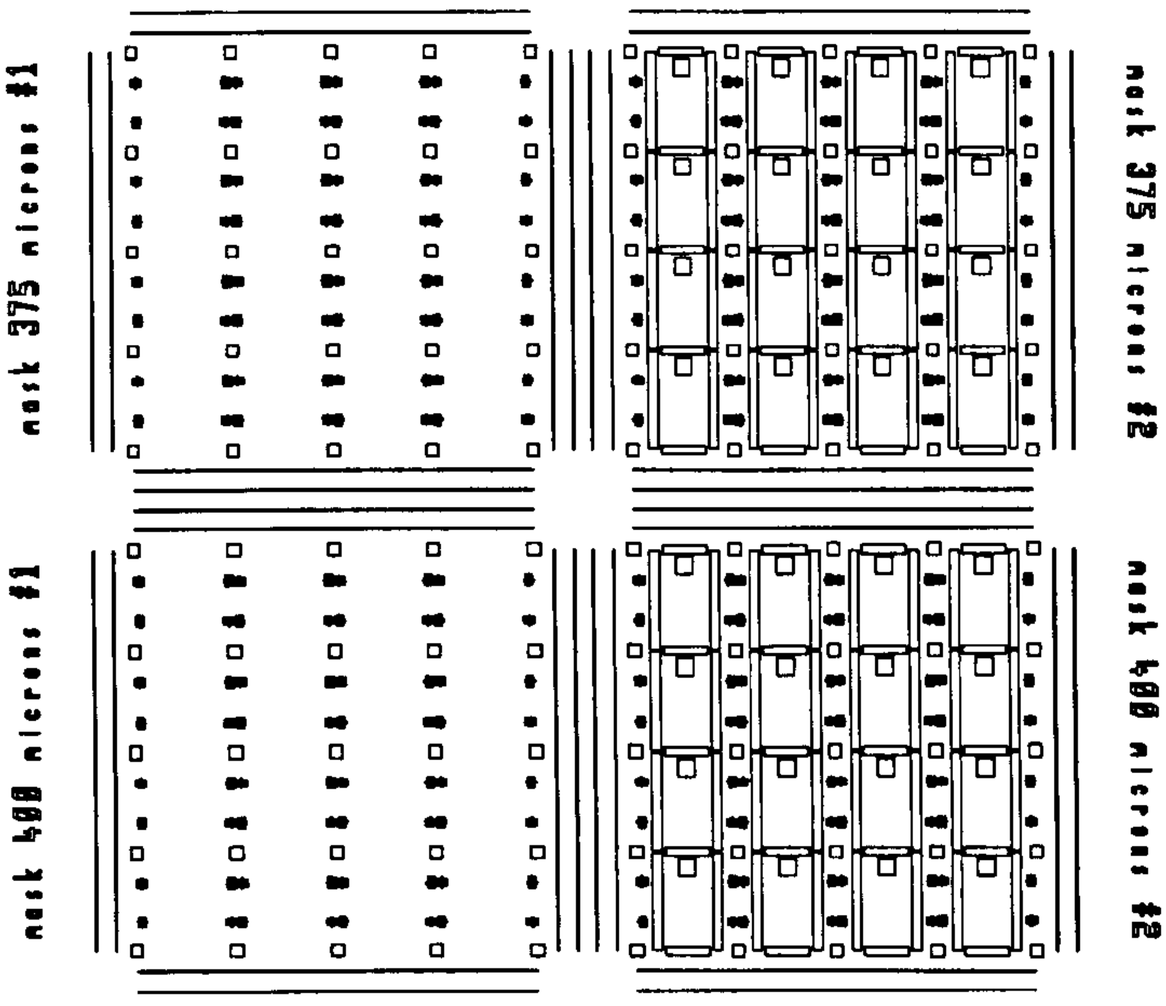


Figure F.3: The layout of one of the first mask plates devised with patterns for both levels on substrates of 375μm and 400μm thickness. Mask pattern name 'amm0001'.

Coarse alignment of the second level mask to the first level mask after the first level wet etch could be carried out by using the openings for the through-wafer alignment markers at the corner of each device, which were the same size in both levels. Fine alignment could be carried out using a 'tree' structure as shown in Figure F.4. This was devised to avoid exposing any convex corners which would be undercut during the first level wet etch [see §3.4.3], such as the more commonly used cross structures. Here all slots are 50μm wide which should terminate at a depth of about 35μm [see Figure 3.7.3]. This is much less than the depth of the first level wet etch [see §G.1] so that this structure would have been etched to completion by the end of the etch. The horizontal slots, which are 200μm long, are spaced out vertically at 150μm intervals on each side with a 50μm gap separating them from the vertical slot, which is 500μm long. The 50μm gap was estimated, and subsequently demonstrated, as being sufficient to accommodate any feature growth that might occur during the photolithography and etching of the {111} etch-stop planes. This was necessary to prevent the slots from merging which would have resulted in the formation of a convex corner and subsequent undercutting of the pattern.

It was also found that the sharp vertex at the bottom of the grooves, formed by the intersection of opposing {111} etch-stop planes, contained higher spatial frequencies which could be seen most clearly by viewing the whole structure in the dark field, *i.e.* the vertices all lit up. This proved to be extremely useful in fine tuning both the lateral

position and rotation of the mask plate with respect to the substrate when used at opposing corners of the wafer.

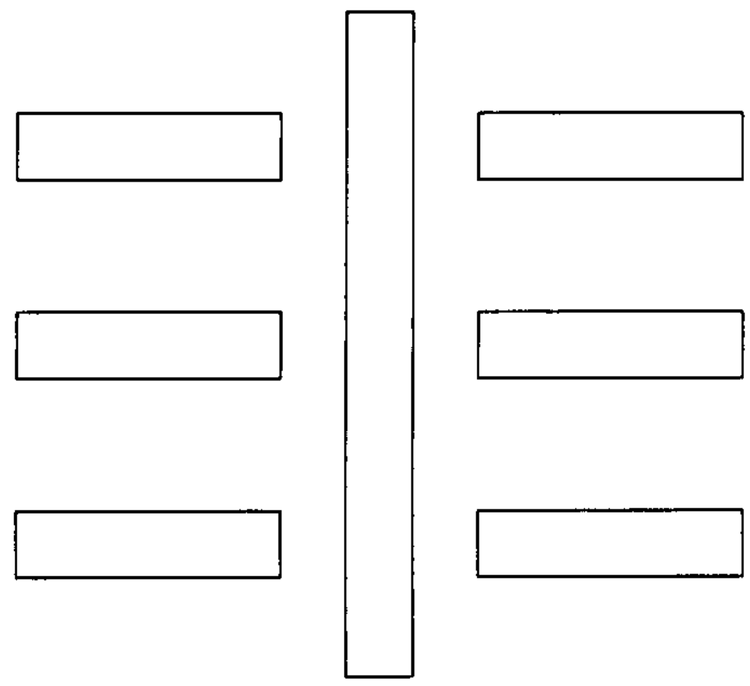


Figure F.4: Fine alignment structure for photolithography.

The details for the 375μm thick wafer pattern were as follows:

(a) The openings used for the through-wafer alignment markers were 580μm squares which were expected to result in 50μm square membrane markers at the other end of the wafer [from Figure 3.7.3].

(b) The device window itself was an 830μm square which was expected to result in a 300μm square membrane, sufficient for cantilevers of a few hundred microns in length.

(c) The vertical cleave lines were 455μm wide which were expected to reach a depth of 315μm before etching terminated.

(d) The horizontal cleave lines were 389μm wide and expected to reach a depth of 275μm upon termination.

The termination depths of the cleave lines were estimates based on the premise that the substrate be cleaved first vertically and then split horizontally into individual devices. The gaps between the vertical cleave lines were 50μm. The pattern for 400μm thick wafers were scaled up in the expected manner.

In practice there were several problems associated with this mask pattern:

(a) The lines at the substrate edges used to align the angle of the first level pattern to the edges of the substrate did not always reach the edge of the substrate (the substrates were sometimes smaller than this pattern required), and the substrate edges were not usually straight all the way along the wafer.

(b) The shapes and sizes of the through-wafer alignment markers upon the completion of the etching turned out to be extremely variable, with some being longer horizontally and some longer vertically, but all markers were 30μm to 50μm larger than the expected 50μm.

(c) Cleaving the substrate into individual devices often resulted in the cleave going through the cantilever membrane, and cleaving was in general very difficult.

(d) The position of the centre of the cantilever membrane from the centre of the horizontal cleave line next to it is 800μm. Upon reflection, this would mean that the cantilevers would be too far recessed into the substrate, as the outer wings of the device would protrude such that a horizontal tilt angle of only 1.5° would result in these wings interacting with the sample, and access of the tip to the specimen would generally be poor.

(e) Upon reflection, sixteen devices from a single one inch square did not appear to make the most efficient use of substrate area.

To tackle these problems another mask plate was designed, this time based solely on a 375μm thick substrate as the wafers being used were consistently around 380μm thick. The solutions attempted for each of the above problems in the order listed were as follows:

(a) A grating was used at the edge of the first level mask pattern, rather than the handful of lines used previously, so that substrates of varying widths could be used.

(b) Two different strategies were employed to tackle the through-wafer alignment marker size problem, but common to both was a reduction in the size of the opening used to 530μm (which should theoretically produce 0μm markers) in case the growth was due to slow etching of the {1 1 1} etch-stop planes. The first strategy was based on the assumption that the markers had grown to be rectangular due to a slight misalignment of the second level mask with the first level mask. If this misalignment was also rotational, then the direction of marker growth would be different at each corner of the substrate. Convex corner undercutting would make sure that by the end of the second level wet etch, the newly exposed area of silicon would have caught up with that exposed in the first level. To prevent this the openings in the *second* level pattern were reduced to 400μm so that they could still be used for coarse alignment without resulting in any marker growth. The second strategy was based on the assumption that some marker growth would occur if the first level pattern was not aligned perfectly with the [110] axes of the silicon substrate. To deal with this a speculative first level pattern based on opening circular holes instead of square holes was created. These circles would be converted (by the wet etch) to squares aligned to the [110] axes and not exceeding the diameter of the hole.

(c) The degree of fragility required to make the substrate easy to cleave without destroying the devices was also not known, so it was decided that

an additional speculative second level mask pattern be used to investigate this by having much larger openings for the cleave lines, to the point where the wafer could fall apart very easily. To offset the risk being taken here, the other second level mask pattern was made less fragile so that both extremes could be explored.

(d,e) It was realised that individual sets of through-wafer alignment markers for each device was an unnecessary waste of substrate area as the first electron-beam lithography level would place accurate markers for individual devices, along with those structures requiring high precision alignment (the tips) [see §4.3]. Thus only a coarse alignment was required to the through-wafer alignment markers to make sure that the tip and cantilever ended up on the membrane area. By placing markers only at the corners of the whole substrate, this could be accomplished whilst making more substrate area available for devices. To deal with the tip access problem, it was decided that the membranes would somehow need to be placed at the edge of the device. It was also realised that for each side of a membrane, 265μm of substrate area was taken up by the slope of a {111} etch-stop plane. It was therefore decided that each membrane have two cantilevers facing each other, sharing a common membrane between them, thus saving on substrate area by removing one {111} etch-stop plane per cantilever. This resulted in a unit cell which consisted of a pair of devices facing each other with the membrane area at the extreme edge of the device, which also improved the access of the tip to the sample. Further improvement to tip access required that the corners of the device at the cantilever end be tapered away from the cantilever. Such a taper could be provided by making *use* of convex corner undercutting. This could be achieved by exposing convex corners at the cantilever edge of the device by merging the cantilever membrane areas across the width of the substrate. However, since it was not known whether the rate of undercutting would yield a sufficient taper for a rectangular edged device, the additional speculative second level mask pattern was also designed so that undercutting would be assisted by tapering the openings at the cantilever edge of the device.

The design for the new mask plate is shown in Figure F.5. This mask, amm0003, was produced in December of 1994. It can be seen that the markers are placed outside the device area because the beamwriter does not reliably expose regions outside the marker boundary. Three markers are placed at each corner at a centre-to-centre spacing of 650μm. This arrangement allows several different combinations of four markers (which must be organised in a strictly rectangular pattern) to be used by the beamwriter, chosen according to which markers have turned out best. Each pair of devices occupies a region

of about 3 x 7mm and means that a 24mm square substrate yields 48 devices in the form of 6 rows of 8 columns, a factor of three increase over the previous design.

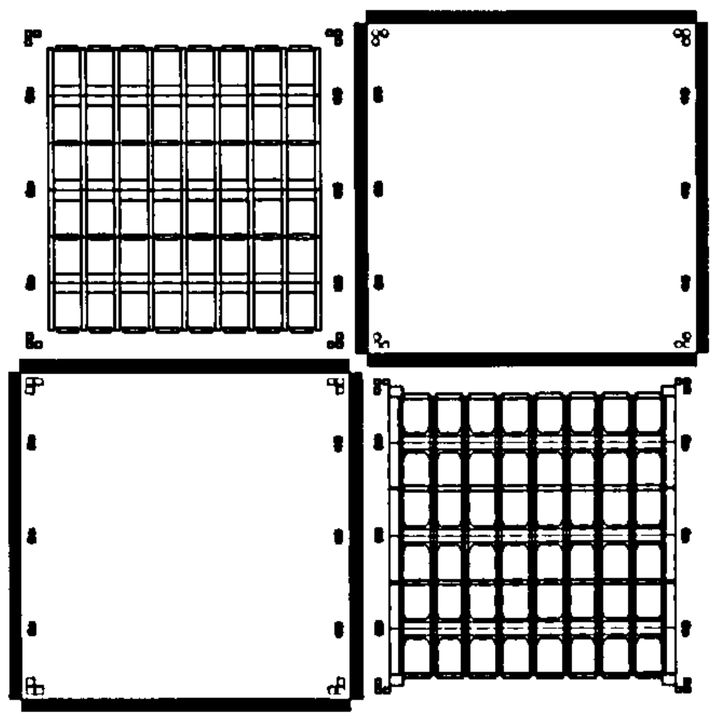


Figure F.5: The layout of the second mask plate devised with a conservative pattern and an ambitious pattern for both levels. Mask pattern name 'amm0003'.

The two first level patterns differ only in that the openings for the through-wafer alignment markers are different shapes, rectangular versus circular. As can be seen from Figure F.6 the second level patterns both have square through-wafer alignment marker openings of reduced dimensions. The more conservative mask pattern has 400μm wide (283μm termination depth) vertical cleave lines which are continuous from one end of the substrate to the other, while the more radical pattern attempts to form membranes running vertically at the edge of the substrate, with rows of devices being supported solely by the gaps between vertical cleave lines. Due to the slope of the {111} planes, these gaps would occupy a substrate area 530μm wider than that defined by the pattern upon the completion of the etch. The width of the central cantilever area is 1500μm, corresponding to about 970μm of membrane (expected to grow to about 1000μm during processing) being shared between two devices. This membrane is continuous across the width of the substrate, and though a membrane this large would normally fracture, this membrane would already have been cut up to form cantilevers prior to being etched fully through the wafer [see §4.4].

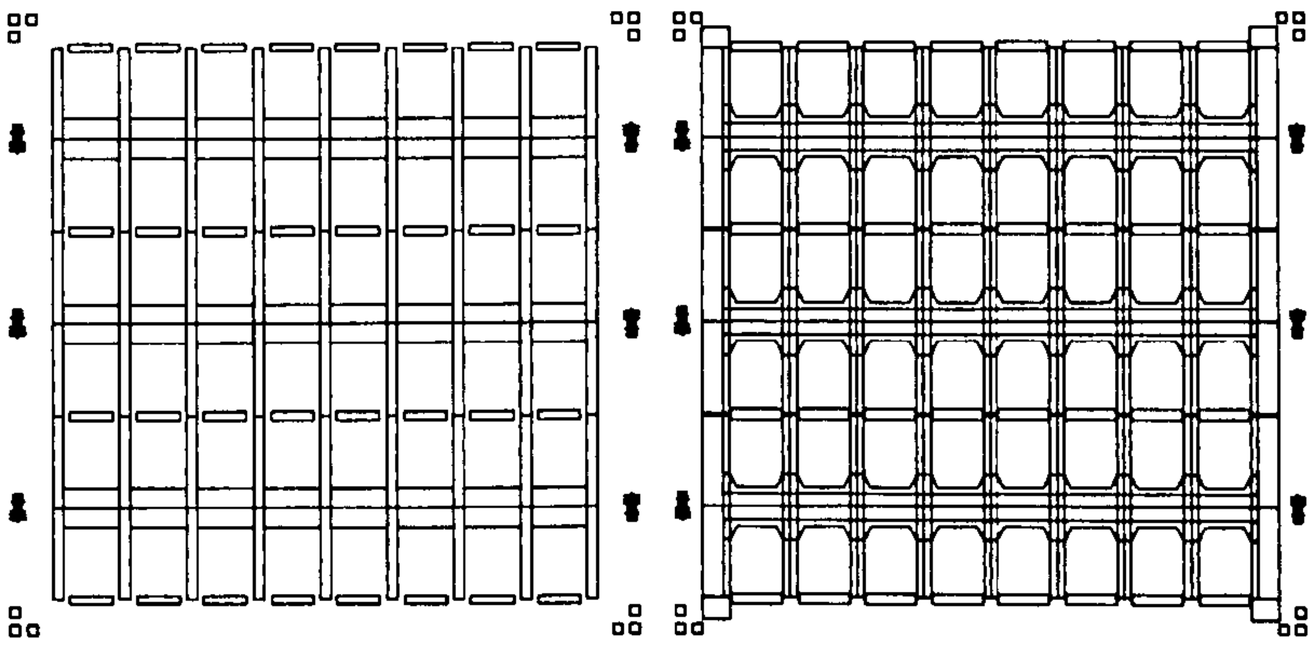


Figure F.6: A close-up view of the second level photolithography patterns from the mask plate ('amm0003') illustrated in Figure F.5.

However, some problems with this design emerged during use:

(a) The area of the gratings used to align the first level patterns to the substrate was large enough to push the pattern extent further out from the centre of the mask plate. This meant that, due to its construction, the optical microscope integrated with the mask aligner could not quite reach the edges of the mask pattern for accurate alignment.

(b) It was found that obtaining five 24mm square substrates from a single 3 inch wafer was unreliable, due to the high level of cleaving control required, so that, on average, only four substrates could be relied upon.

To address these issues, very shortly afterwards an additional mask plate was designed with the following modifications:

(a) The area of the gratings was reduced to bring the patterns closer together and therefore closer to the centre of the mask plate.

(b) The substrate size was reduced to 21.5mm by 24mm forcing the number of devices to be reduced by one column from 48 to 42 per substrate arranged as 7 columns of 6 rows (this reduced the number of devices per 3 inch wafer to 210 from 240 previously).

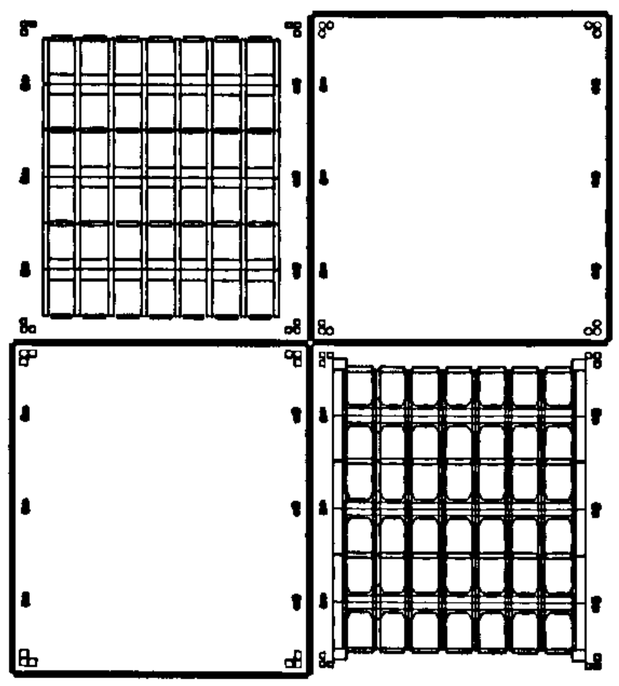


Figure F.7: The layout of the third mask plate devised with a conservative pattern and an ambitious pattern for both levels. Mask pattern name 'amm0004'.

All other aspects of the mask pattern remained unchanged. The result of this change is illustrated in Figures F.7 and F.8. This mask, amm0004, was written in January 1995. However, the problem of bad through-wafer alignment markers persisted despite the changes, although all the markers were smaller, and some did not even make it through the substrate. This problem was so bad that very few, approximately 25% of substrates, had markers of a quality sufficient for use in the first electron-beam lithography level. It was hypothesised that because the resist tends to be thicker very close to the edge of a rectangular substrate than anywhere else, probably due to centripetal forces during spin coating and surface tension effects at the edge of the substrate, placement of the markers so close to the edges of the substrate may have led to their elongation and enlargement.

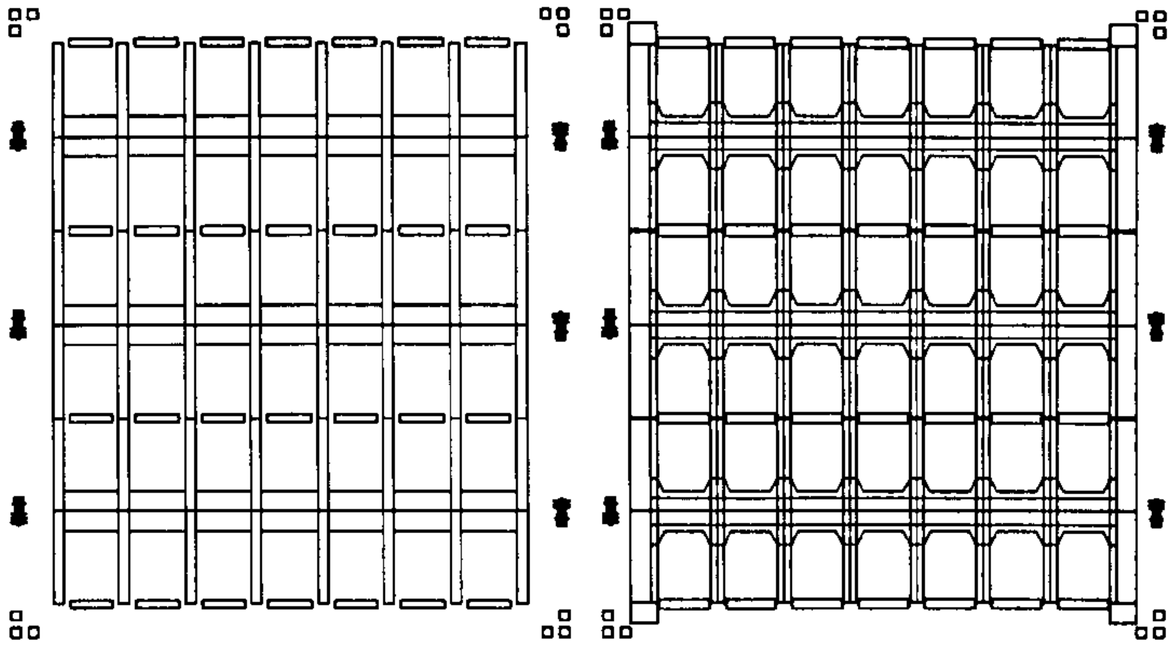


Figure F.8: A close-up view of the second level photolithography patterns from the mask plate ('amm0004') illustrated in Figure F.7.

To resolve this problem and since the conservative second level pattern did not seem to require further modifications at this time (convex corner undercutting of the tapered device edges of the speculative pattern proceeded too quickly so that the conservative pattern was being used exclusively), it was decided that the number of substrate edges be reduced whilst increasing throughput from the first level by carrying out the first level on an entire 3 inch wafer. This could then be cleaved up after the first level wet etch for patterning with the current conservative second level pattern.

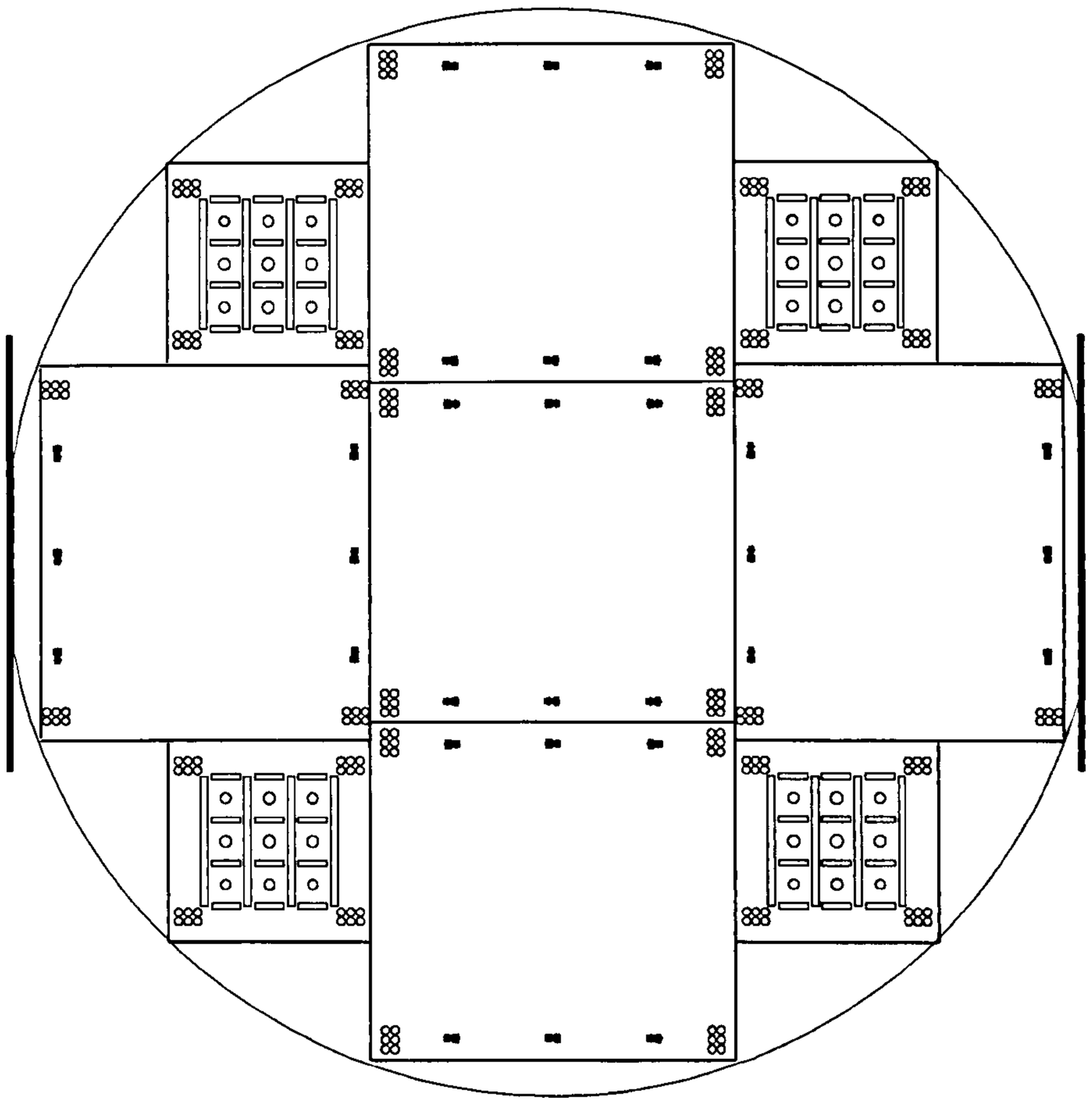


Figure F.9: The layout of the first whole 3 inch wafer mask plate devised for first level photolithography only. Mask pattern name 'amm0005'.

The resulting 4 inch mask plate, amm0005, is shown in Figure F.9 and was written in February 1995. The main differences were: the pattern was designed explicitly for a 380μm thick wafer, since all the 3 inch wafers seemed to have thicknesses very close to this value; the openings for the through-wafer alignment markers were increased to 560μm circles which should yield a marker size of 22.6μm for a 380μm thick wafer, and; the number of markers at the corner of each substrate was increased to six to increase the probability of a good set of markers being found. [The previous mask-plate had also shown that circular openings tended to produce smaller markers than square openings, indicating that the edge of substrate alignment was not sufficiently good, probably due to the poor form of the substrate edges after the scribing and cleaving process.]

To design this pattern, a 3 inch wafer was traced onto a piece of paper and measured as accurately as possible. Account was taken of the large and small flats of the wafer inasmuch as a substrate could be placed much closer to the edge of the wafer on the large flat side as opposed to the small flat side. The six through-wafer alignment markers could not be placed arbitrarily since the pattern had to be compatible with the existing second level pattern of the previous mask plate, *i.e.* the three marker openings at each substrate corner in the previous plate's second level pattern had to overlap with three of the six openings at each substrate corner in the new first level pattern.

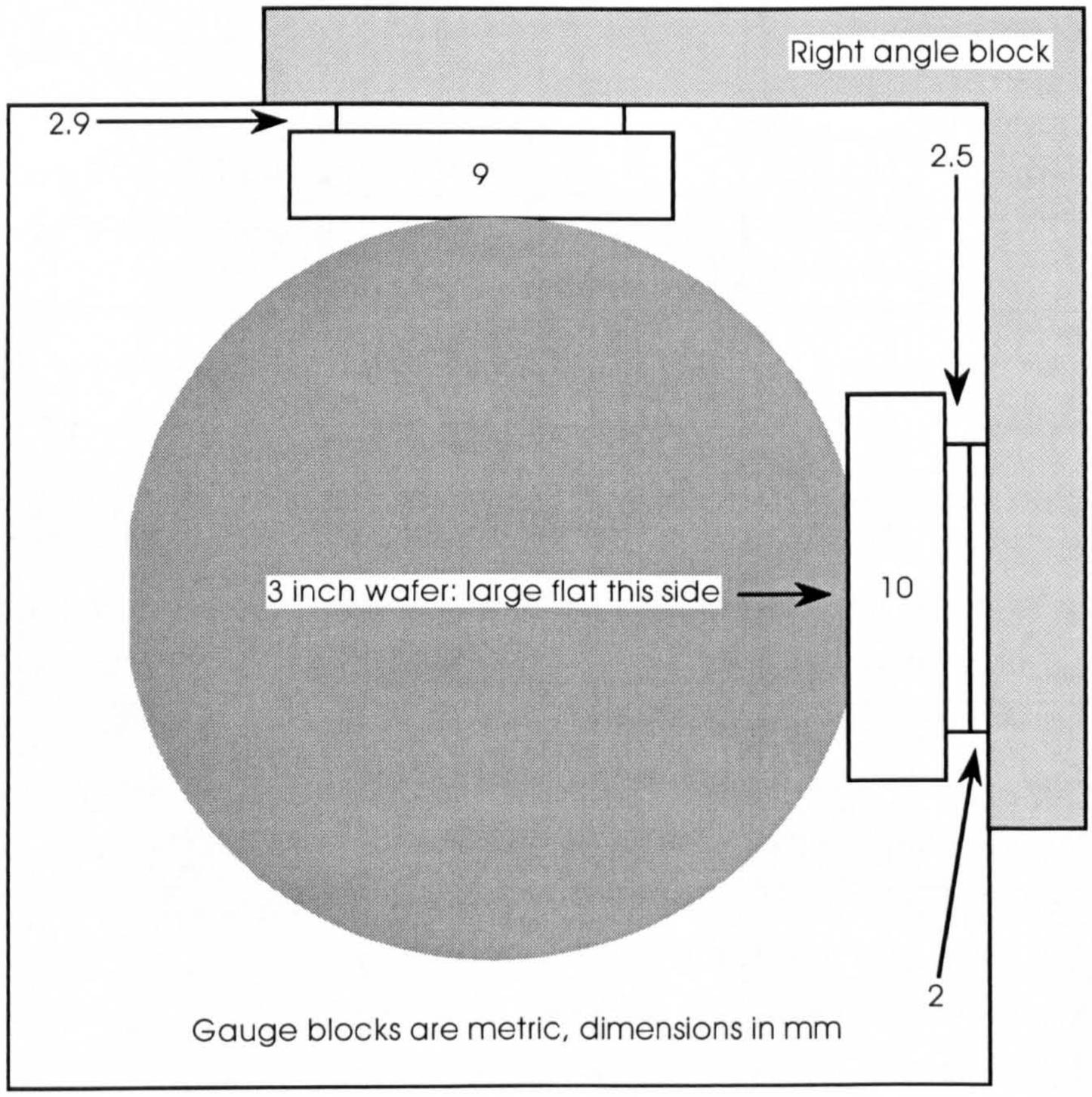


Figure F.10: The gauge block technique devised to align the mask plate ('amm0005') of Figure F.9 to a 3 inch wafer with rotational accuracy. Drawn to scale.

An additional feature of this mask is the patterning of the corner pieces to make samples with alignment markers and membranes for use as transmission electron microscope (TEM) substrates. These substrates were requested by other users for their projects and as the cost of patterning these was minimal, the request had been accommodated within the pattern.

50μm wide substrate cleave 'guide' lines were exposed by this pattern so that the substrates would be easier to cleave accurately prior to the second level. Since two of these guide lines were continuous from one side of the wafer to the other, a distance of around 70mm, a misalignment of only 0.1° with respect to the crystallographic planes would result in these lines growing by around 120μm in width. To prevent this a 40μm 1:1 mark:space ratio grating was placed at each of the wafer-flat edges of the pattern. As these flats are very cleanly defined by the manufacturer, reproducible alignment of 10μm along the large flat, which is around 15mm long, could be achieved, corresponding to a rotational reproducibility of 0.04°. However, at the time this plate was written, it had been found that good ferric oxide copies of 4 inch mask plates (from the chromium originals produced directly by electron-beam lithography) could not be made due to the use of light sources with an inadequate intensity uniformity [Harkins[1] (1993-1996)]. This meant that the first level pattern had to be exposed using the chromium plate directly which meant that the grating-based rotational alignment scheme could not be used. To circumvent this problem, a scheme making use of a contact printer and wafer positioning using right angle and gauge blocks as shown in Figure F.10 was devised by trial and error. The wafer would be placed on the surface of the mask plate and positioned with respect to its edges using these blocks, before the mask plate was illuminated from beneath, exposing the pattern on the wafer.

The direct use of a chromium plate meant that the pattern was a mirror image of that which would be produced using a ferric plate. This presented a problem for the second level ferric oxide plate as the fine alignment 'tree' markers were arranged with rotational symmetry around each substrate and the mirroring resulted in the handedness of this symmetry being reversed. This meant that the second level ferric oxide plate had to be mirrored by copying the pattern onto another 2.5 inch ferric oxide plate.

The number of substrates usable in the development of the electron-beam lithography levels increased dramatically as a result of this mask plate, even though the size and shapes of the markers seemed not to have improved noticeably. The improvement in substrate yield was solely because the additional number of markers at the corner of each substrate had dramatically increased the number of marker combinations possible. A partial explanation for the variability of marker size was the use of a non uniform illumination source for the first level exposure (the mask aligner could not be used with the gauge block scheme) so that the centre had to be overexposed in order that the edges of the wafer received a sufficient dose. However, this did not explain why the markers were not square.

An additional problem with this mask plate was that upon cleaving the wafer for the second level, the substrates at the wafer-flat edges of the wafer, which had markers very close to the cleave guide lines, would sometimes be destroyed during the cleaving process. However, the four remaining markers on these substrates were usually sufficient.

This mask plate allowed more rapid development of the first two electron-beam lithography levels [see §4.3 and §4.4]. However, as the development of the electron-beam lithography levels proceeded apace, the demand for greater throughput from the photolithography levels eventually meant that the second photolithography level became the bottleneck in the development process. At this time, a method for producing 4 inch ferric oxide copies of chromium mask plates became available due to some in-house development work [Harkins, MacIntyre and Thoms[2] (1993-1996)].

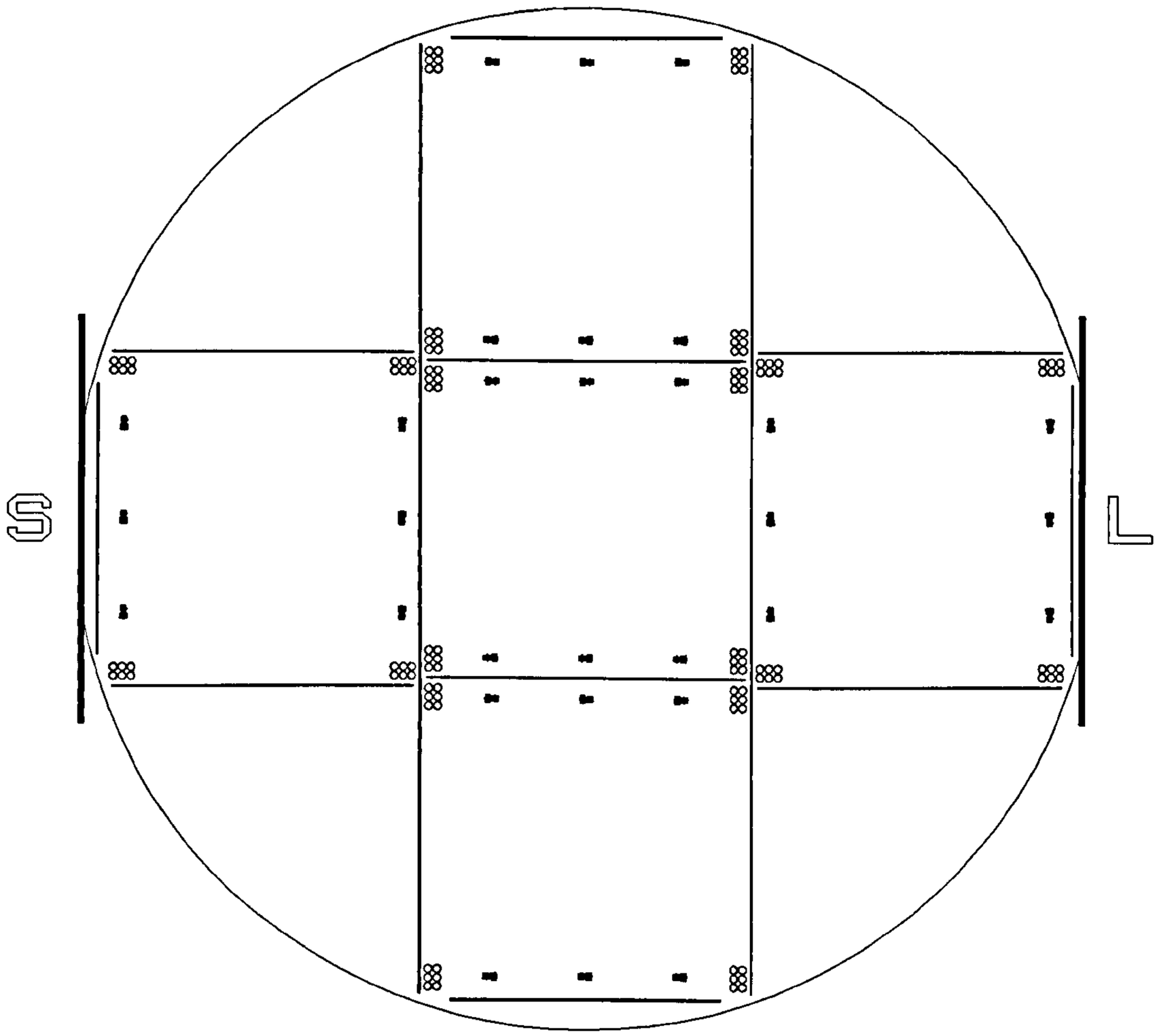


Figure F.11: The second whole 3 inch wafer mask plate devised for the first photolithography level. Mask pattern name 'amm0006'.

Thus it was decided that two 4 inch mask plates be produced so that both levels could be carried out on whole 3 inch wafers. These wafers would then be cleaved after the second level wet etch, prior to the electron-beam lithography levels. To accomplish this the cleave guide lines had to be replaced with actual cleave lines (deeper 'v'-grooves) in the second level. From the position of the grating pattern exposed in the first level, a more accurate rendition of the outline of a 3 inch wafer was produced. The resulting mask plates for

both levels are shown in Figures F.11 and F.12. The corner pieces were left unpatterned here, as their effect on the cleaving of the wafer after the second level was not known and could endanger device substrates. These plates, amm0006 and amm0007, were written in June 1995.

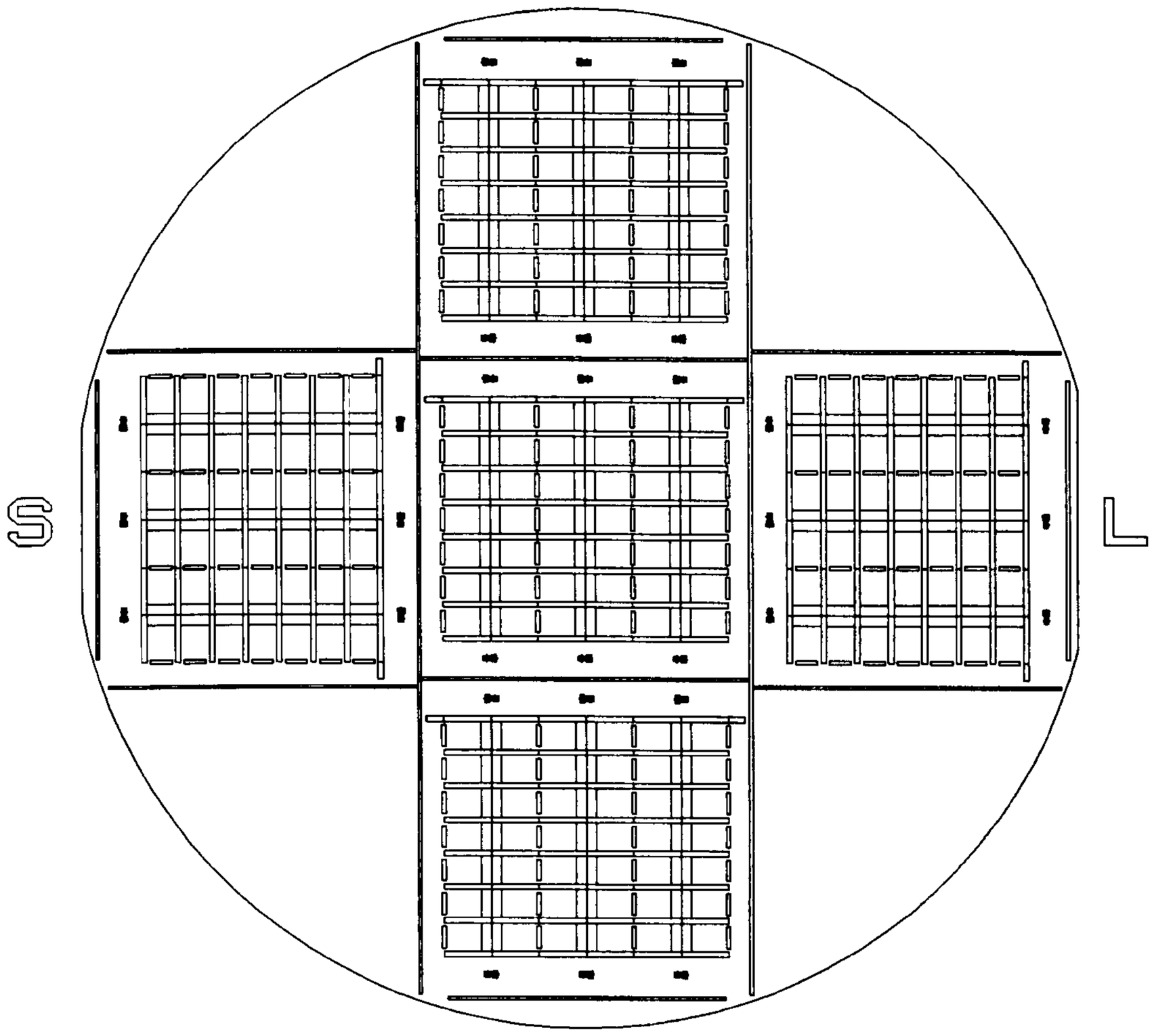


Figure F.12: The first whole 3 inch wafer mask plate devised for the second photolithography level. Mask pattern name 'amm0007'.

For the first level, the wafer alignment grating was retained (but positioned better). The full length cleave guide lines were widened to 75μm with all other features remaining largely unchanged from the last mask plate. An 'S' and 'L' was also written onto the plate (off the wafer area) to indicate the large and small flat sides of the pattern. The six markers at each corner were also better placed with respect to the cleave lines. For the second level, the full length cleave lines were widened to 250μm (termination depth 177μm), with the shorter ones being widened to 200μm (termination depth 141μm). To make the substrates easier to cleave once the devices had been made, the vertical cleave line at one side of each substrate was made weaker by extending its length by 1250μm at either end. The grating was not written in the second level pattern to protect the wafer edges from further damage and the through-wafer marker openings were also omitted from the second level to eliminate the possibility that variability in marker size was due to the resist around these markers being exposed through the smaller second level openings. The remaining features were broadly unchanged from the previous pattern.

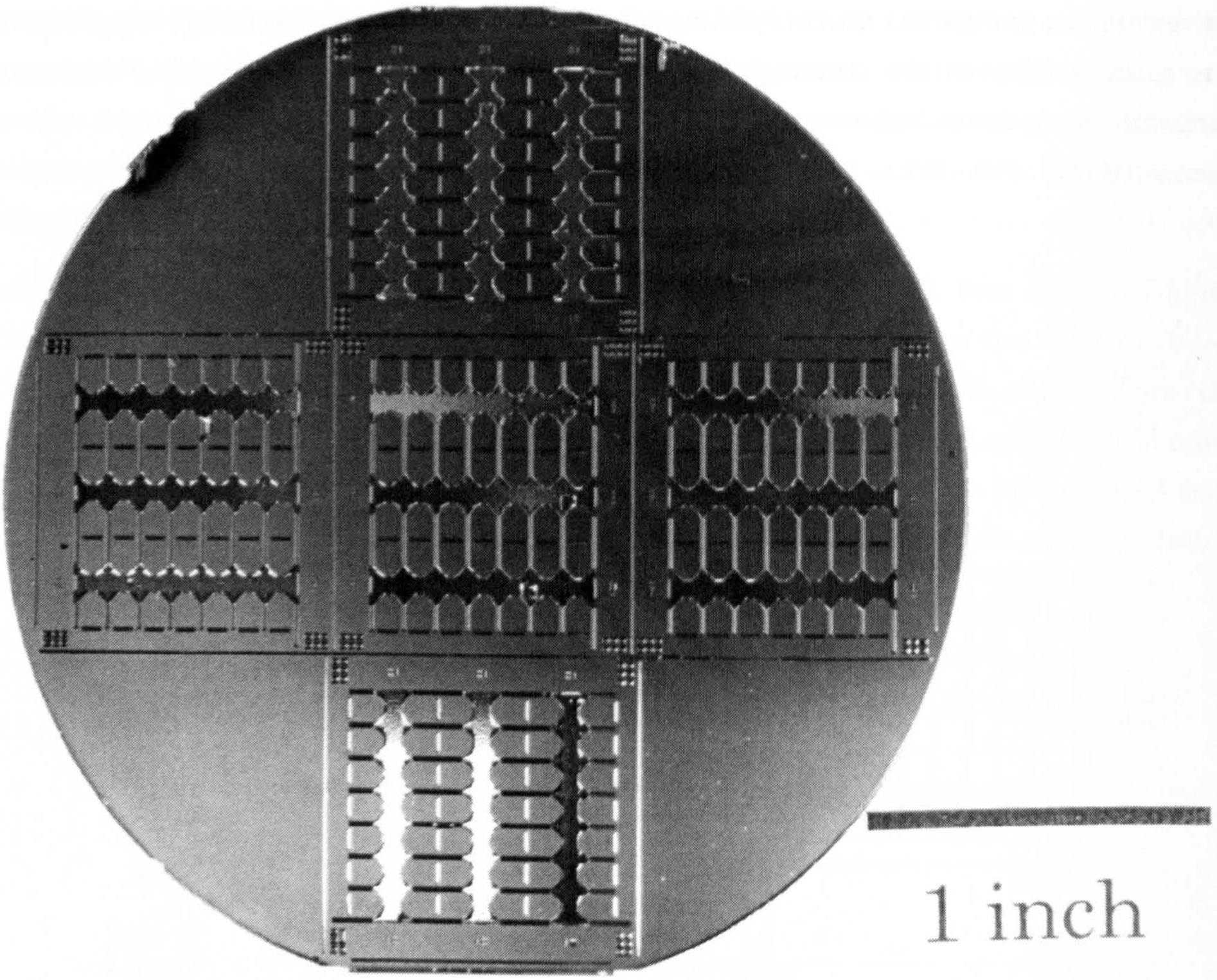


Figure F.13: Photograph of a whole 3 inch wafer after the second wet etch using mask patterns 'amm0006' and 'amm0007' of Figures F.11 and F.12.

These plates helped to improve throughput significantly. A photograph of a 3 inch wafer after the second wet etch is shown in Figure F.13. The dimensions of the 3 inch wafer were now so accurately known that the grating pattern, which consisted of 10 lines (20 pairs) at 40μm pitch at each wafer-flat edge, lined up with the wafer so that the middle pair of each grating was directly over the flats. However several problems still remained:

(a) The sizes and shapes of the through-wafer alignment markers were still not consistent, despite all efforts.

(b) The additional weakening of one of the substrate vertical cleave lines of each substrate meant that when the wafer was cleaved, the substrates would often cleave through this instead of along the wafer 'v'-grooves. This was because the substrates were much weaker at that edge combined with the close proximity of these weaker lines to the wafer cleave lines, and the amount of force required to cleave the wafer.

It was only at this point that the true cause of the through-wafer marker enlargement and distortion was discovered. The first clue was that the shape of the marker seemed to depend on the radial distance of the marker from the wafer centre, with the edges furthest from centre becoming the most elongated. This suggested a problem related to the spin coating of resist. The second clue was that all the through-wafer alignment

markers and membranes on the patterns exposed at the wafer corner pieces (for TEM substrates) using the previous mask plates seemed to turn out remarkably square and much closer to the calculated value. The final clue was that the widths of the wafer cleaving lines seemed to be closer to the mask pattern size than the widths of the through-wafer alignment marker openings after etching. This all indicated that the photoresist was not adequately coating the edges of the marker openings after the first level wet etch, due to the three dimensional nature of the etched marker geometry. This was confirmed by comparing the size of the marker openings before and after the second level wet etch: the growth in size was primarily occurring during the second level etch.

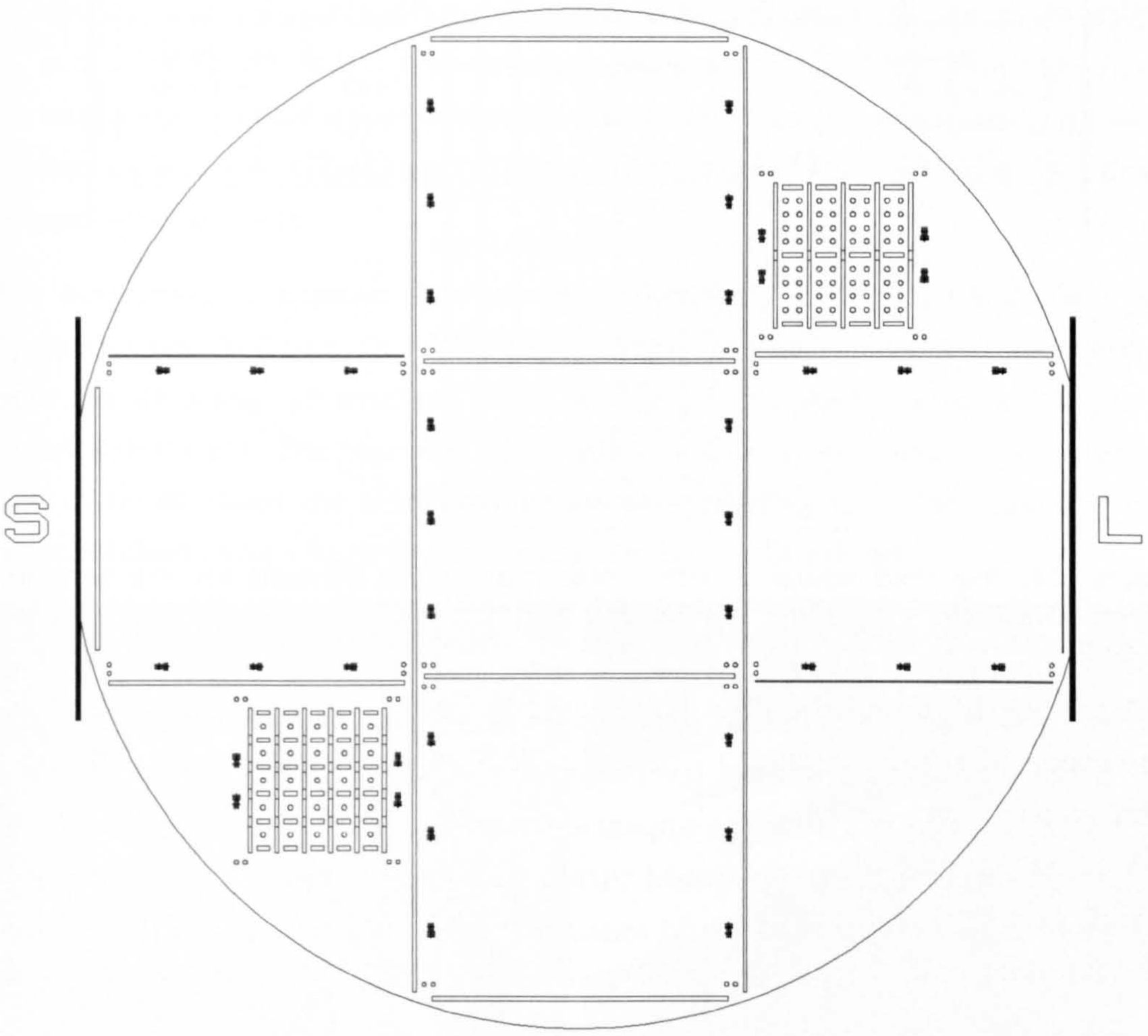


Figure F.14: The third whole 3 inch wafer mask plate devised for the first photolithography level. Mask pattern name 'amm0008'.

There was an obvious solution to this problem: to make the openings smaller in the first level than the second level so that the area of resist around the opening which is non-uniform is removed and the marker opening size is fully redefined by the second level pattern. Thus another set of 4 inch mask plates were required. The patterns for these are illustrated in Figures F.14 and F.15. These plates, amm0008 and amm0009, were written in January 1996.

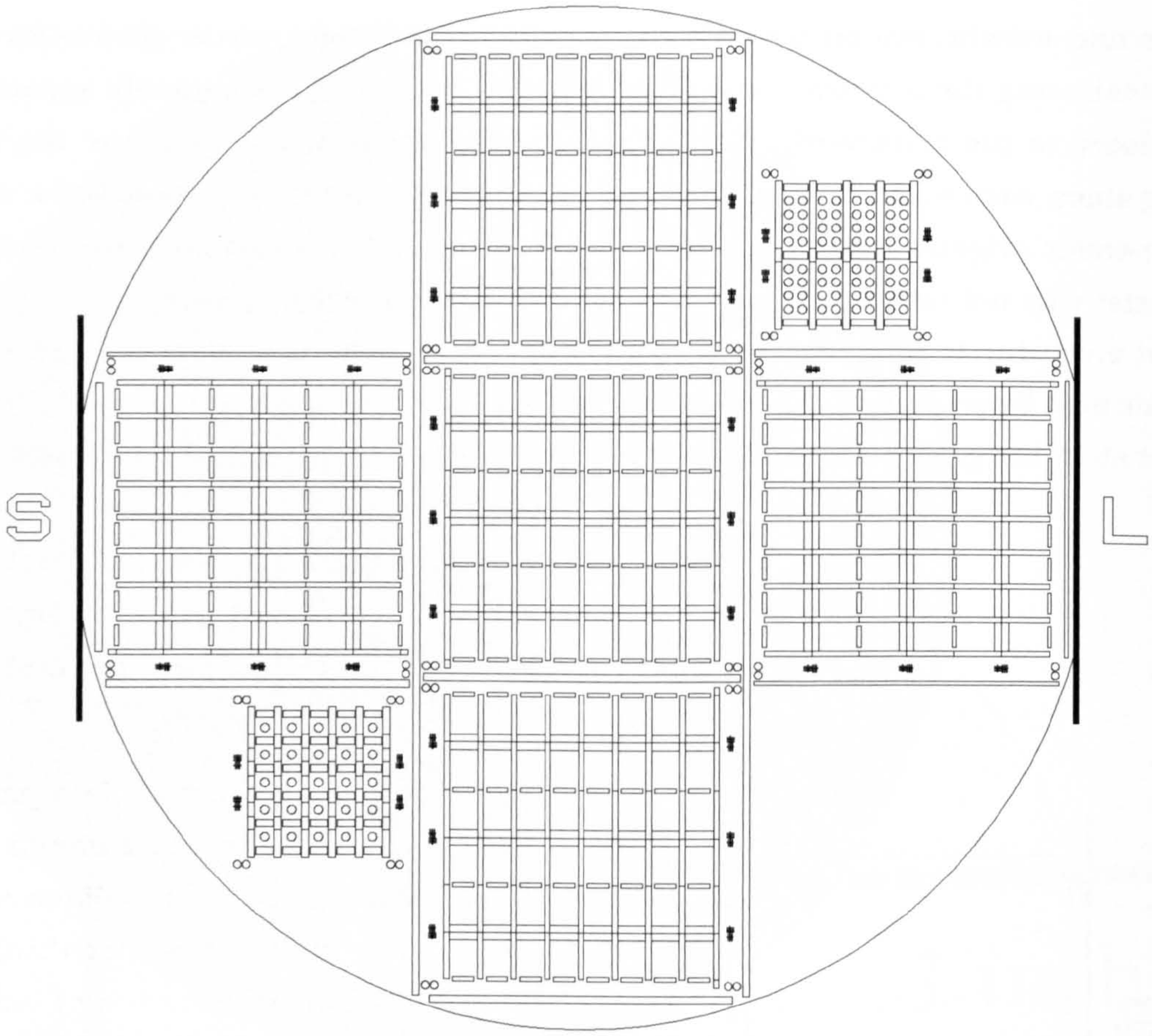


Figure F.15: The third whole 3 inch wafer mask plate devised for the second photolithography level. Mask pattern name 'amm0009'.

The main changes to the pattern were:

(a) The first level through-wafer alignment marker openings became 310μm circles with 570μm second level openings, the increase of 10μm being in anticipation of the markers appearing closer to their calculated size. The 310μm circular openings would turn into 310μm square openings during the first level etch which would mean that the distance between the corner of these markers and the second level 570μm openings would be about 66μm. Also, these 310μm openings would not have terminated by the end of the first etch so that no additional etching time would be required for the formation of the through-wafer alignment markers during the second etch by making use of convex corner undercutting.

(b) The number of markers at the corner of each substrate was reduced to two, again in anticipation of improved markers. The result of this being that the number of devices per substrate was increased to 48 by the addition of an extra column. This now brought the number of devices per wafer back up to 240.

(c) The vertical cleave line at one end of each substrate was made weaker than the others by a reduced extension of 600μm at each side (compared to 1250μm previously).

(d) The wafer cleave lines were also converted into a two stage scheme, with the full length cleave line openings being 250μm and 500μm wide in the first and second levels respectively (termination depth 354μm), and the shorter cleave line openings being either 300μm and 530μm (termination depth 375μm) or 125μm and 250μm (termination depth 177μm) in the first and second levels respectively, depending upon their location.

(e) A pattern for TEM membranes was reintroduced at one of the wafer corners, with another corner being occupied by a membrane sample for NSOM aperture tests (described in Chapter 6), leaving two corners free for use in other tests. These patterns were also both made to use a two stage scheme with the cleave line openings being 300μm and 530μm in the first and second levels respectively and the membrane openings being 375μm and 640μm (corresponding to 103μm membranes) respectively in the first and second levels.

All other dimensions remained approximately as in the previous mask plate pattern. These patterns resulted in a dramatic improvement in the quality of the through-wafer alignment markers with all markers being square and between 35μm and 45μm in size (calculated size 33μm). The tapering of the devices by convex corner undercutting also appeared to be of about the correct dimensions, resulting in a flat region of between 200μm and 400μm across from which the cantilevers could extend.

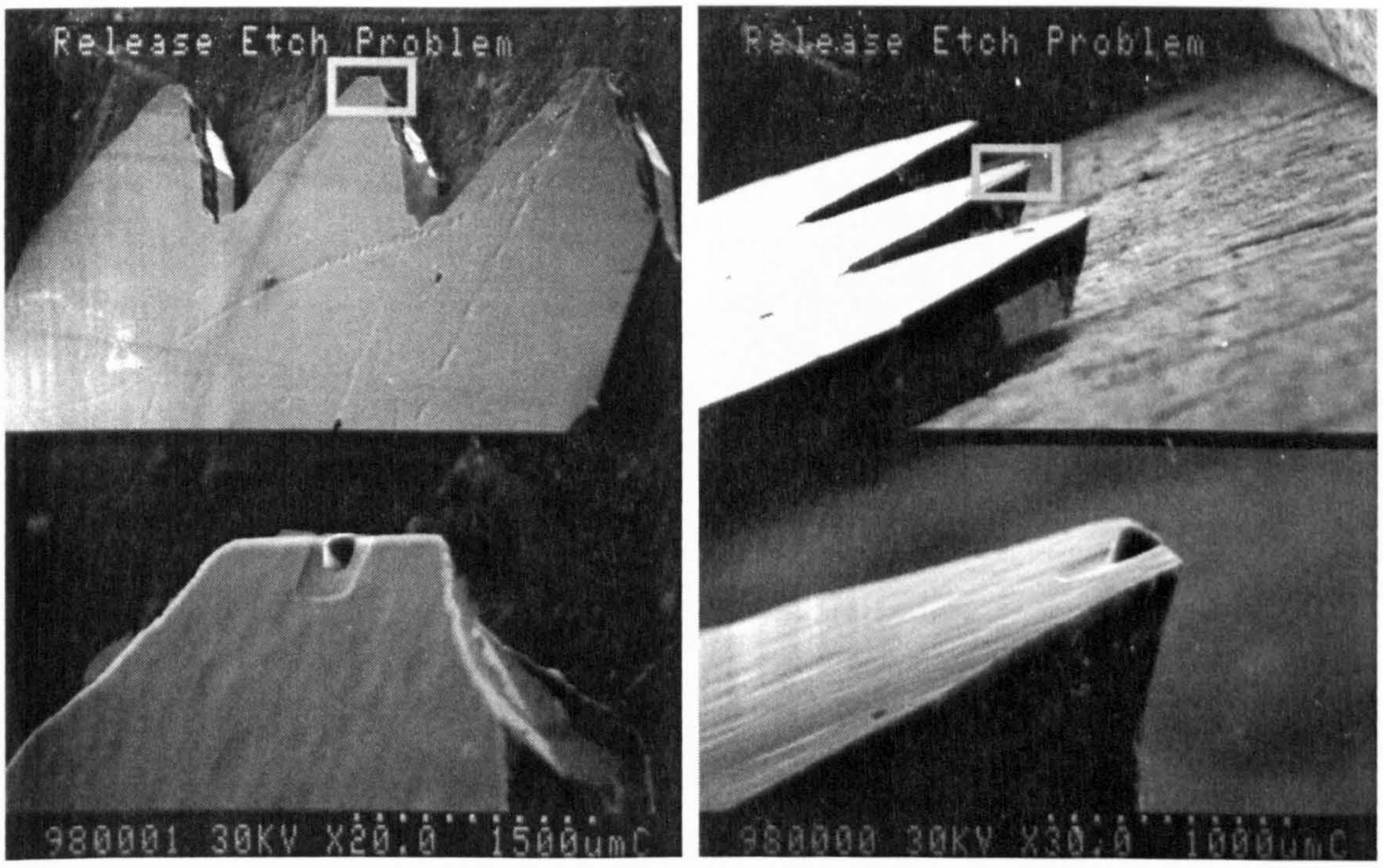


Figure F.16: Micrographs showing how the pyramidal tips were ending up on the substrate rather than at the end of a cantilever (not present here). The bottom half of both micrographs are a 10x magnification over the displayed scale which applies to the top half.

However, some new problems emerged:

(a) The 'cliff-to-cliff' spacing did not match the positions of the tips on the other side as it turned out that the etching was behaving much closer to theory than originally expected. Figures F.16 and F.17 show how the pyramidal tip was located on the substrate instead of being off the edge by a few hundred microns after a release etch (this etch is discussed in §4.4.2).

(b) There was some evidence of non-uniformity of dimensions across the wafer, most likely to have been caused by distortions introduced during the copying of 4 inch chromium mask plates to 4 inch ferric oxide mask plates.

(c) The vertical substrate cleave lines tended to result in cracks on the device side of the substrate, most probably induced by stresses caused from the clamping of the substrate during the pyramidal tip etching, electron-beam lithography and evaporation steps. These cracks would propagate until they reached the centre of the cantilever region of the substrate where they would then proceed across the middle of this region, ruining several pairs of devices. The cracks also reduced the alignment accuracy of the electron-beam lithography levels as the marker separation was altered. [This was despite measures being taken in the first electron-beam lithography level to prevent this. §4.3 discusses these measures in more detail.]

(d) Individual devices were still proving hard to cleave.

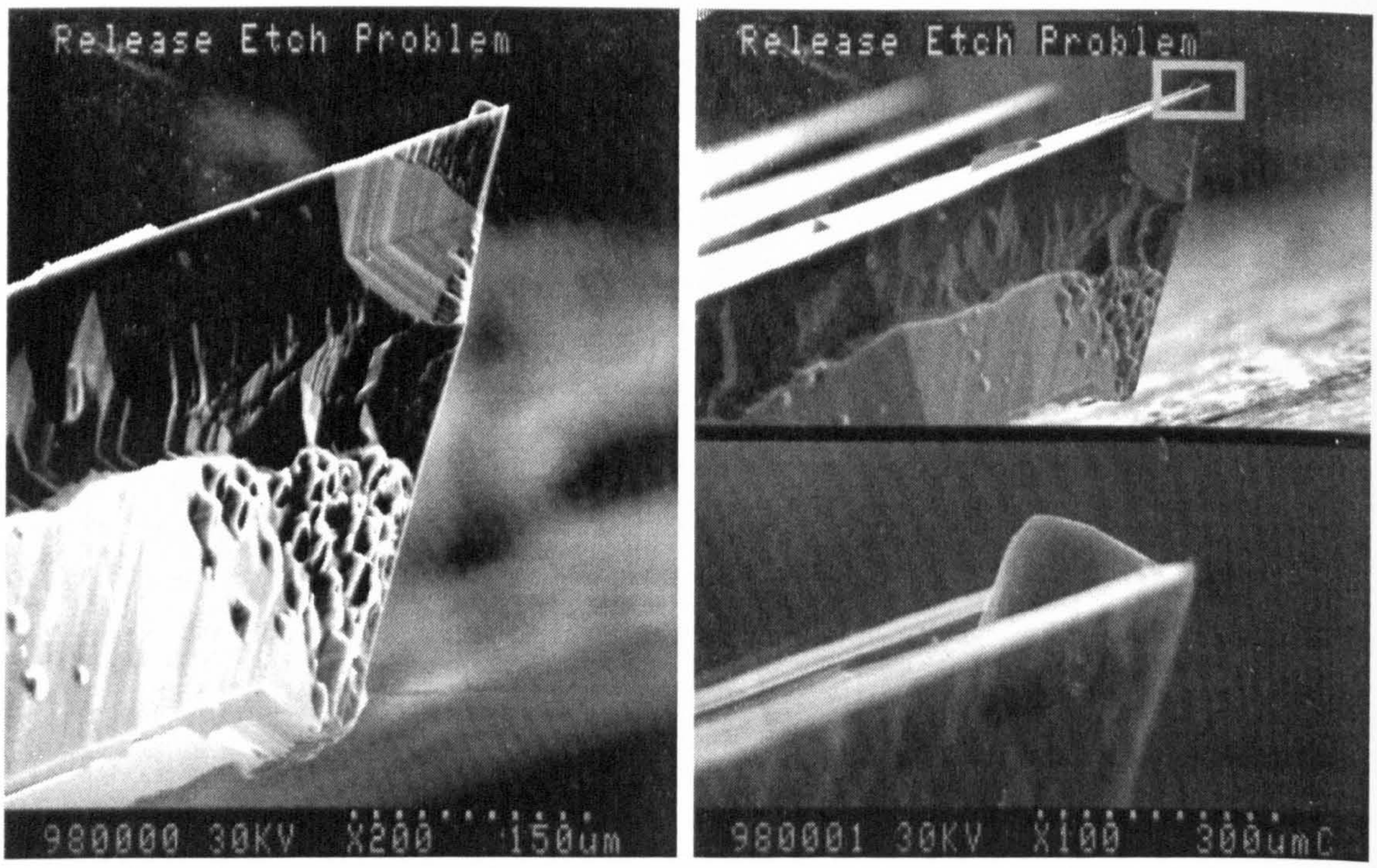


Figure F.17: Micrographs showing how the pyramidal tips were ending up on the substrate rather than at the end of a cantilever (not present here). A more oblique view of the sample shown in Figure F.16. The bottom half of the right micrograph is a 10x magnification over the displayed scale which applies to the top half.

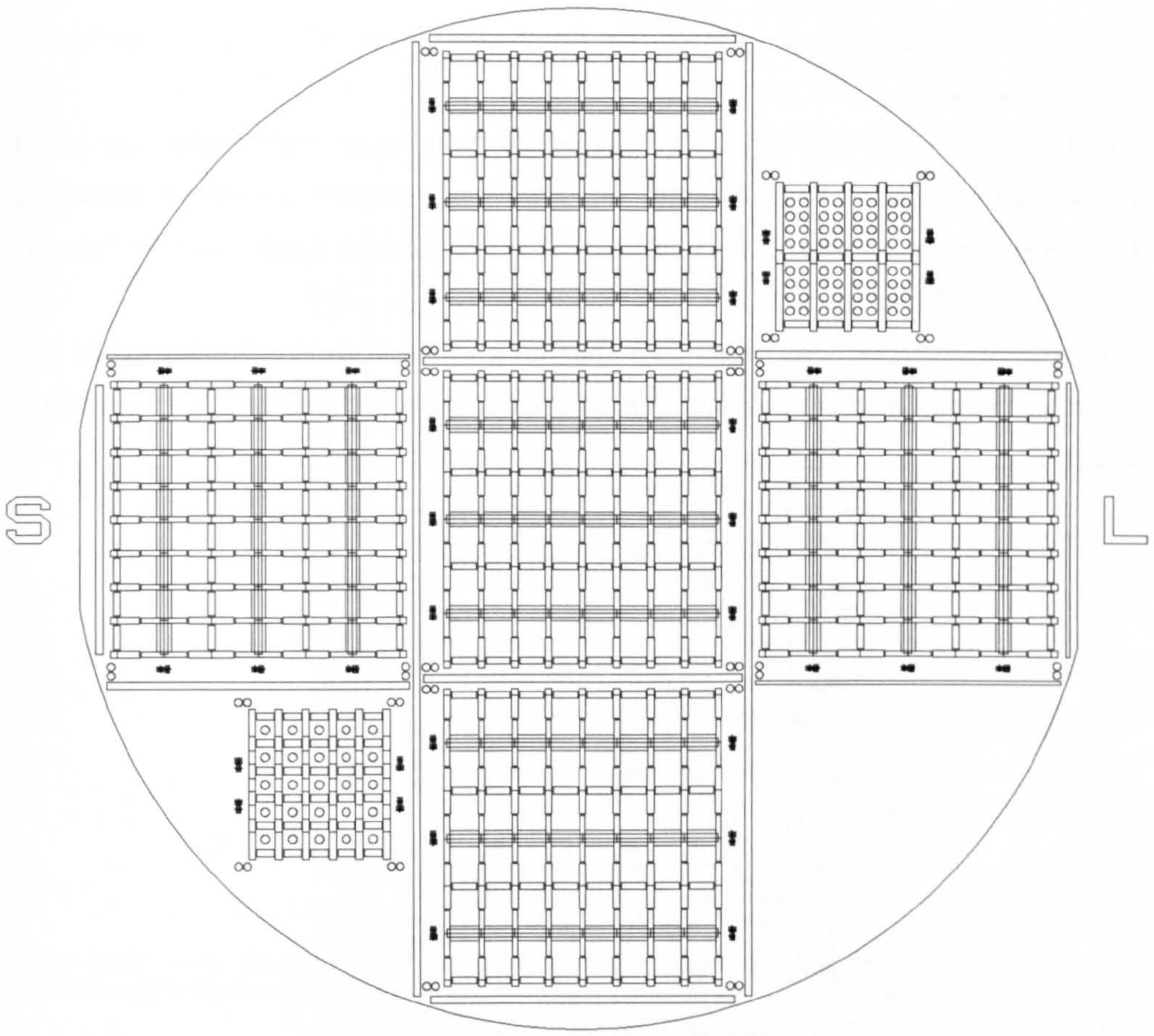


Figure F.18: The fourth whole 3 inch wafer mask plate devised for the second photolithography level. The logos at the back of each device are not shown here for clarity. Mask pattern name 'amm0011'.

To resolve some of these issues, as a significant number of devices were being lost, especially due to problem (c), another second level mask plate was produced. The pattern for this is shown in Figures F.18 and F.19. This mask plate, amm0011, was written in July 1996. The main changes are:

(a) A reduction in the cliff-to-cliff spacing to 1100μm (590μm membrane length per pair of devices) from 1500μm. This would allow cantilevers up to 250μm long to be produced whilst reducing the area of the fragile part of the substrate.

(b) The introduction of a 100μm wide (637μm at the other side of the wafer after etching through) crack arrester at the midpoint of each device along the vertical device cleave lines. This was meant to stop cracks from propagating along these cleave lines.

(c) The widening of the vertical device cleave lines below the crack arrester *i.e.* not the tapered edge of the device, and the horizontal device cleave lines to 500μm (termination depth 354μm) to make cleaving of individual devices easier. The vertical device cleave line above the crack arrester remained at 400μm wide as this determined the amount of tapering by convex corner

undercutting and the length of the flat region from which the cantilevers would extend.

(d) As the devices at the top and bottom rows of each substrate had been particularly difficult to cleave, all vertical device cleave lines were extended by 500μm beyond the centre line of the horizontal device cleave line at each end of the substrate.

(e) The university logo was placed at the back of each device, digitised in the form of 10μm squares on a 20μm pitch (approximately) for branding purposes.

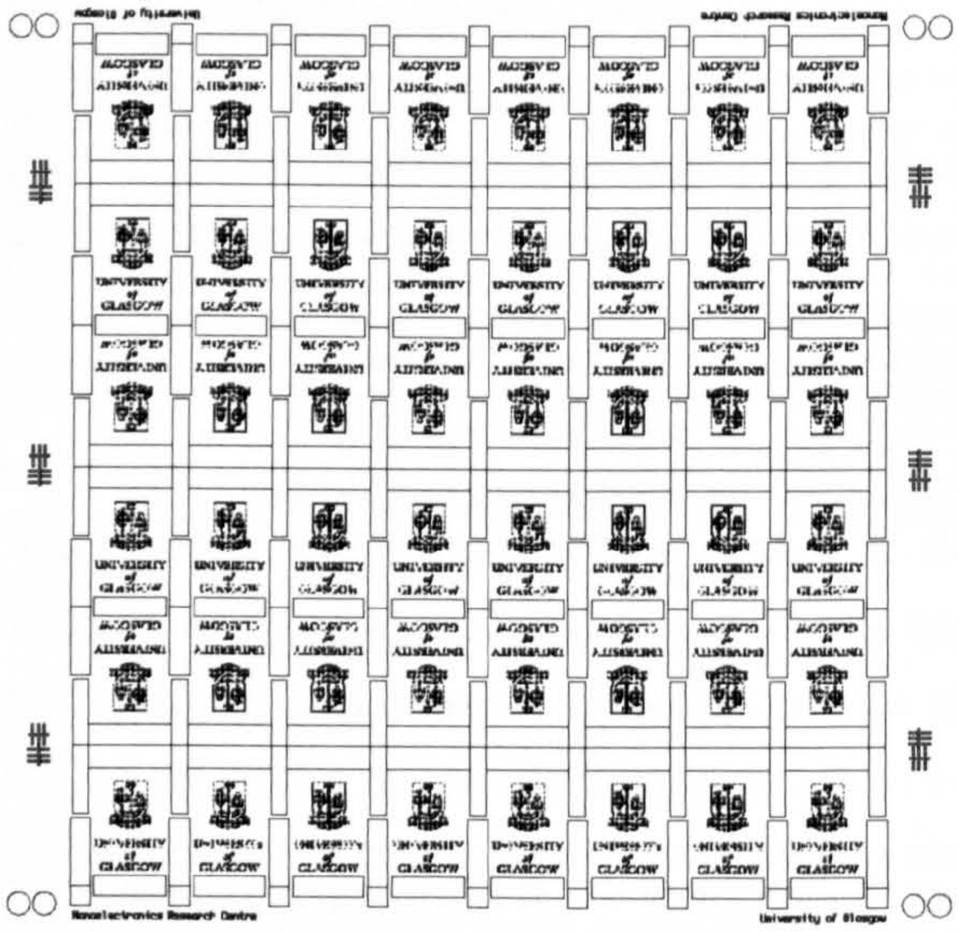


Figure F.19: A close-up of one of the substrate patterns of the mask plate ('amm0011') illustrated in Figure F.18, showing the logos at the back of each device.

This mask set became the most successful to date, however additional problems emerged:

(a) The wafer cleave lines were possibly a little too weak as the wafer would occasionally cleave during the second wet etch.

(b) The maximum cantilever length of about 250μm was too restrictive in terms of future device development.

(c) The non-uniformity of dimensions due to the use of a 4 inch ferric oxide plate was still causing problems.

(d) The pattern layout was not compatible with full 3 inch electron-beam lithography as the markers would have to be outside the patterned region and structures would have to be defined to enable the markers to be found easily on a 3 inch wafer.

Thus a final mask set was designed to address all of these issues. The first level would be carried out with a 4 inch ferric oxide mask plate while the second level, which defines all the opening sizes, would be carried out directly with a 4 inch chromium mask plate. The resulting plate patterns are shown in Figures F.20, F.21 and F.22. These plates, amm0012 and amm0013, were written in October 1996.

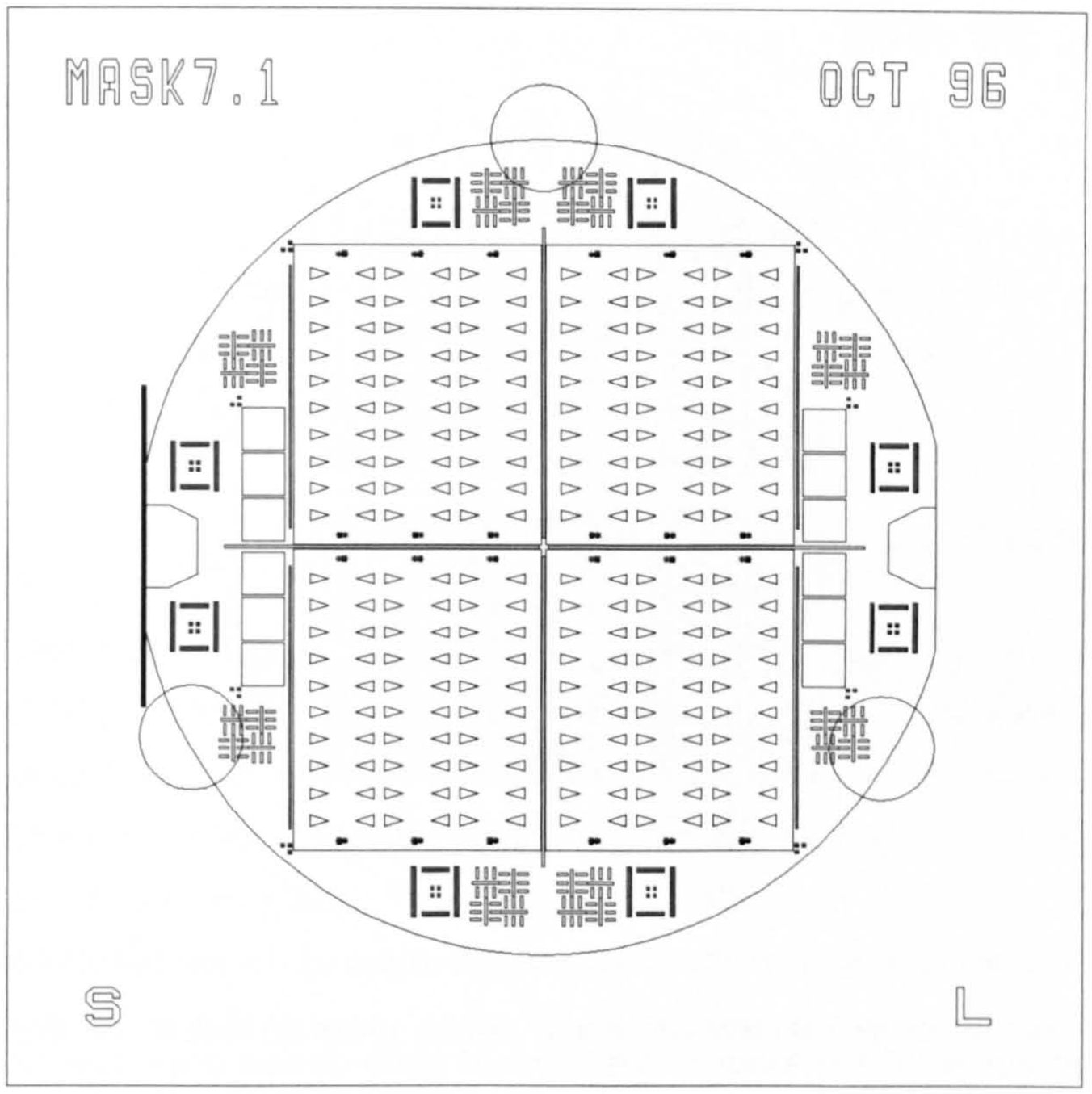


Figure F.20: The final whole 3 inch wafer mask plate devised for the first photolithography level. Mask pattern name 'amm0012'.

The positions of height reference blocks and wafer clamps as used by the beamwriter have been taken into account during design and can be seen in Figure F.20. The layout has been changed so that the first electron-beam lithography level must be carried out on a whole wafer. The pattern is based on dividing the wafer into quarter wafer substrates, each with 6 rows and 10 columns of devices, yielding 240 devices per wafer. The ferric oxide mask pattern is based on a mirrored version of the chromium mask pattern, as required for this scheme. The wafer cleave lines have been narrowed, the wider ones being 250μm and 400μm in the first and second levels respectively, the narrower ones being 150μm and 300μm in the first and second levels respectively. A grating has been placed along the small flat only (and only in the first level pattern) as the large flat is used by the beamwriter for coarse rotational location in the wafer holder.

To allow the second level alignment to be carried out with a chromium mask plate directly, a 5mm window was opened around the wafer (to make the plate transparent) with a 2mm boundary around the edge of the wafer as shown in Figure F.21. This allowed coarse placement to within ±2mm. A 5 times enlarged version of the 'tree structure' (approximately 4mm across) could then be used to align the plate laterally to within 125μm in each direction. Finer alignment was then provided by the first level through-wafer alignment marker openings, which remained at 310μm, with the finest alignment still being provided by non enlarged 'tree structures'. The second level through-wafer alignment marker opening was reduced to 560μm (calculated marker size 23μm).

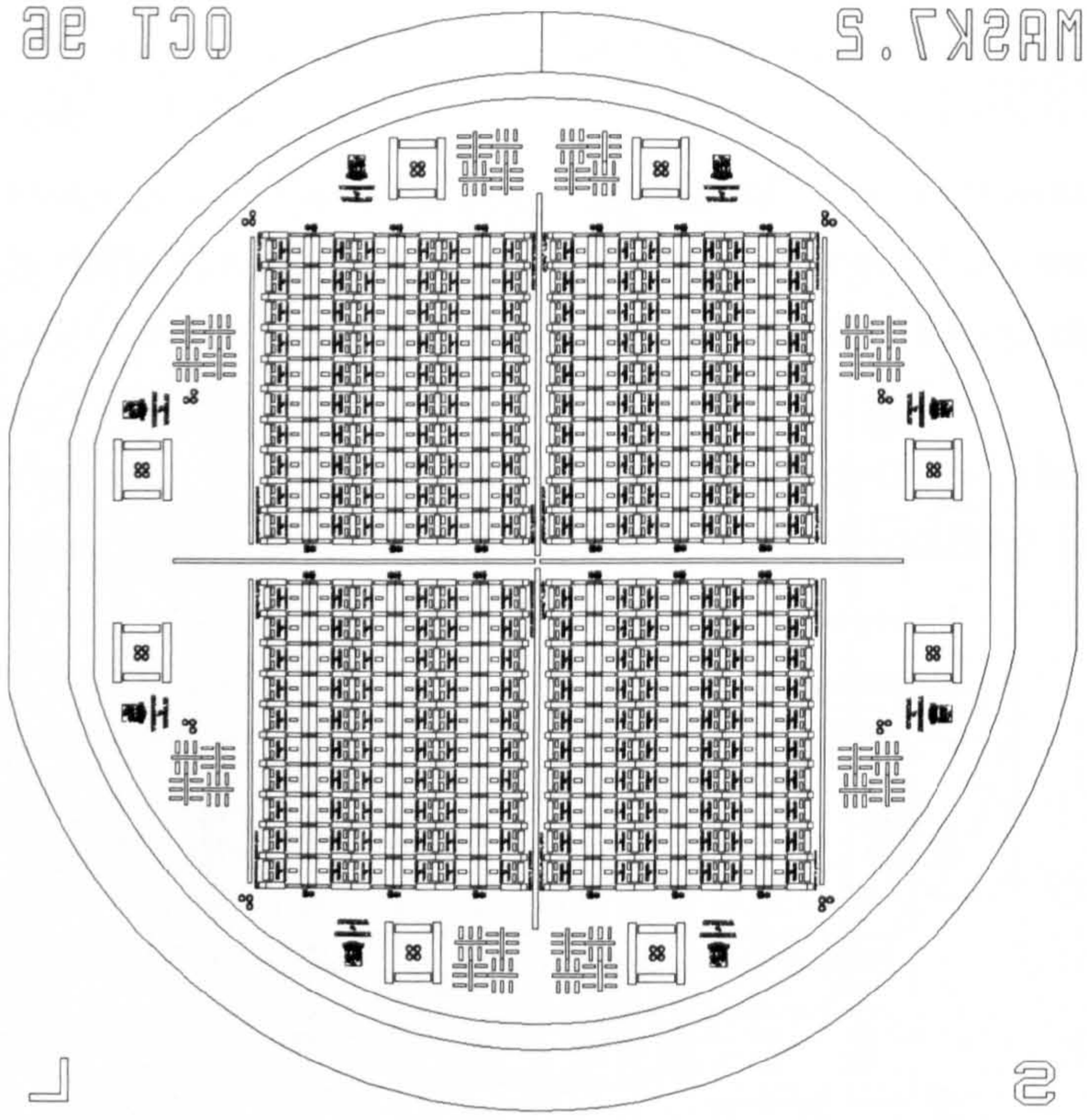


Figure F.21: The final whole 3 inch wafer mask plate devised for the second photolithography level. The logos at the back of each device are shown here. Mask pattern name 'amm0013'.

The 'cliff-to-cliff' spacing was increased by 200μm to 1300μm (allowing up to 350μm long cantilevers to be produced). A change was also made to the width of the vertical device cleave lines above the crack arrester. This was reduced by 100μm to 300μm since the previous change in the 'cliff-to-cliff' spacing had resulted in the taper of each device becoming so strong that there was no flat platform from which the cantilever could extend. The crack arrester was also widened to 150μm to make each substrate even more resilient against cracks, as an increased number of devices would now share the same fate.

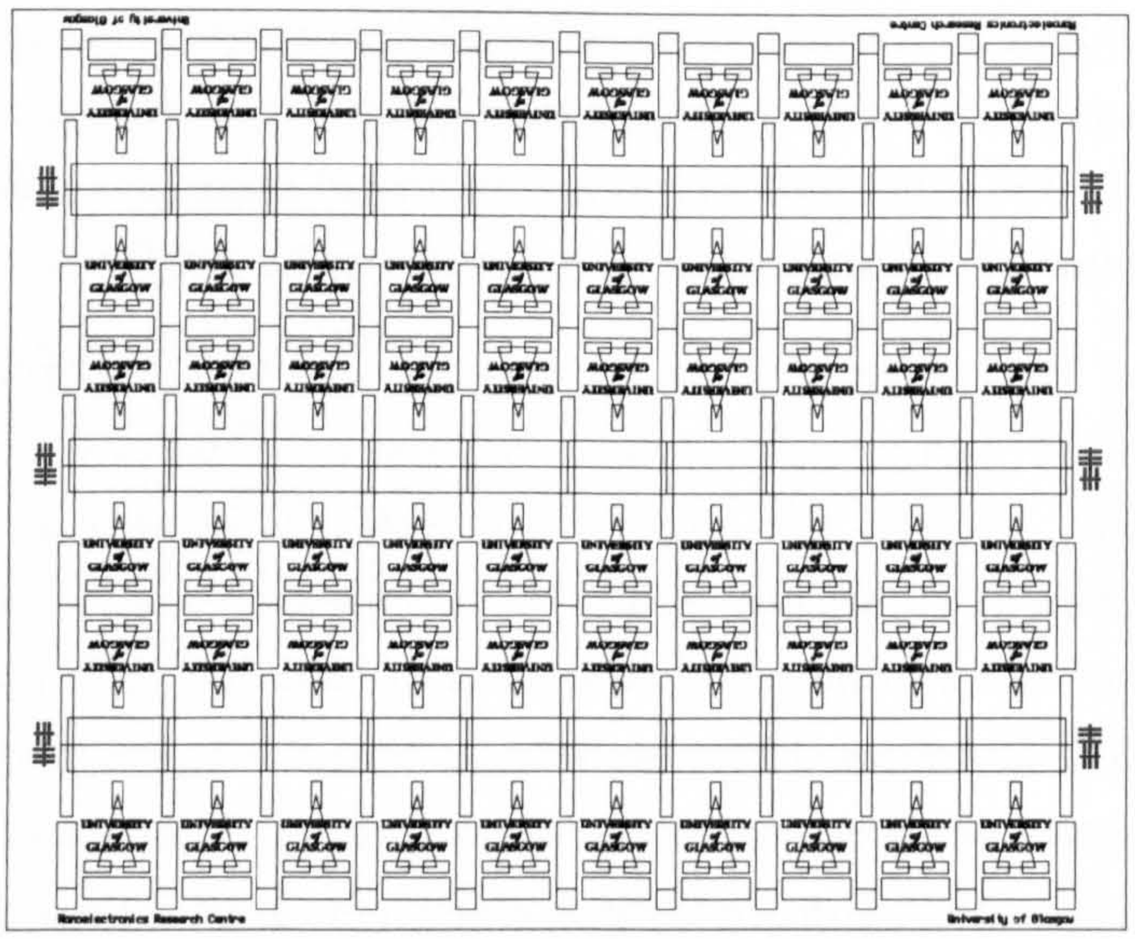


Figure F.22: A close-up of one of the substrate patterns of the mask plate ('amm0013') illustrated in Figure F.21 showing the logos at the back of each device.

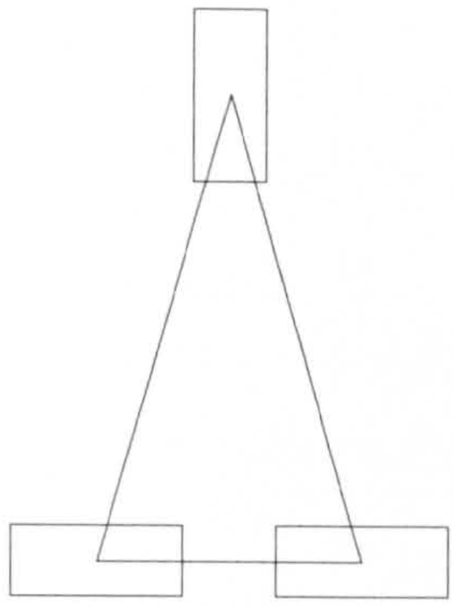


Figure F.23: A close-up of the etched grooves used for kinematic location. The triangle connecting the centres is only shown as a guide and is not written.

The full university logo at the back of each device was replaced with a pseudo-kinematic arrangement of 'v'-grooves for use in device mounting and a reduced university logo consisting of only the text without the emblem. This is shown in greater detail in Figure F.23. The width of the groove is based on the size of commercially available (from Goodfellow, Cambridge, England) ruby spheres, 300μm in diameter. Other features, including an additional set of through-wafer alignment markers, were also created as an aid to the electron-beam lithography levels. Their purpose is discussed in §4.3.2.3.

With this mask set, the main development phase of the photolithography levels ended. [A photograph of a whole 3 inch wafer after the second wet etch can be seen in Figure 4.2.4.]

REFERENCES

1. R. Harkins, *Private Communication*, (1993-1996).

2. R. Harkins, D. MacIntyre and S. Thoms, *Development of an in-house 4 inch chromium mask to ferric oxide mask copying process.*, (1993-1996).

APPENDIX G

THE SUBSTRATE FABRICATION PROCESS

G.1 THE PHOTOLITHOGRAPHY LEVELS

The final photolithographic process, using 3 inch silicon wafers with 100nm of LPCVD silicon nitride on both sides, is as follows:

(a) Clean the 3 inch wafer using an inorganic followed by organic clean [App. D].

(b) Evaporate 100nm of nichrome (60% nickel, 40% chromium) onto the polished side of the wafer (for protection) [§4.2.1].

(c) Clean the wafer using an organic followed by an inorganic clean [App. D].

(d) Carry out first level photolithography on the unpolished side of the wafer using the 4 inch ferric plate (mask 7.1 = 'amm0012') [§4.2.3].

(e) Dry etch unpolished side LPCVD silicon nitride for 5 minutes or more [App. E.2].

(f) Strip the resist and carry out an inorganic clean [App. D].

(g) Deoxidise the wafer in unbuffered hydrofluoric acid:water 1:4 for 1 minute [App. A].

(h) Wet etch in 7 molar (+5%) potassium hydroxide solution at 80°C for 80 minutes (etch depth of about 100μm expected) [App. A].

(i) Neutralise the wafer in sulphuric acid:water 1:4 for 1 minute followed by a thorough water rinse [App. A].

(j) Clean the wafer using an organic followed by an inorganic clean [App. D].

(k) Carry out second level photolithography on the unpolished side of the wafer using the 4 inch chromium plate (mask 7.2 = 'amm0013') [§4.2.3].

(l) Dry etch the unpolished side LPCVD silicon nitride for 5 minutes or more [App. E.2].

(m) Strip the resist and carry out an inorganic clean [App. D].

(n) Deoxidise the wafer in unbuffered hydrofluoric acid:water 1:4 for 1 minute [App. A].

(o) Wet etch in 7 molar (+5%) potassium hydroxide solution at 105°C for 70 minutes (etch depth of about 300μm expected) [App. A].

(p) Neutralise the wafer in sulphuric acid:water 1:4 for 1 minute followed by a thorough water rinse [App. A].

(q) Clean the wafer using an inorganic clean [App. D].

After this process the wafer is ready for the first electron-beam lithography level.

G.2 THE FIRST EBL LEVEL

The final first level EBL process using whole 3 inch wafers is as follows:

(a) The wafer is etched using chrome etch for ten minutes to remove the nichrome protection layer [§4.2.1].

(b) The wafer is cleaned using an inorganic clean followed by an organic clean [App. D].

(c) 8% equivalent thickness Aldrich is then spin coated at 5000 rpm for 60 seconds onto the polished side of the wafer [Table 3.5.1].

(d) The resist is dried by baking for one hour at 180°C.

(e) 4% Elvacite is then spin coated at 5000 rpm for 60 seconds onto the previous layer of resist [Table 3.5.1].

(f) The wafer is then prebaked at 180°C for at least six hours (usually overnight).

(g) The pattern (AM3INRPDEVS???_07 for contact mode devices or AM3INATDEVS???_07 for non-contact mode devices, with '???' replaced by the required cantilever length of 100μm, 150μm, 200μm, 250μm, 300μm and 350μm for the first matrix; AM3INMARKERS_15 for the second matrix; and, AM3INPYRCAL_18 for the optional third matrix) is then exposed at 50kV using a dose of 300 to 350μC/cm^2 and a spot size of 112nm for the first matrix (pyramidal tips and marker inside edges), a dose of 300 to 350μC/cm^2 and a spot size of 400nm for the second matrix (crosses, crack arresters and marker plateaux) and optionally a dose of 200 to 450μC/cm^2 and a spot size of 400nm for the third matrix (in-situ calibration pattern).

(h) The pattern is then developed in MIBK:IPA 1:1 for 35 seconds.

(i) 50nm of nichrome (60% nickel/40% chromium) or 10nm nichrome, 20nm gold and 100nm titanium is then evaporated onto the resist using the Plassys MEB450 [§3.6.2].

(j) Lift-off is then performed using acetone, followed by a rinse in IPA and blow dry.

(k) The wafer is then etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 5 minutes or more [App. E.3] in the Plasma Technology BP80 [§3.6.4].

(l) A calibration sample is cleaned using an inorganic clean [App. D] and deoxidised for 60 seconds in unbuffered hydrofluoric acid:water 1:4 [§4.3.1].

(m) The calibration sample is then etched for 30 to 60 minutes in 20% (by weight) potassium hydroxide solution:IPA 3:1 at 70°C [§4.3.1].

(n) The calibration sample is then removed from the etch, rinsed in water and neutralised in nanostrip:water 1:1 for several minutes.

(o) The calibration sample is then rinsed in water followed by IPA and blown dry.
(p) The required etching time is calculated from measurements of the calibration sample using an optical microscope [§4.3.1].
(q) Step (l) is repeated for the wafer. If in-situ calibration is required, the titanium layer will have been stripped off during this step, leaving a gold layer [§4.3.2.3].
(r) Step (m) is repeated for the wafer using the calculated etch time, or in-situ calculated time.
(s) Step (n) is repeated for the wafer.
(t) The wafer is then rinsed in water and placed in unbuffered hydrofluoric acid:water 1:4 for 1 to 3 hours until the LPCVD silicon nitride has been completely etched away [§4.3.1].
(u) The sample is then rinsed in water thoroughly before being blown dry directly from water (hydrophobic surface).

After this process the wafer is then ready for the second EBL level.

G.3 THE SECOND EBL LEVEL

The final second level EBL process using whole 3 inch wafers is as follows (assuming that for non-contact mode devices, the cantilever membrane has been thinned to between 10μm and 20μm between the second photolithography level and the first EBL level using an additional etch in 7 molar (+5%) potassium hydroxide solution at 65°C):

(a) Between 250nm and 750nm of stress free PECVD silicon nitride is deposited onto the (pyramidal tip side of the) wafer [§3.6.6].
(b) 100nm of chromium is sputtered onto the wafer.
(c) 8% equivalent thickness Aldrich is then spun onto the wafer at 2500 rpm for 60 seconds [Table 3.5.1].
(d) This resist is then dried by prebaking at 180°C for at least 6 hours (usually overnight)
(d) 4% equivalent thickness Aldrich is then float coated onto the wafer [§5.2].
(e) This resist layer is then dried by baking at 80°C for 30 minutes.
(f) The pattern is then exposed at 50kV using one of the global EBL markers with a spot size of 400nm and a dose of 350μC/cm^2. This pattern is comprised of two layers: the layer (layer 2) containing the cantilever pattern (including the shoulder and tapered region below) is subtracted from a layer (layer 9) containing a pattern consisting of long rectangular strips enclosing the cantilever patterns.
(g) The pattern is then developed in MIBK:IPA 1:1 35 seconds.

(h) The chromium layer is then etched using chrome etch for 5 to 10 minutes or until it is clear [§3.7.3].

(i) The resist is then stripped in acetone and the sample inspected under the microscope: if the chromium at the apices of the pyramidal tips is still in tact, then processing continues from step (t) below, otherwise the intermediate steps must also be carried out.

(j) 12% equivalent thickness Aldrich is then spin coated onto the substrate at 2500 rpm for 60 seconds [Table 3.5.1].

(k) The resist layer is then prebaked for at least 6 hours (usually overnight) at 180°C.

(l) 4% equivalent thickness Aldrich is then float coated onto the wafer [§5.2].

(m) This resist layer is then dried by baking for 30 minutes at 80°C.

(n) 2% Elvacite is then float coated onto the wafer [§5.2].

(o) The resist is then dried by baking for 30 minutes at 80°C.

(p) A pattern consisting of 15μm squares placed at the apices of each pyramidal tip is then exposed at 50kV by using one of the global EBL markers for alignment and with a spot size of 400nm, a dose of 500μC/cm^2 and defocus of the pyramidal tip height [§5.1].

(q) The pattern is then developed in MIBK:IPA 1:1 for 35 seconds.

(r) 50nm of nichrome (60% nickel/40% chromium) is then evaporated onto the resist using the Plassys MEB450 [§3.6.2].

(s) Lift-off is then performed using acetone, followed by a rinse in IPA and blow dry.

(t) The wafer is then etched with 15mT of C_2F_6, 100W of r.f. power, a flow of 20 sccm and a self bias of -400V for 15 minutes or more in the Plasma Technology BP80 [§3.6.4].

(u) The chromium layer is then stripped using chrome etch for approximately 10 minutes or until clear [§3.7.3].

contact mode devices:

(v) The wafer is then cleaned using an inorganic clean followed by an organic clean [App. D].

non-contact mode devices:

(v) The wafer is then cleaned using an inorganic clean [App. D].

both devices:

(w) The wafer is then etched in 7 molar (+5%) KOH at 65°C to release the cantilevers [§4.4.2 and §4.4.3].

(x) The wafer is then removed from the etch, rinsed in water and neutralised in sulphuric acid:water 1:4 for 60 seconds followed by a thorough water rinse [App. A].

(y) The wafer is then cleaned using an inorganic clean followed by an organic clean [App. D].

Notice that in step (w) no deoxidation is carried out because the PECVD silicon nitride is attacked extremely rapidly (several hundred nanometres per minute) by hydrofluoric acid.

APPENDIX H

BEAMWRITER ALIGNMENT MARKERS

H.1 FIRST LEVEL EBL ALIGNMENT

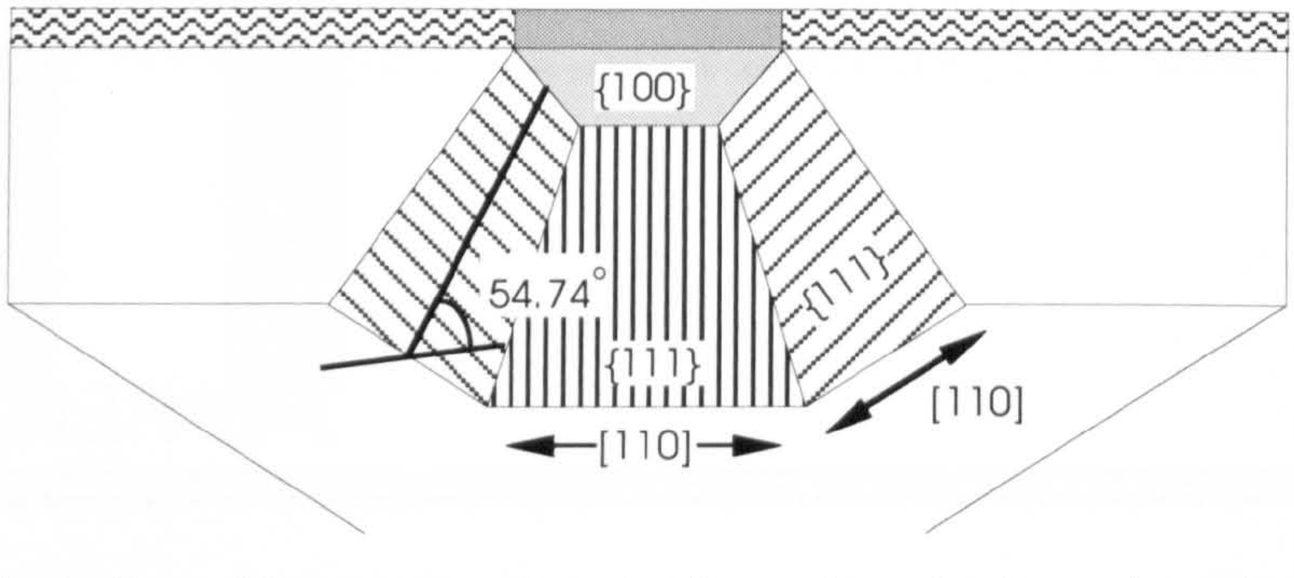


Figure H.1: Illustration of the cross-sectional oblique view of a through-wafer alignment marker. The thickness of the silicon nitride layer has been exaggerated for clarity.

One problem that was encountered during the development of the first EBL level was related to the ability of the beamwriter to find through-wafer markers. It is clear that these markers are rectangular and of negative contrast from Figure H.1, as the substrate is extremely thin over the marker resulting in a marked drop in the number of backscattered and secondary electrons over the surrounding material. However, the beamwriter had many problems finding these markers, until it was discovered that the beamwriter measurement of the marker size differed substantially from the optical size. This discovery, made by Dr. Steve Thoms, explained the high failure rate of automatic marker alignment at this level. The explanation for this effect comes from the geometry of the through-wafer markers and the penetration depth of 50kV electrons into the substrate. As the edges of the marker slope away from the centre, along the silicon {111} planes, and the penetration depth of 50kV electrons into the substrate is of the order of 10μm [Thoms[1] (1993-1996)], the edge of the marker is not where it appears optically but is around 7.5μm further out from each optical edge. Thus it was found that by adding 10μm to 15μm to the optical marker size, the beamwriter was able to find these markers without difficulty. For 100kV alignment, the penetration depth is approximately double that of 50kV [Thoms[1] (1993-1996)] so that adding 25μm to 30μm to the optical marker size seemed to work well.

H.2 SECOND LEVEL EBL ALIGNMENT

Similar problems to those encountered with first level EBL alignment markers [see §H.1] were also encountered with alignment to the topological markers for the second EBL level. Figure H.2 shows the pattern used to create topological markers. These markers are simply inverted pyramidal etch pits located in a plateau so that they can be defined simultaneously with the pyramidal tips. This ensures accurate placement of markers with respect to the tips. The isolated plateaux are located several millimetres from the tips because the tips need to become the most extreme point on the substrate in order to have good access to samples during AFM.

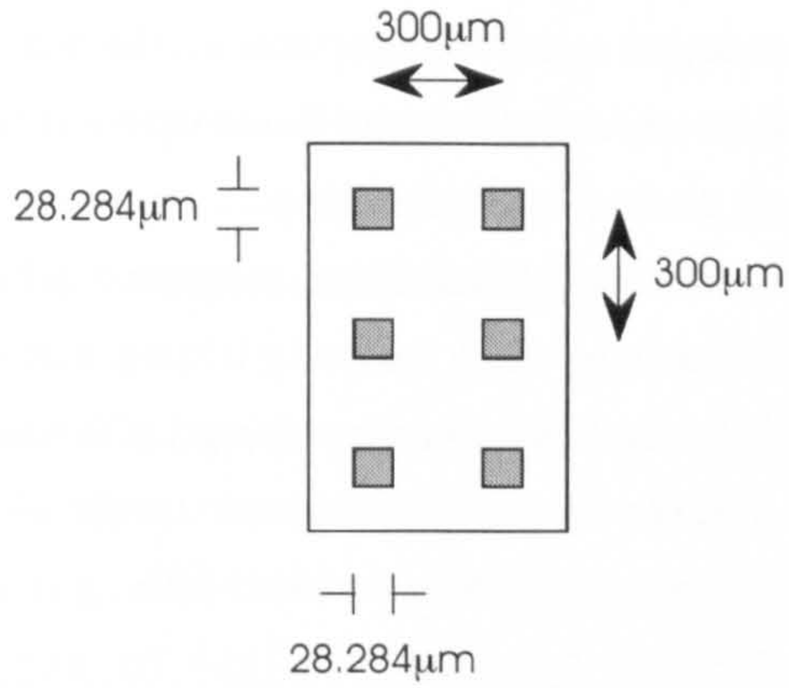


Figure H.2: Diagram illustrating the pattern used to form topological alignment markers. The large rectangle represents the masked area which forms the plateau, in which the six grey windows are opened as markers. These form inverted pyramids upon wet etching.

As mentioned previously in §3.5.2, the beamwriter only accommodates one fixed size of topological marker, 20μm, and one contrast setting, negative. Originally the beamwriter seemed to have great difficulty finding these markers automatically. For the second level EBL pattern only three markers, located globally at substrate corners, were required to be found, but this more often than not proved to be too many. The result was that manual alignment had to be used for this level, or Dr. Steve Thoms would have to adjust beamwriter parameters that would not normally require altering until the markers could be found automatically.

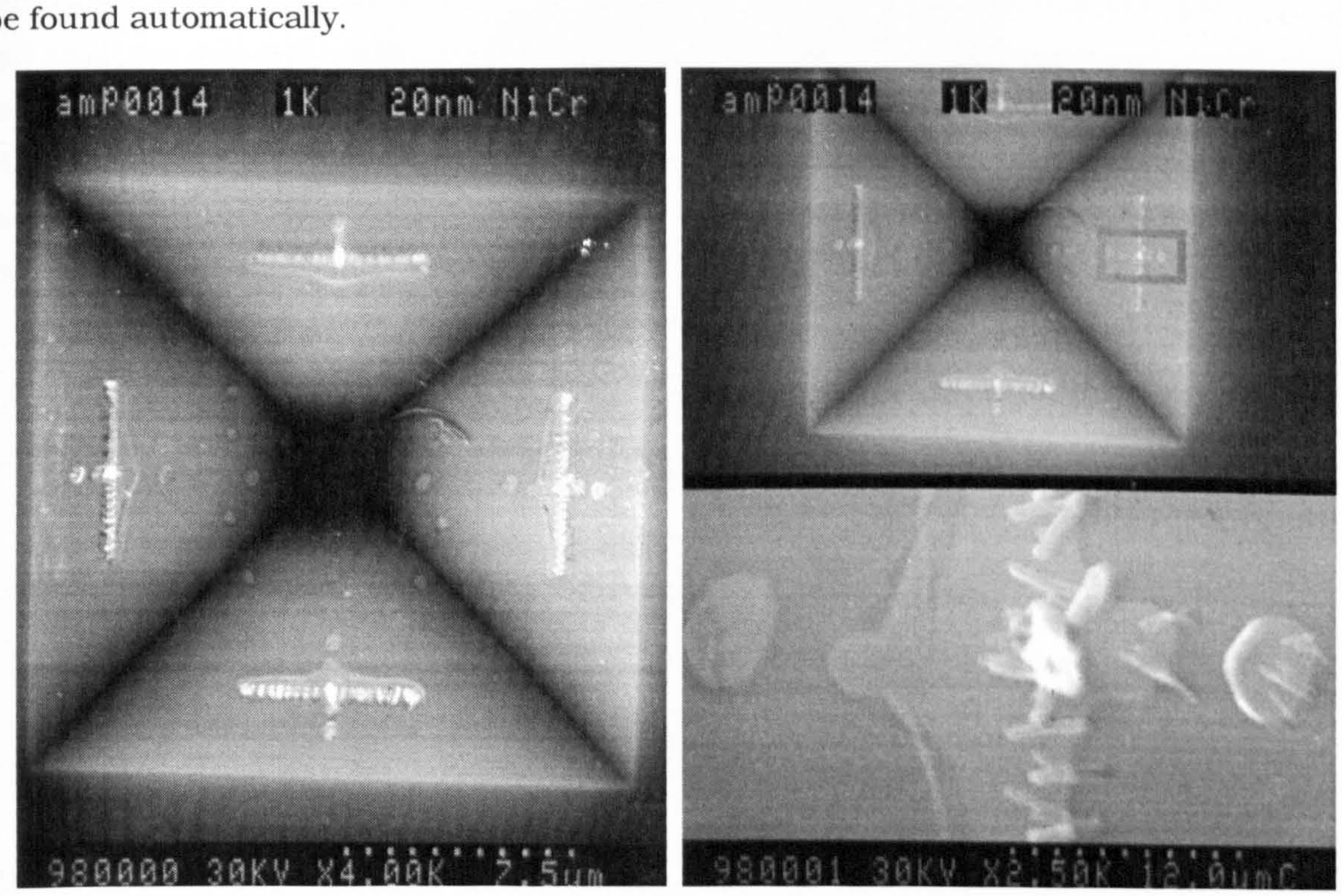


Figure H.3: Micrographs showing where the beamwriter defines the edge of a topological marker to be. During alignment the beamwriter exposed the marker, the subsequent lift-off highlighted the problem. The bottom half of the right image is a 10x magnification over the displayed scale which applies to the top half.

As with the through-wafer alignment markers the first indication that something odd was occurring was that the beamwriter measurement of the size of these markers was 14μm.

This was confirmed by examining a used marker after a lift-off level had been performed, as shown in Figure H.3. It was clear that a similar slope problem as with the through-wafer markers [see §H.1] was occurring here also. This was caused to the beamwriter using the point at which the electron signal level drops by a certain level as the edge of the marker [Rishton, Beaumont and Wilkinson[2] (1984); Chisholm[3] (1988)]. Thus the solution was to increase the physical marker opening by around √2 to 28μm. Following this change there was some improvement in the percentage of successfully auto-aligned jobs, but unfortunately, the majority of jobs were still failing.

The strangest aspect of this problem was that another user of topological markers was experiencing no significant alignment problems after some initial alterations of the video and contrast levels by Dr. Steve Thoms [Khamsepour[4] (1993-1996)]. The main difference between his markers and those used by the author was that his were 20μm etch pits in an otherwise flat substrate, *i.e.* no plateau. Thus the first possibility was that the auto-focusing mechanism of the beamwriter, based on a laser bouncing off the substrate, meant that the spot would be focused onto the base of the plateau, rather than on the marker itself. To overcome this, the beamwriter was deliberately defocused by the 20μm associated with the height of the plateaux. However, this did not improve the situation appreciably.

Finally, it was decided that some serious tests be carried out to eliminate process-specific possibilities. Two samples were fabricated together by the author, one with markers as used by the other user, recessed into the substrate, and one with markers as used by the author, located on 20μm tall plateaux. Two beamwriter jobs were constructed to force the beamwriter to align to all 16 markers on each sample. The results were dramatic. For the markers recessed into the substrate, there was a 100% success rate in finding the markers, while for the markers located on plateaux there was a 100% failure rate.

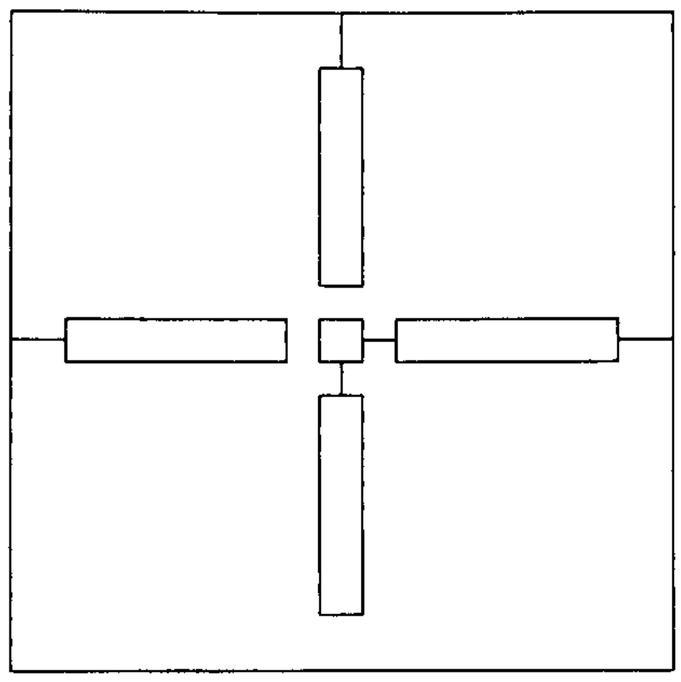


Figure H.4: The pattern used to produce crosses for coarse alignment by beamwriter operators. This design prevents any convex corners from being exposed other than those at the corners of the plateau. However, the degree of undercutting will not be severe enough to destroy the cross.

For this test the plateaux had been made 150μm across, with a single 20μm etch pit at the centre of each one. Altering this to 500μm did not improve the situation. However, upon one occasion when Dr. Steve Thoms was attempting to adjust some parameters for

a sample which suffered from the same problem, he succeeded in interrupting the beamwriter's marker search routine. [Normally the beamwriter delays responding to interrupts until after the routine has finished.] At this point, when he switched the beamwriter to SEM mode, he noticed that not all the secondary electron detectors were switched on. This was presumably part of the beamwriter's routine for finding the marker. This led to Dr. Steve Thoms discovering a parameter, 'detset', which allows the number and position of secondary electron detectors used in a marker search routine to be set. This parameter is normally zero, indicating that the beamwriter will use the default number of detectors to find the current marker. Then it is clear that for topological markers that number is not four. Dr. Steve Thoms then set this parameter to force the beamwriter to use all four detectors to find the marker. Surprisingly, this dramatically increased the success rate of finding topological markers. Dr. Steve Thoms also discovered that the detector signals were not all balanced correctly with respect to each other. After adjusting these levels and making 'detset = 1' (all detectors switched on) the default setting for topological markers, the auto-alignment success rate was close to 100%.

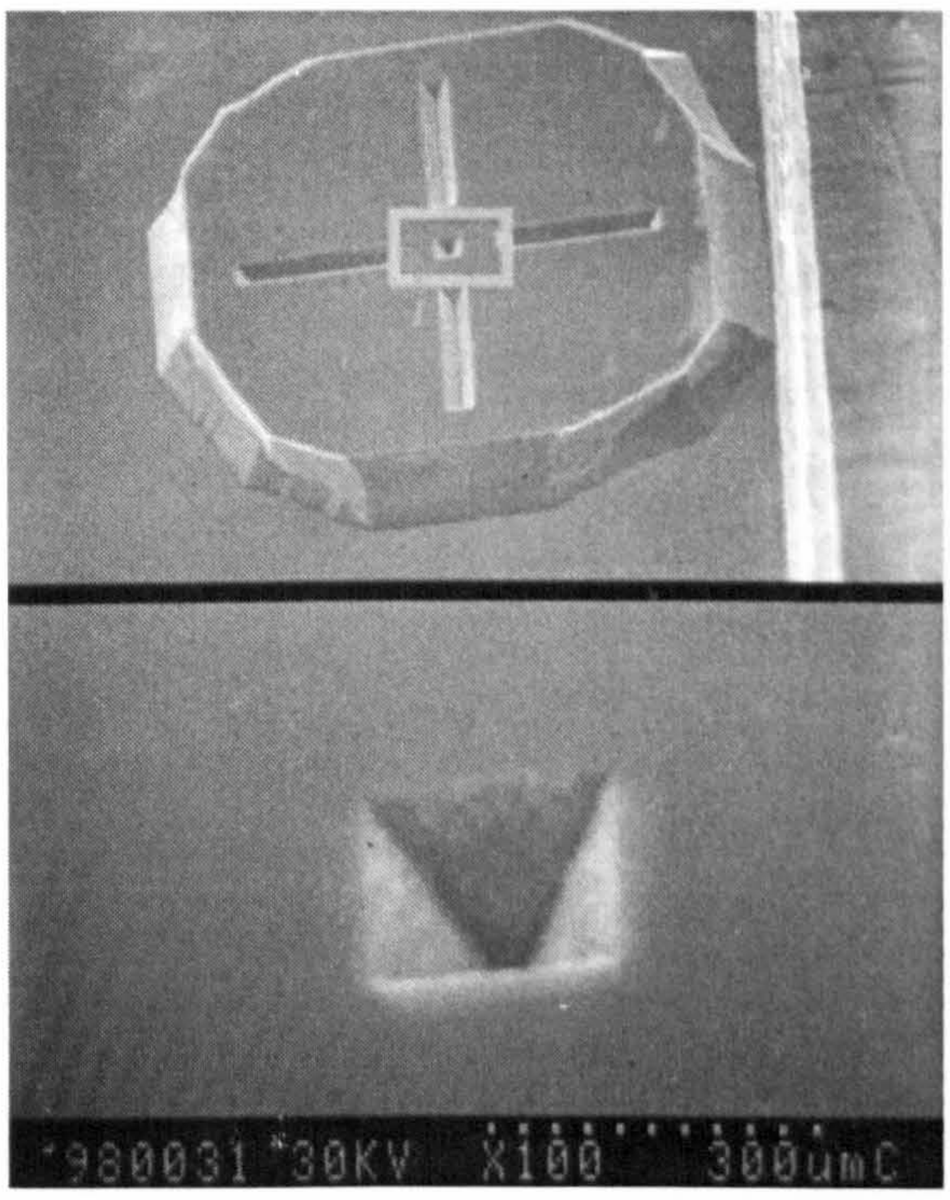


Figure H.5: Micrograph showing an etched cross structure used for coarse alignment by beamwriter operators. This helps to protect the first alignment marker from overexposure. The bottom half of the right image is a 10x magnification over the displayed scale which applies to the top half.

In addition, patterned crosses can be used to help limit the damage caused to the first bottom marker by being exposed to the electron-beam in SEM mode. This mode overexposes the resist on the first marker until it has become cross-linked to the point where the resist is difficult to remove. Fabrication steps following this can then damage the marker further. With the cross method, a set of crosses or distinctive marks are placed on the substrate. Since the distance from these structures to the first marker is

accurately known, the SEM mode only destroys these cross structures when the machine operator centres on them, rather than the first marker itself.

Due to the constraints imposed by silicon wet etching and convex corner undercutting, the crosses used consisted of a 'cross-like' arrangement of etched v-grooves combined with a topological marker, as illustrated in Figure H.4. After the tip definition etch, these crosses look as shown in Figure H.5. Twenty crosses were placed at the edge of each quarter wafer of devices.

H.3 REFERENCES

1. S. Thoms, *Private Communication*, (1993-1996).

2. S. A. Rishton, S. P. Beaumont and C. D. W. Wilkinson, *MEASUREMENT OF THE PROFILE OF FINELY FOCUSED ELECTRON BEAMS IN A SCANNING ELECTRON MICROSCOPE*, Journal of Physics E: Scientific Instrumentation **Vol. 17,** p. 296 (1984).

3. T. Chisholm, *SPOT-SIZE MEASUREMENT IN AN ELECTRON-BEAM PATTERN GENERATOR*, Journal Of Vacuum Science & Technology B **Vol. 6,** No. 6, pp. 2066-2069 (1988).

4. B. Khamsepour, *Private Communication*, (1993-1996).

APPENDIX J

TIP DEVELOPMENT

J.1 TIP PROFILE DEVELOPMENT

A pattern was designed for the exploration of possible solutions to the 'egg-timer' profile problem of §3.7.4, experienced when wet etching the tips [see Figure 3.7.8a]. This pattern consisted of 20μm, 30μm, 40μm and 50μm circles and squares with orientations along the projections of the {111}, {212}, {313/311} planes onto the (100) surface. The {212}, {313} and {311} planes have variously been identified as the fast etching corner undercutting planes in silicon [Lee[1] (1969); Declercq, Gerzberg and Meindl[2] (1975); Bean[3] (1978); Ju and Hesketh[4] (1992); Hesketh, *et al.*[5] (1993)]. Simple geometry shows that the projection of the {212} planes on the (100) surface are along the eight [012] axes which make an angle of 18.435° and 71.565° with the four [011] axes (the projection of the {111} planes onto the (100) surface) which are parallel to the flats of the wafer. Similarly the projection of the {313} and {311} planes on the (100) surface are along the eight [013] axes which make an angle of 26.565° and 63.435° with the four [011] axes. Thus the squares were aligned parallel with the substrate edges for the {111} case, rotated by 18.435° with respect the substrate edges for the {212} case and rotated by 26.565° with respect to the substrate edges for the {313/311} case.

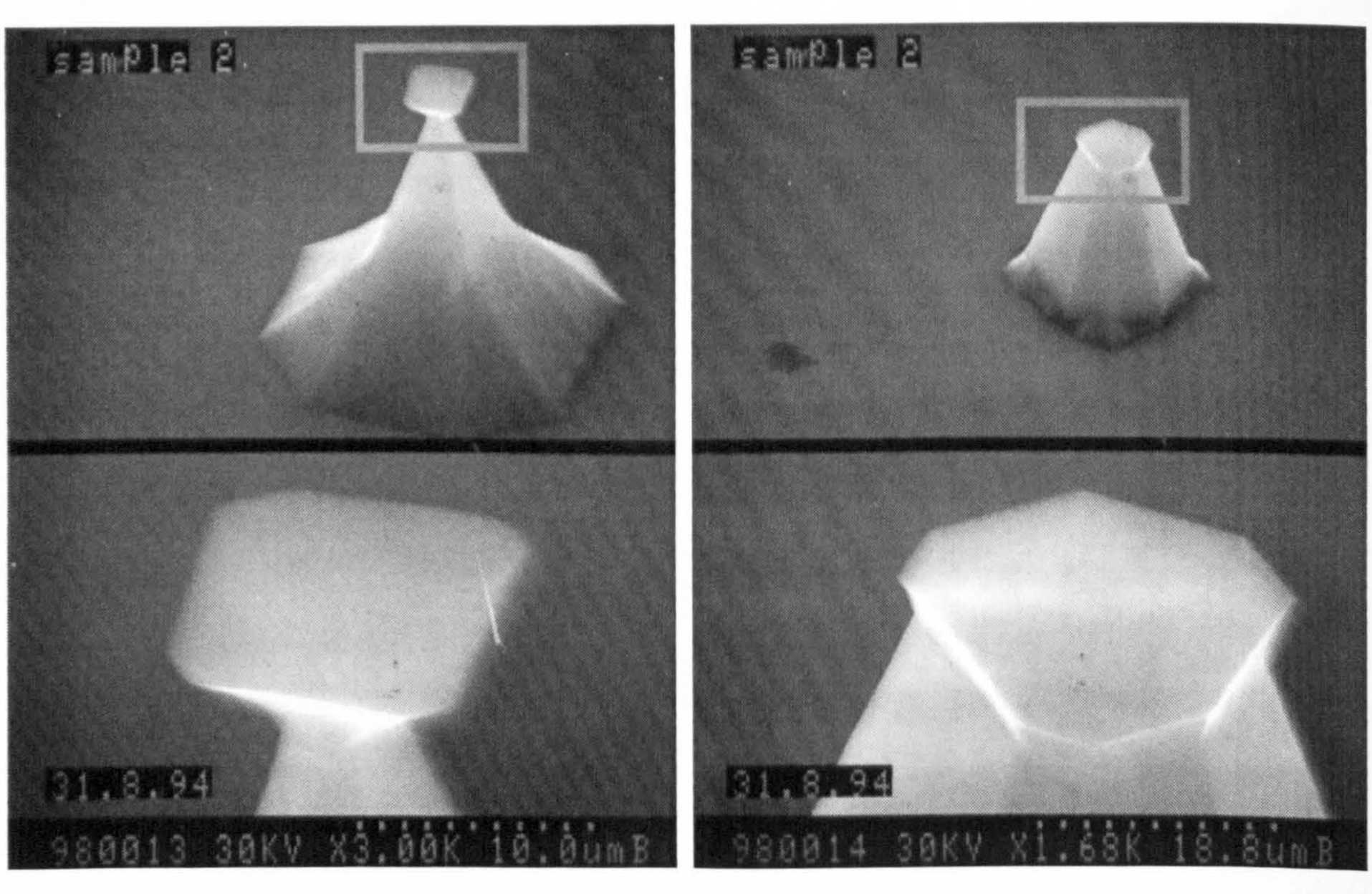


Figure J.1: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide. (a) Left Image: about half of the 40μm [011] oriented squares resulted in this structure; (b) Right Image: about half of the 40μm [011] oriented squares resulted in this structure. The bottom images are a 5x magnification over the displayed scales which apply to the top images.

These samples were prepared using the HRN technique described in Appendix E (§E.3) as this method was still working fine at this point. The pattern was written nine times on a whole 3 inch wafer (coated with 50nm of LPCVD silicon nitride) to make sure of the

alignment to the edges of the wafer. The silicon nitride was patterned using the orthophosphoric acid technique rather than dry etching (so features were smaller than intended due to undercutting – see Appendix E, §E.3).

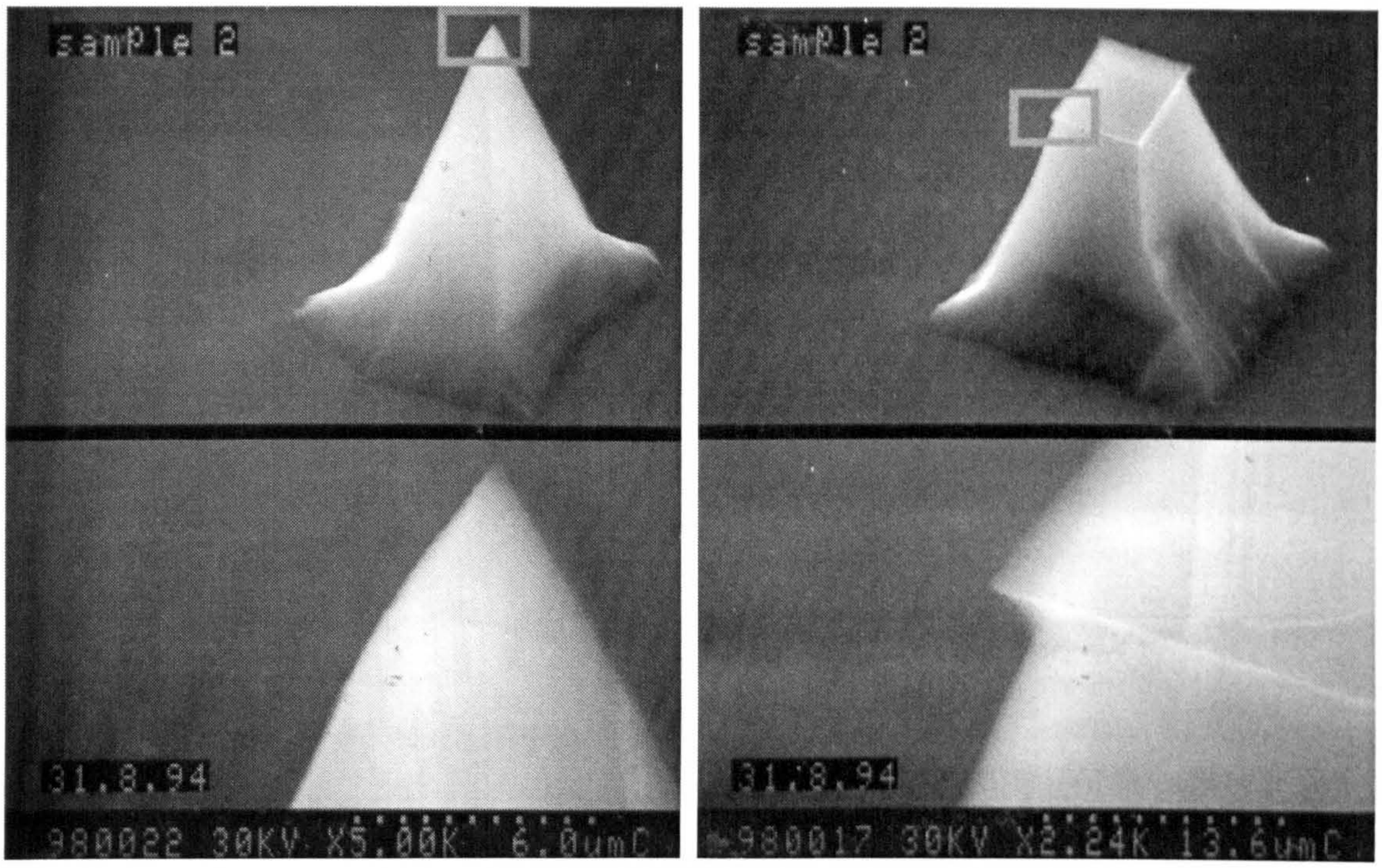


Figure J.2: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide. (a) Left Image: more than half of the 40μm [012] oriented squares resulted in this structure; (b) Right Image: less than half of the 40μm [012] oriented squares resulted in this structure. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

The first tests were carried out to evaluate the effects of stirring and wafer orientation in the etching solution. The etching solution was 7 molar (+5%) potassium hydroxide and etching was carried out for 10 minutes at 80±2°C (closed loop temperature control was not available at this time). Deoxidation was carried out with silicon dioxide etch 4:1 and after the tip definition etch the silicon nitride was stripped off using a long soak (90 minutes or so) in this deoxidation etch. Stirring and substrate orientation in the etching solution seemed to have no obvious impact on the resulting features. Some very bizarre looking structures were produced as shown in Figures J.1 through J.3. The etch depth in all cases was in the range of 10μm to 14μm. The results for the [013] oriented squares was similar to the [012] oriented squares. The variability of these tips was extremely undesirable and in particular, the change in etching to higher index planes partially through the etch, which only seemed to affect some tips, gave rise to structures which were much too steep for reliable electron-beam lithography. The apices of many tips were completely unsuitable and the variation in fracture point suggested that a consistent tip height across the sample would not be possible.

Thus an alternative etch had to be found. In the literature the addition of IPA to potassium hydroxide is known to change the etch selectivity and relative etch rates of

various planes [Bean[3] (1978); Lambrechts and Sansen[6] (1990-1993); Mastrangelo and Tang[7] (1994)].

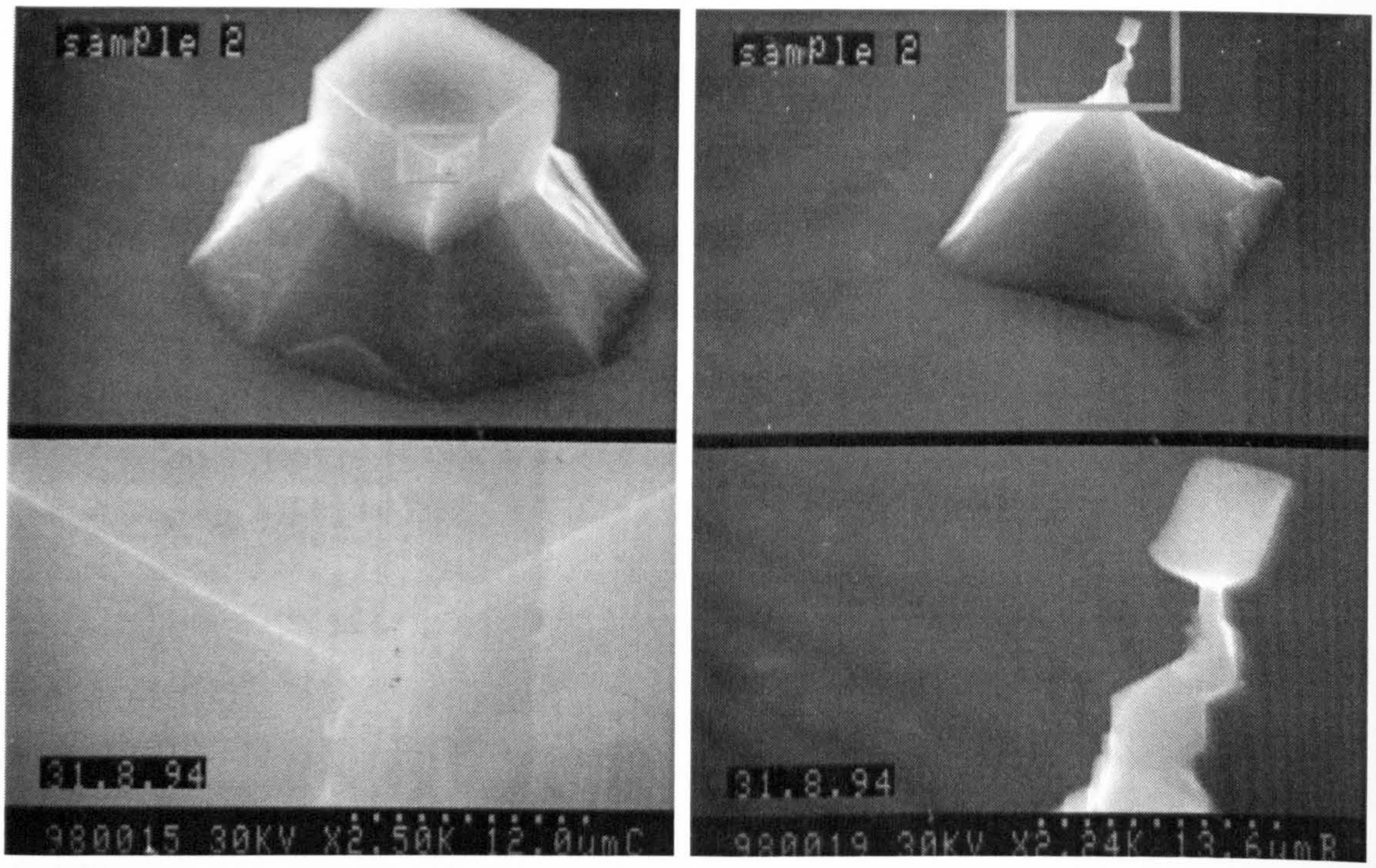


Figure J.3: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide. (a) Left Image: a few of the 40µm [011] oriented squares resulted in this structure. The bottom image is a 10x magnification over the displayed scale; (b) Right Image: a very few of the 40µm [012] oriented squares resulted in this structure. The bottom image is a 5x magnification over the displayed scale.

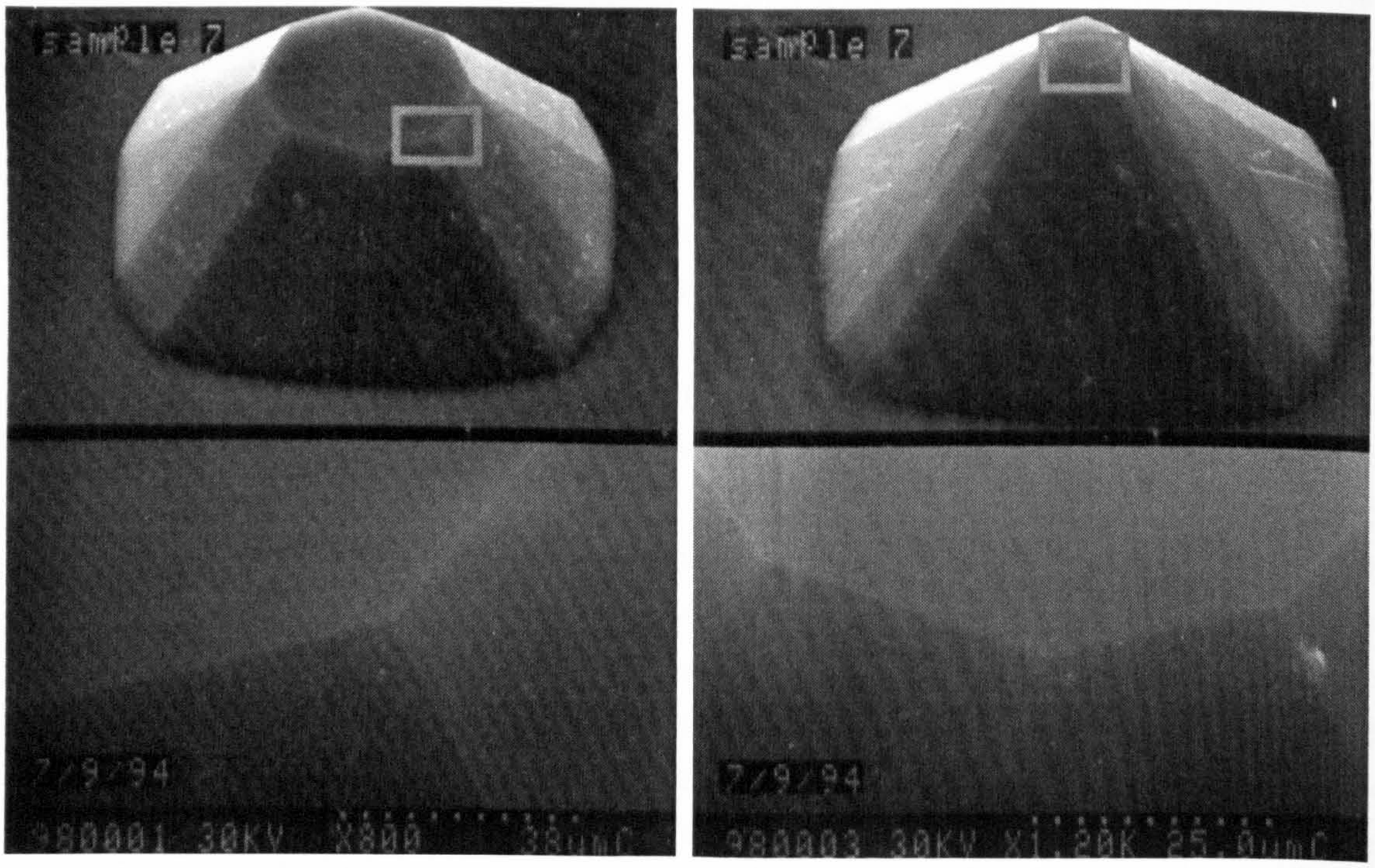


Figure J.4: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide and isopropanol. (a) Left Image: tips from 50µm [011] oriented squares; (b) Right Image: tips from 20µm [011] oriented squares. The bottom images are a 10x magnification over the displayed scales.

This was then investigated by adding 50ml of IPA to 250ml of the potassium hydroxide solution. [Note that as IPA is not fully miscible with water, some IPA remained as a separate layer above the potassium hydroxide solution.] This solution behaved in a very different manner as can be seen in Figures J.4 through J.7. Etching was carried out for 30 minutes at 80±2°C as before. This temperature seemed very close to the boiling point of this solution. The first difference is that the 'egg-timer' profile has completely disappeared. The second difference is that the rate of convex corner undercutting has decreased dramatically, as even at an etch depth of 30μm, the smallest pattern (20μm) has not been overetched till extinction of the plateau on the tip.

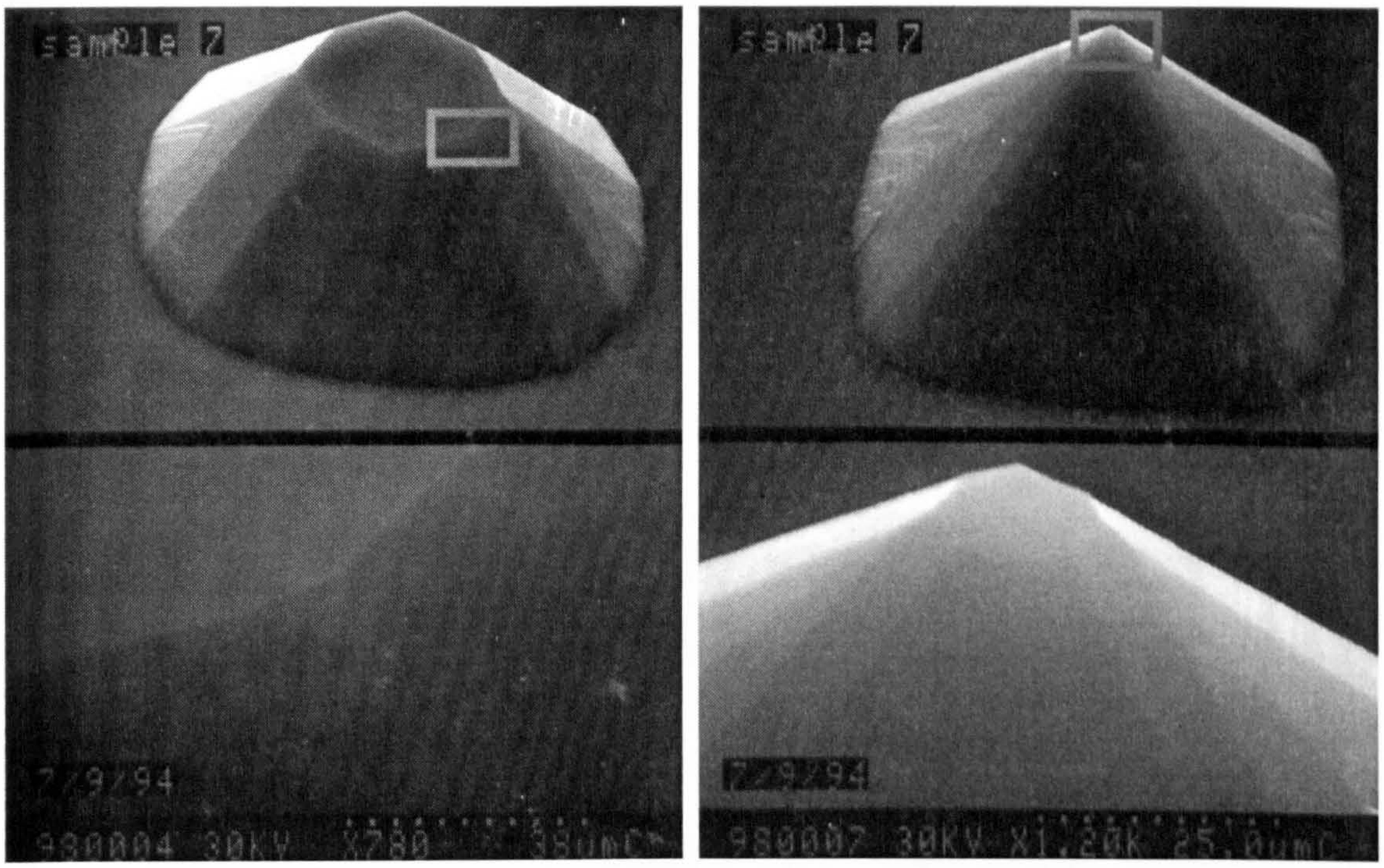


Figure J.5: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide and isopropanol. (a) Left Image: tips from 50μm circles; (b) Right Image: tips from 20μm circles. The bottom images are a 10x magnification over the displayed scales which apply to the top images.

Another observation is that the [011] oriented squares undercut the least, followed by the circles, then the [012] oriented squares and the fastest undercut occurs with the [013] oriented squares. This indicates that the {313} or {311} planes are the dominant convex corner undercutting planes in this etch. Both the circles and [011] oriented squares resulted in octagonal edged apices, presumably bounded by the [013] axes once the [011] projections had been completely eroded (in the case of the [011] oriented squares). By comparison, both the [012] and [013] oriented squares resulted in square shaped apices. In all cases, the undercut planes were the smoothest while the {111} planes were the most rough, however, it was clear that as the etch progressed, the size of the apex decreased along with the surface area occupied by these smooth planes. The roughness along the {111} planes were presumably due to edge roughness in the original masking layer in an analogue of the ledges seen when a concave structure is slightly misaligned with the [011] axes of a silicon wafer [Kendall[8] (1975)].

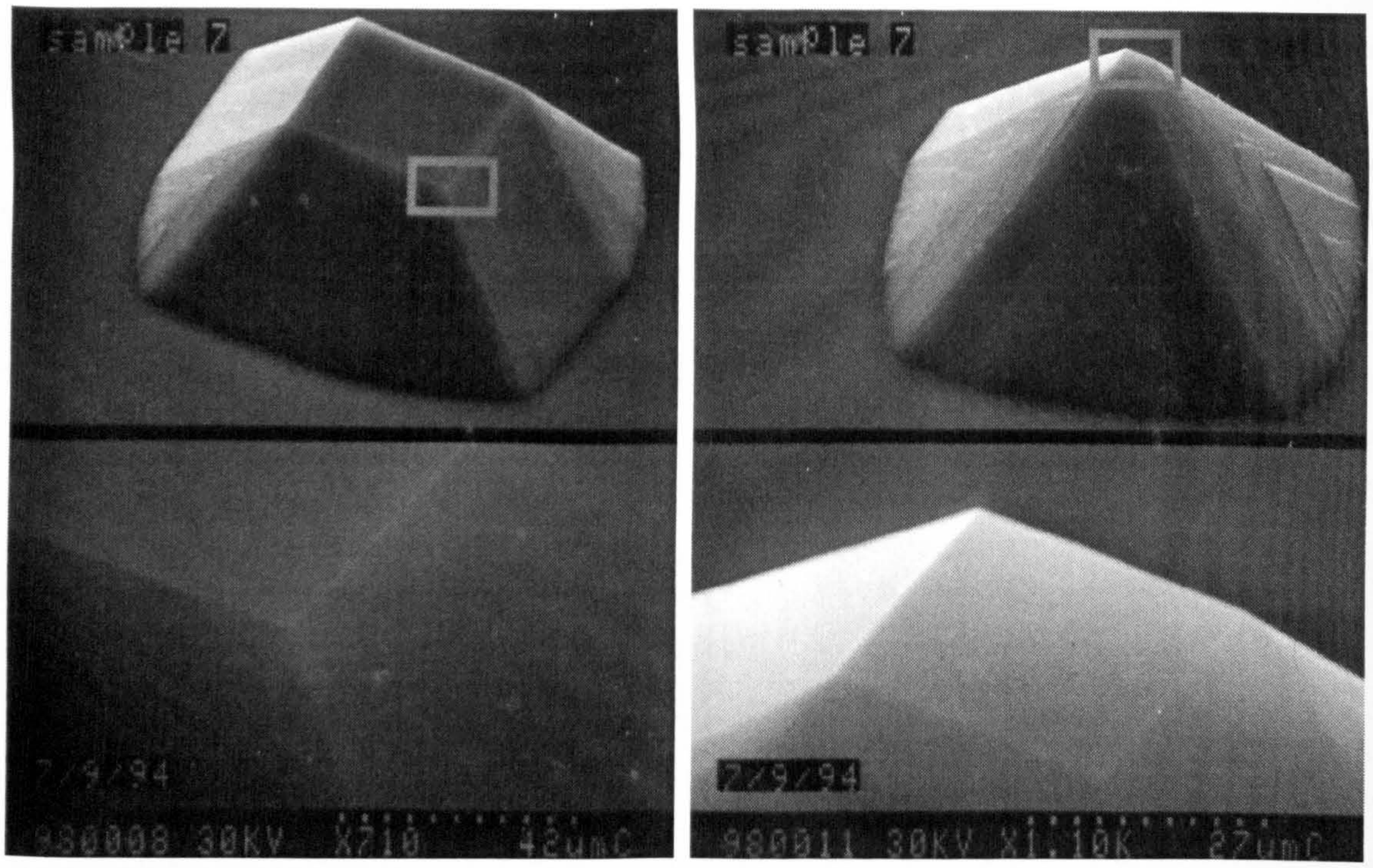


Figure J.6: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide and isopropanol. (a) Left Image: tips from 50μm [012] oriented squares; (b) Right Image: tips from 20μm [012] oriented squares. The bottom images are a 10x magnification over the displayed scales.

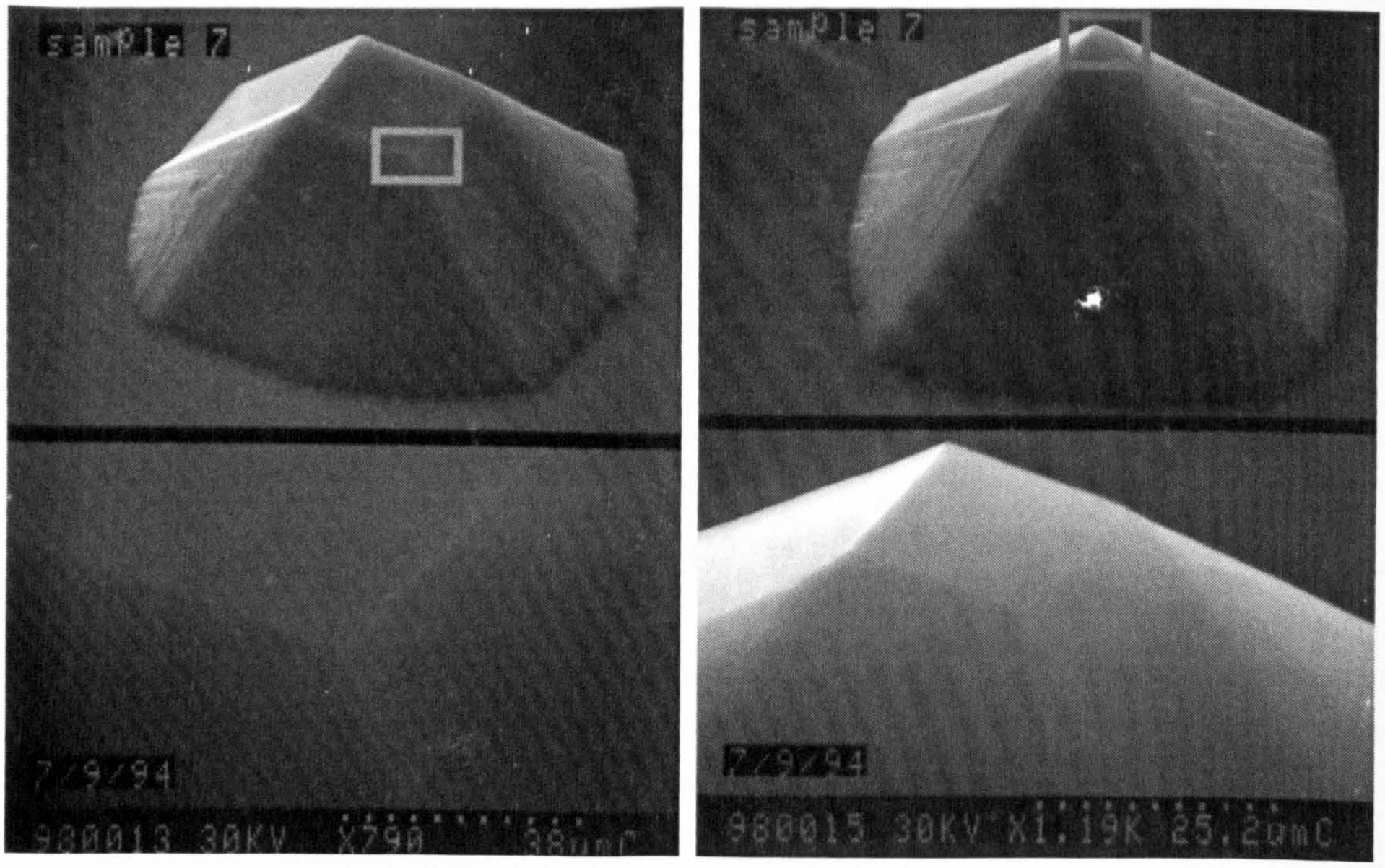


Figure J.7: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide and isopropanol. (a) Left Image: tips from 40μm [013] oriented squares; (b) Right Image: tips from 20μm [013] oriented squares. The bottom images are a 10x magnification over the displayed scales.

Since the undercutting of the resist by the orthophosphoric acid could not be relied upon to be uniform, the conclusion that the {313} or {311} planes were the fastest undercutting required further verification. To this end a sample was produced using a bi-layer of PMMA (8% BDH, 4% Elvacite both spun at 5000 rpm for 60 seconds – see Table 3.5.1) to

lift-off 100nm thick titanium as a dry etch mask for the silicon nitride. This was carried out on a whole 3 inch wafer to ensure orientational accuracy. The pattern consisted of squares oriented from 0° to 45° with respect to the flats of the wafer in 0.25° increments. This sample was wet etched for 15 minutes in the same potassium hydroxide-isopropanol solution to a depth of about 15μm and the silicon nitride stripped off with hydrofluoric acid as before. Figures J.8 and J.9 show the tips produced from masks oriented along selected angles. From this it was clear that the {313} or {311} planes were the fastest undercutting planes.

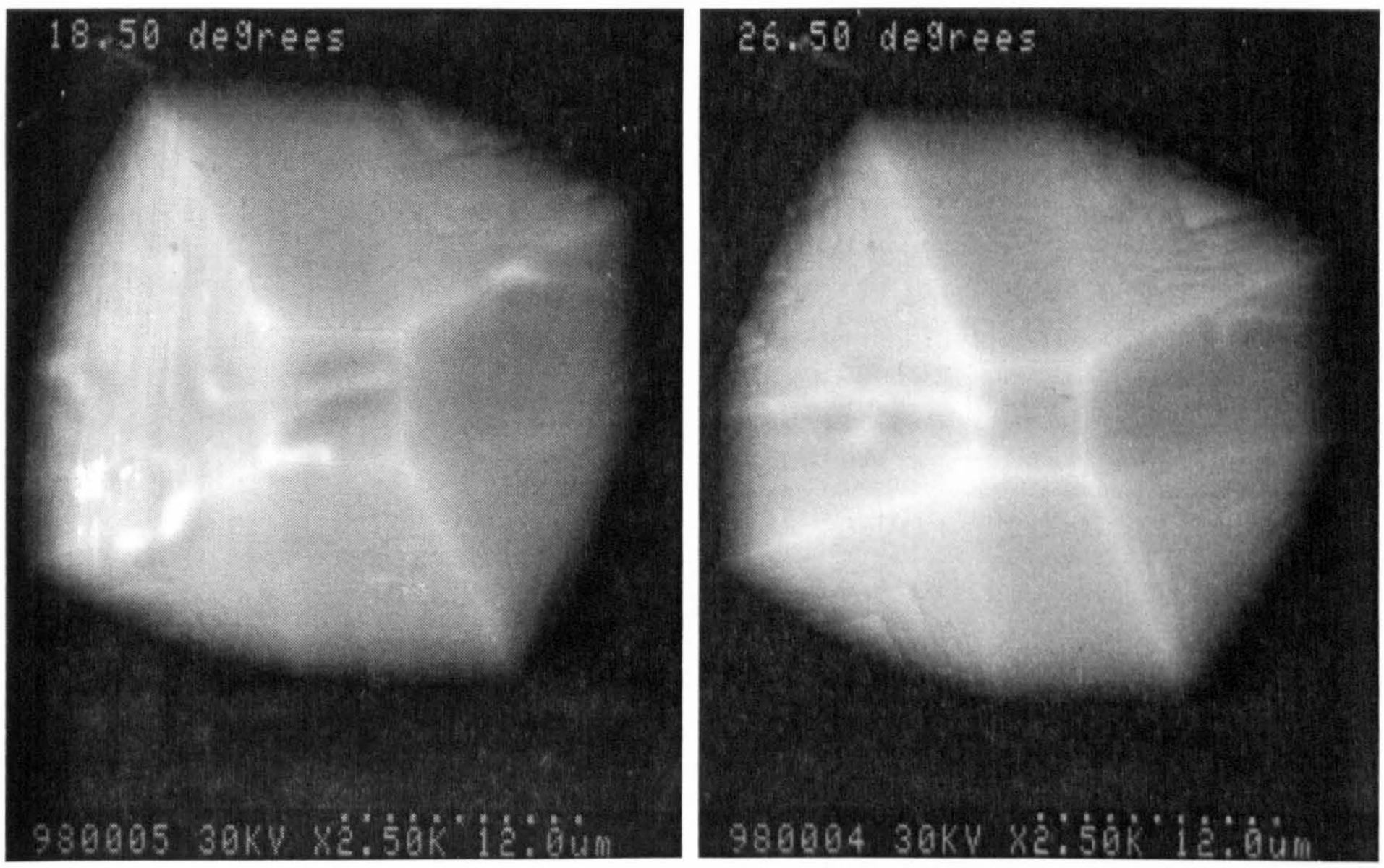


Figure J.8: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide and isopropanol. These tips were produced from square masks oriented at 18.5° (left image) and 26.5° (right image) to the [011] axes. This corresponds to a mask orientation approximately along the [012] and [013] axes respectively.

To decide between the {313} and {311} planes a simple calculation of the angle between the inclination of the pyramidal plane and the horizontal plane of the substrate was carried out. This corresponds to the angle between the undercutting planar direction and the [001] substrate direction. Then the approximate angle the {111} planes make with the substrate is 54.74°, the {212} planes make with the substrate is 48.19°, the {313} planes make with the substrate is 46.51° and the {311} planes make with the substrate is 72.45°. From Figure J.7 it is clear that the inclination of the undercutting planes is quite close to that of the {111} planes. This rules out the 72.45° inclined {311} planes and suggests that the 46.51° inclined {313} planes are the most likely candidate.

It is clear from Figure J.9b that aligning the mask with the [010] axes yields the structure most geometrically suitable for structures such as thermocouples, because the edges and smooth planes are all co-linear, however, because the {313} planes are the fastest undercutting planes, it is to be expected that the most reproducible pyramidal tips will be

those aligned with the [013] axes because the mask undercutting rate has a local maximum with respect to mask orientation along the [013] axes. This would result in the reduction of any effects due to misorientation of the mask with respect to the crystallographic axes. Thus it was decided that this mask orientation be used to produce all future pyramidal tips.

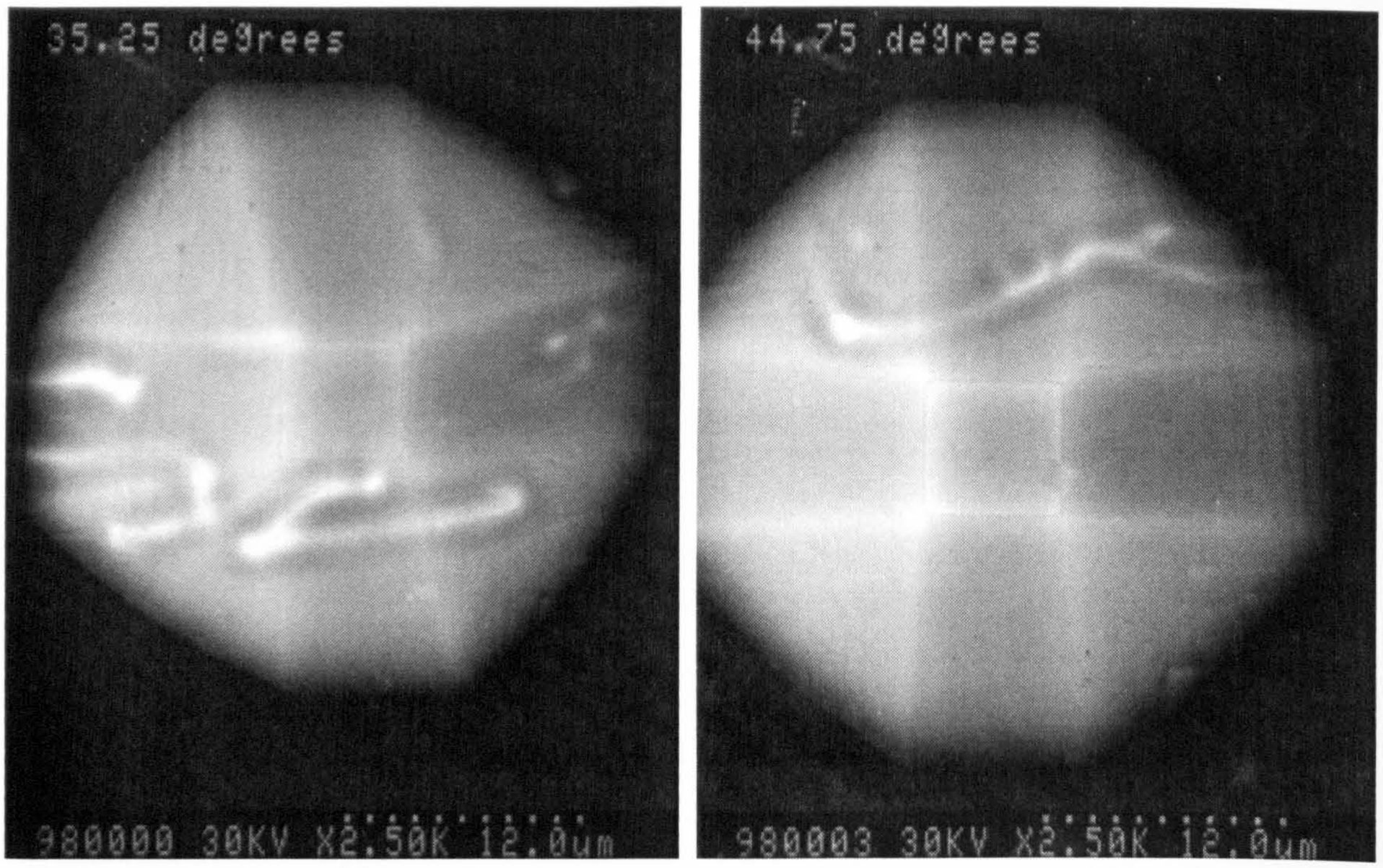


Figure J.9: Micrographs showing the mask orientation dependence of wet etching pyramidal tips with potassium hydroxide and isopropanol. These tips were produced from square masks oriented at 35.25° (left image) and 44.75° (right image) to the [011] axes. This corresponds to a mask orientation approximately along the [016] and [010] axes respectively.

However, an etch rate of 1μm per minute along the [100] directions was too fast for accurate control of the plateau size of the tips, due to the larger relative timing error in a 15 minute etch. Therefore it was decided that the etch temperature be reduced to 65°C in order to slow the etch rate down. So that whole 3 inch wafers could also be etched, it was decided that the etching solution be changed to 50ml of IPA added to 700ml of 7 molar (+5%) potassium hydroxide solution. It was expected that although the apparent relative concentration of IPA would change dramatically by doing this, the effect on the etching solution would be minimal because, as noted earlier, the previous solution appeared to be saturated with IPA to the point where a layer of IPA formed above the etching solution. This etch then gave a {100} etch rate of about 22μm per hour and a 12μm square mask oriented along the [013] directions required about 48 minutes of etching to yield a tip with a sub-micron plateau size. The tip height was about 18μm.

To control the plateau size, it was found that an etch rate calibration was necessary. The calibration sample simply consisted of [013] oriented squares of sizes from 5μm to 25μm in 0.25μm increments. This sample would be etched, typically for 30 minutes, immediately before the real sample and then used to calibrate the undercutting rate so

that a more accurate etch time could be estimated. However, it was found that using test samples which had been written on the corner pieces left over from 3 inch wafers was not accurate enough, as illustrated by Figure J.10 so that these test pieces had to be written on a whole 3 inch wafer.

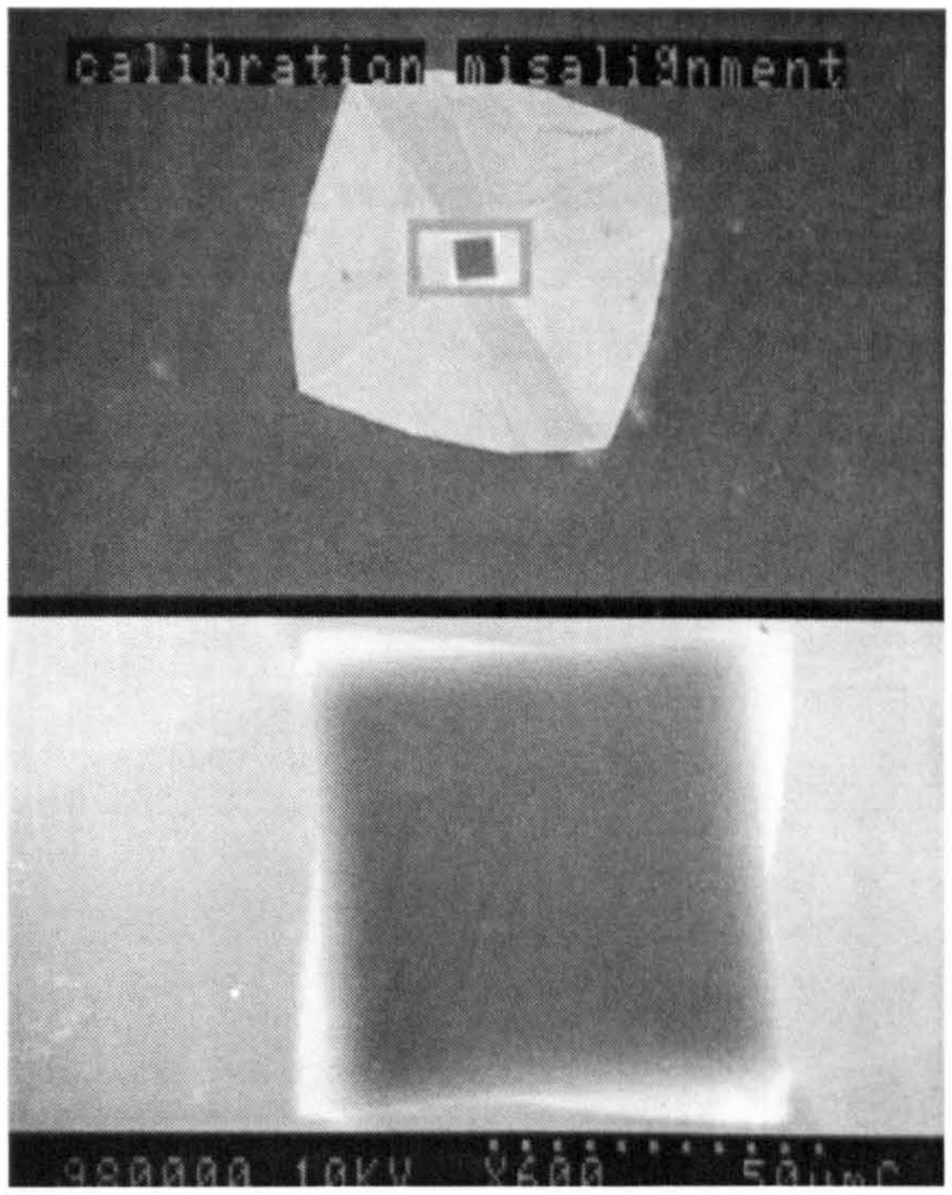


Figure J.10: Micrograph showing misorientation of the mask used for calibration of the pyramidal tip wet etch. The bottom half is a 10x magnification over the displayed scale which applies to the top half.

J.2 TIP AND SURFACE QUALITY IMPROVEMENT

One problem which arose using the etch of §J.1 was the variability in the etching characteristics of this solution. Although it appeared that the solution was saturated with IPA, it was noticed that the rate of undercutting would vary from etch to etch. At first this was attributed to changes in the IPA concentration due to evaporation. It was found that without any IPA, the {100} etch rate was about 28μm/hour, so a new scheme for correcting the quantity of IPA by measuring the {100} etch rate was devised based on a speculative linear decrease in etch rate from 28μm/hour to 22μm/hour with the addition of IPA (to 1 part in 14). However, this did not improve the situation. This coupled with erratic roughness problems where the pyramidal tips would sometimes be surrounded by myriads of miniature tips, or 'hillocks', between 5μm and 15μm high, meant that further development of this process was required. Hillocks as high as 10μm have been known to form using potassium hydroxide solutions less than 30% concentrated [Mastrangelo, *et al.*[7] (1994)] but as the current etching solution was almost 30% concentrated and these hillocks were not seen when etching with the same solution without any IPA, the cause was not obvious. However, it had also been noted by Lambrechts and Sansen that the IPA-potassium hydroxide etching solution results in 'hillock formation' during the first

few hours of etching [Lambrechts, *et al.*[6] (1990-1993)]. Unfortunately no reason or solution had been given.

As a specific potassium hydroxide/IPA combination had been mentioned in one particular reference [Mastrangelo, *et al.*[7] (1994)], it was decided that this etch be evaluated in the hope that it was more stable and consistent. This etch consisted of 20% by weight potassium hydroxide solution mixed with 250ml of IPA per litre. Initial attempts using samples with mask square sizes ranging from 10μm to 18μm showed that it was no better. It was thus decided that the temporal evolution of the tip in the etch be investigated.

It was then that it was discovered that samples which had been etched for a short time, around 5 minutes, seemed to be coated in a white material upon removal. Upon examination with an SEM it was concluded that some sort of polymer was forming on the surface of the sample. This polymer seemed to act as a mask for a short period of time, before being eroded away, and resulted in excessive sample roughness. The origin of the polymer was not known, so a series of experiments were carried out to isolate the polymer constituents from those used prior to, during and after the etch.

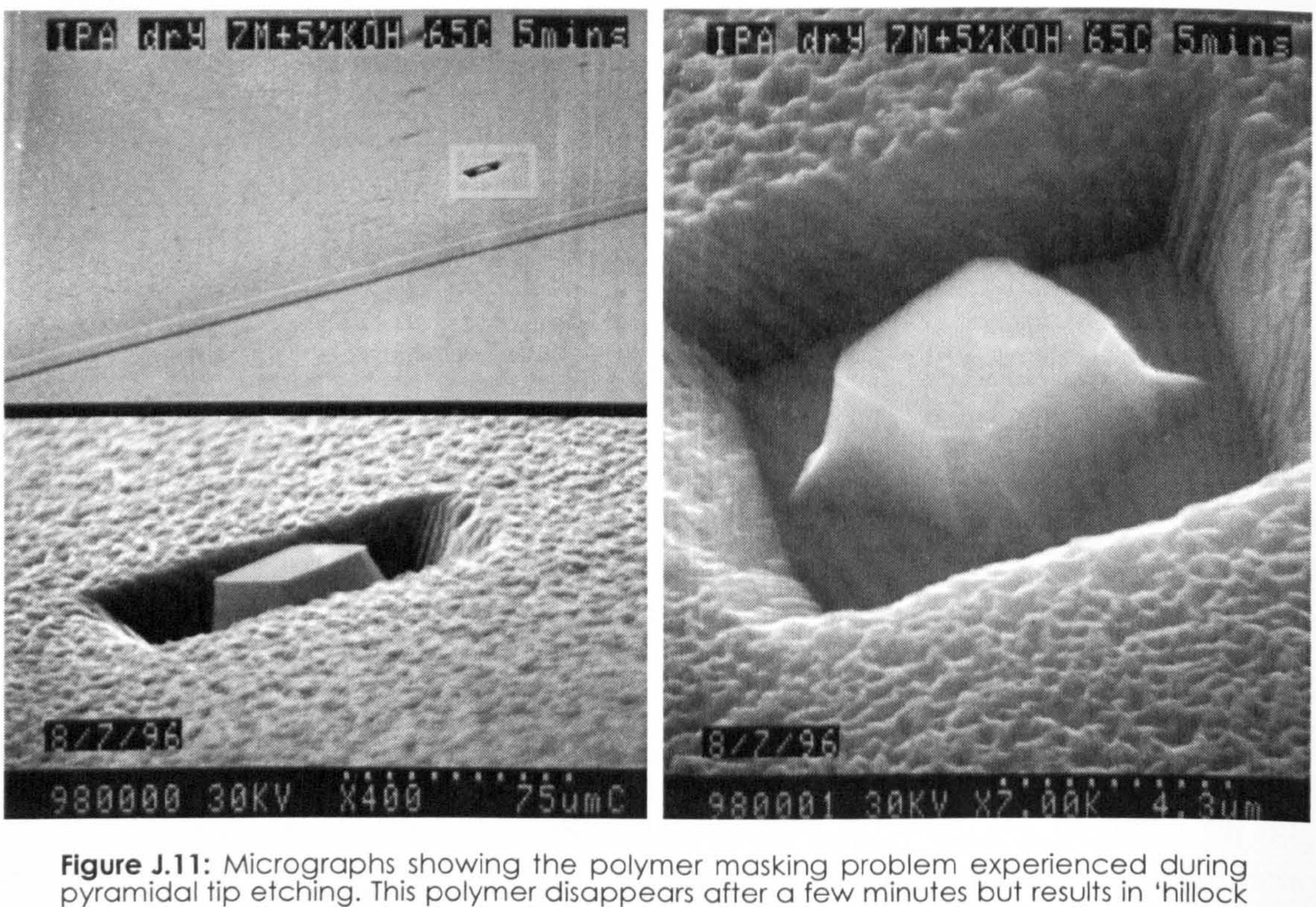


Figure J.11: Micrographs showing the polymer masking problem experienced during pyramidal tip etching. This polymer disappears after a few minutes but results in 'hillock formation' on the sample. The bottom half of the left image is a 10x magnification over the displayed scale which applies to the top half.

These samples had been treated as follows:

(a) Nanostrip clean.

(b) One minute deoxidation in silicon dioxide etch 4:1.

(c) 5 minute etch in 20% potassium hydroxide:IPA 3:1.

(d) 1 minute neutralisation in sulphuric acid:water 1:4

(e) 5 minute rinse in ultra-pure water

(f) 90 minute silicon nitride strip in silicon dioxide etch 4:1.

It was found that the same roughness could be created in potassium hydroxide solution without any IPA, provided the sample was rinsed in IPA and blown dry immediately before etching. Until this point all the potassium hydroxide without IPA etches had produced surfaces with a near-mirror finish. Figure J.11 shows the effect of the polymer in masking the etch for a short time on the silicon surface. The cause was then limited to a combination of silicon dioxide etch, IPA and potassium hydroxide solution. When the silicon dioxide etch was substituted with unbuffered 48% hydrofluoric acid, all signs of the polymer immediately disappeared, as illustrated in Figure J.12.

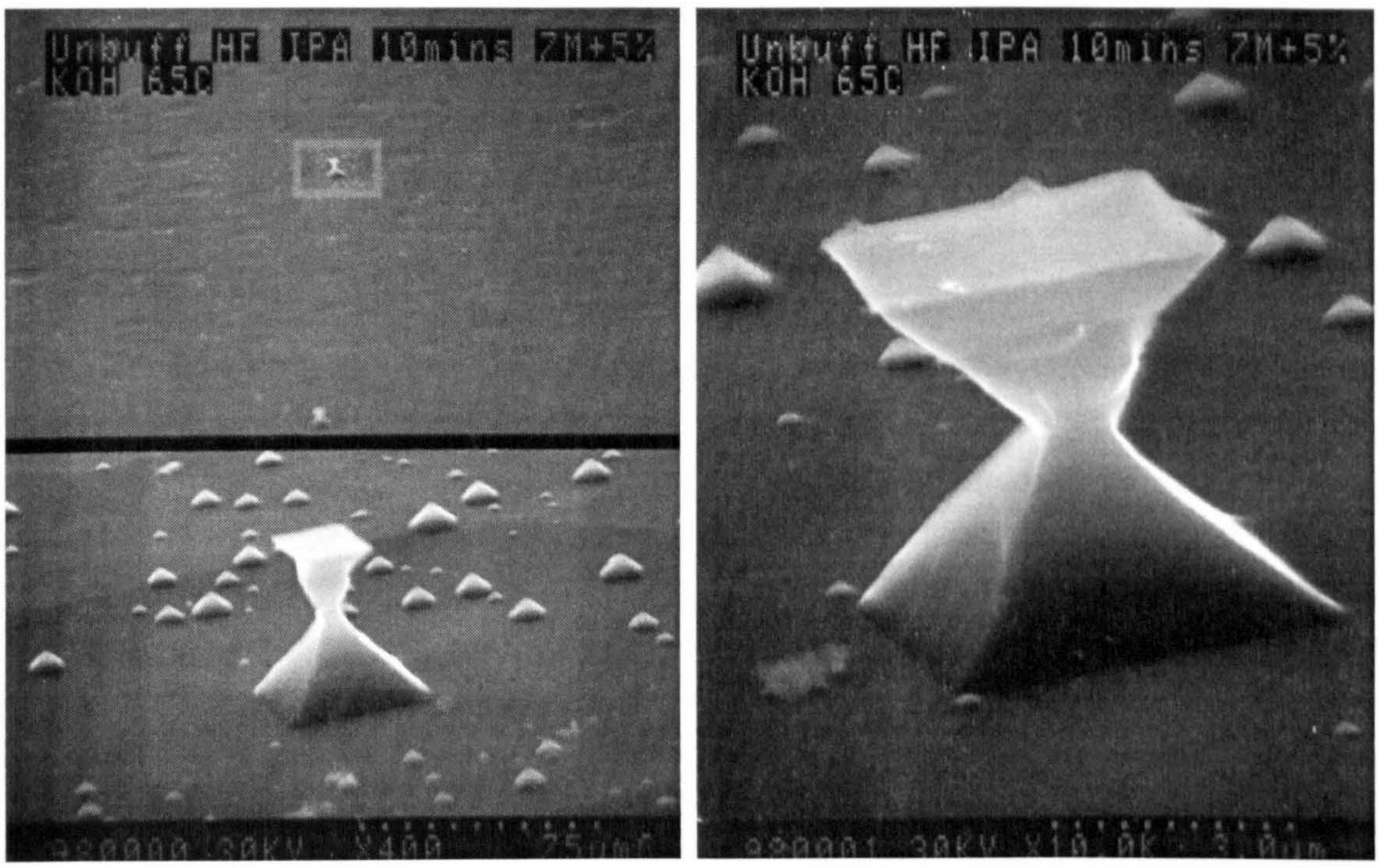


Figure J.12: Micrographs showing the elimination of the polymer masking problem experienced during pyramidal tip etching. The etch here is potassium hydroxide without IPA. Although there are spurious pyramids, the polymer layer of Figure J.11 has disappeared. The bottom half of the left image is a 10x magnification over the displayed scale which applies to the top half.

Eventually it was discovered that the polymer could be produced on a bare silicon surface by simply deoxidising it for one minute in silicon dioxide etch, rinsing it in water, dipping it into IPA and blow drying. The polymer formed only on the silicon surface and its quality could be improved by preceding the deoxidation with a clean in nanostrip. It was also found that the polymer could be produced on glass slides in the same manner (although the nanostrip clean seemed to be necessary for an adequate film). It was speculated that the buffering agent in the deoxidation etch, ammonia, combined with hydrofluoric acid, in some way left silicon in a condition which was favourable for the polymerisation of IPA. Whilst it is known that hydrofluoric acid leaves a bare silicon surface in a hydrophobic state, there was insufficient time to investigate the interaction of this surface with ammonia and IPA. However, it would appear that the use of silicon

dioxide etch may explain the 'hillock formation' during the early stages of etching using potassium hydroxide with IPA as observed by Lambrechts, *et al.*[6] (1990-1993).

At this point it was decided that unbuffered hydrofluoric acid:water 1:4 be used exclusively for deoxidising silicon prior to all etching steps. In addition, it had been noted that during the stripping of silicon nitride using silicon dioxide etch, a brown coloured deposit would occasionally appear on the silicon surface which could not be removed in any concentrated acid or by cleaning. This deposit was found to virtually disappear when unbuffered hydrofluoric acid was used in place of silicon dioxide etch for stripping the silicon nitride. The resistance of this deposit to both hydrofluoric acid and the acid clean step of the inorganic clean leads the author to speculate that this deposit may be porous silicon. A brownish film, which has been identified as porous silicon, is known to form during the electrochemical etching of silicon with low current densities, or with hydroxide ion starved solutions such as concentrated hydrofluoric acid [Petersen[9] (1982)] However, it is not clear what the mechanism of its formation in the above process is.

An additional investigation was also carried out to determine how the etch changed when the sample was immersed in the solution above and below the potassium hydroxide/IPA interface (the IPA rich region was above the interface, the potassium hydroxide rich region was below the interface). At 65°C it was found that the {100} etch rates were 13μm/hour and 18μm/hour for the regions above and below the interface respectively, while the projection of the {313} etch rate onto the (100) surface (*i.e.* the mask undercutting rate per edge) was 6.5μm/hour and 4.75μm/hour for the regions above and below the interface respectively. Since this {100} etch rate was perceived as being too slow, the same tests were repeated at 70°C with the following results: {100} etch rates of 19μm/hour and 28μm/hour, [013] mask undercutting rates on the (100) surface of 6.25μm/hour and 6μm/hour for the regions above and below the interface respectively.

J.3 REFERENCES

1. D. B. Lee, *ANISOTROPIC ETCHING OF SILICON*, Journal Of Applied Physics **Vol. 40,** No. 11, pp. 4569-4574 (1969).

2. M. J. Declercq, L. Gerzberg and J. D. Meindl, *OPTIMIZATION OF THE HYDRAZINE-WATER SOLUTION FOR ANISOTROPIC ETCHING OF SILICON IN INTEGRATED CIRCUIT TECHNOLOGY*, Journal of the Electrochemical Society **Vol. 122,** No. 4, pp. 545-552 (1975).

3. K. E. Bean, *ANISOTROPIC ETCHING OF SILICON*, IEEE Transactions On Electron Devices **Vol. ED-25,** No. 10, pp. 1185-1193 (1978).

4. C. Ju and P. J. Hesketh, *MEASURMENTS OF THE ANISOTROPIC ETCHING OF A SINGLE-CRYSTAL SILICON SPHERE IN AQUEOUS CESIUM HYDROXIDE*, Sensors and Actuators A-Physical **Vol. 33,** pp. 191-196 (1992).

5. P. J. Hesketh, C. Ju, S. Gowda, E. Zanoria and S. Danyluk, *SURFACE FREE ENERGY MODEL OF SILICON ANISOTROPIC ETCHING*, Journal of the Electrochemical Society **Vol. 140,** No. 4, pp. 1080-1084 (1993).

6. M. Lambrechts and W. Sansen, *Chapter 4: PLANAR TECHNOLOGIES FOR MICROELECTROCHEMICAL SENSORS* in Biosensors: Microelectrochemical Devices, published by Institute of Physics, Leuven, p. 98 (1990-1993).

7. C. H. Mastrangelo and W. C. Tang, *Chapter 2: SEMICONDUCTOR SENSOR TECHNOLOGIES* in Semiconductor Sensors, edited by S. M. Sze, published by John Wiley & Sons, New York, pp. 17-95 (1994).

8. D. L. Kendall, *ON ETCHING VERY NARROW GROOVES IN SILICON*, Applied Physics Letters **Vol. 26,** No. 4, pp. 195-198 (1975).

9. K. E. Petersen, *SILICON AS A MECHANICAL MATERIAL*, Proceedings of the IEEE **Vol. 70,** No. 5, pp. 421-457 (1982).

APPENDIX K

THE COMPLIANCE OF SELECTED CANTILEVER GEOMETRIES

K.1 DERIVATION OF COMPLIANCE EQUATIONS

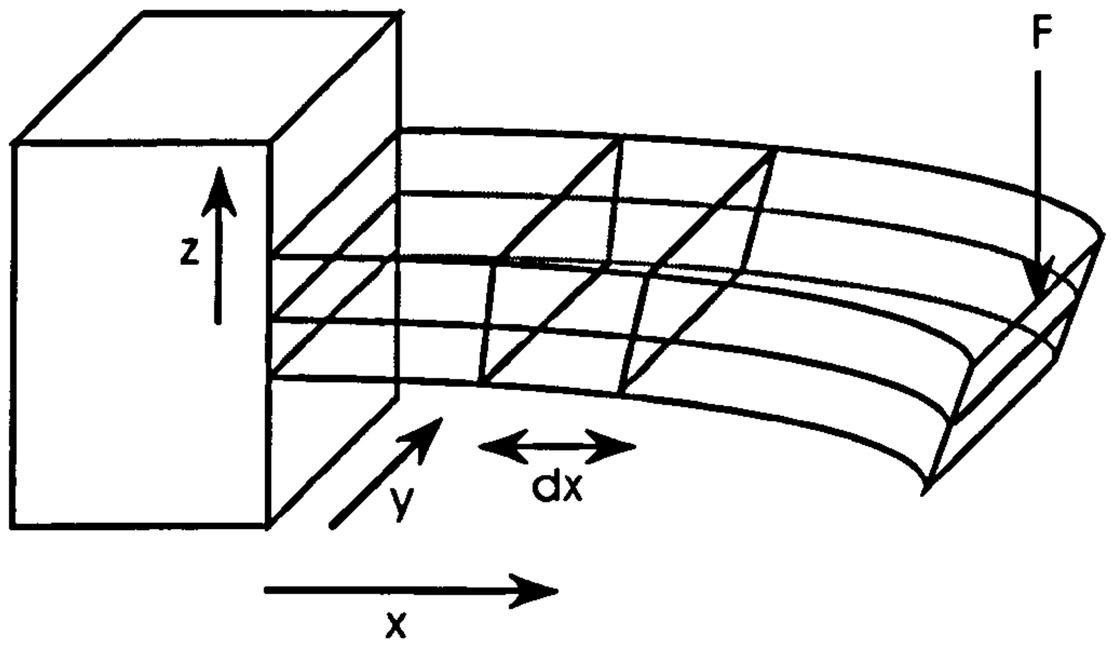


Figure K.1: Diagram of a beam being bent by the application of a force.

A means of relating the cantilever stiffness to its dimensions and material properties is required. Only a simple derivation will be given here, valid for small deflections such that no cross sectional area of the cantilever makes an angle, θ, with the undeflected position large enough for the approximations:

$$\cos\theta \sim 1, \tan\theta \sim \sin\theta \sim \theta \tag{K.1}$$

to be invalid.

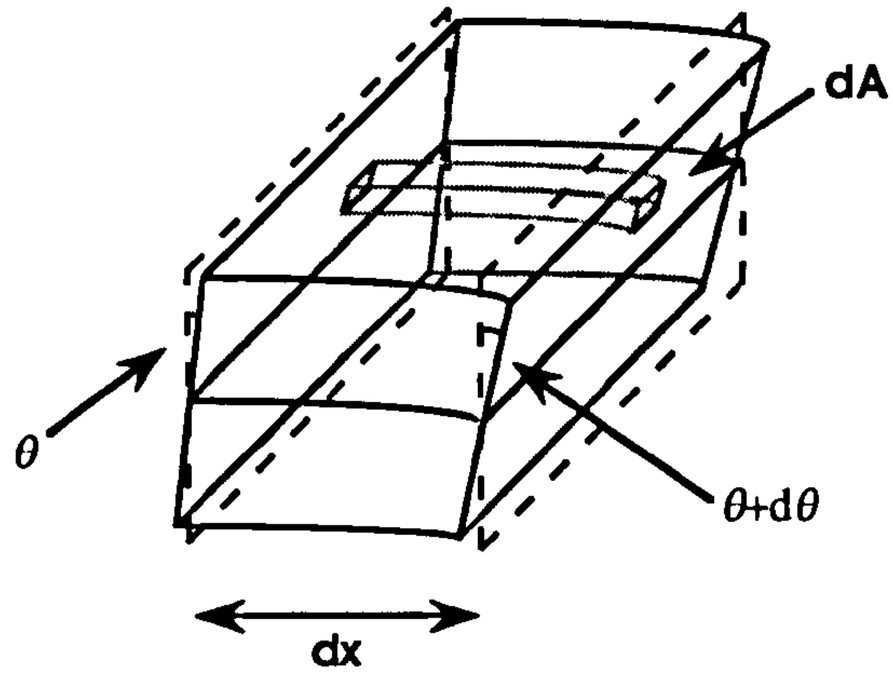


Figure K.2: Diagram of a beam section experiencing a bending moment.

Figure K.1 shows a beam being deflected by the application of a force. Figure K.2 shows a close-up of a small section of the beam. The external force results in parts of the beam on one side of the unperturbed axis being compressed, while parts on the other side are expanded. In equilibrium the moment applied by the external force about the unperturbed axis of the beam at its anchor point is balanced by the restoring moments generated by each segment of the beam.

Consider a small element of length, dx, and area, dA, of the beam section at position x from the anchor plane of the beam. Let this element be located at a position z from the unperturbed axis and let the vertical position of the unperturbed axis with respect to the

undeflected beam position be given by $z_c(x)$. Then it is clear that this element has been stretched by an amount dl given by:

$$dl = z\sin(\theta + d\theta) - z\sin\theta = z\sin\theta\cos d\theta + z\cos\theta\sin d\theta - z\sin\theta. \tag{K.2}$$

Using the approximations of Equation K.1, this becomes:

$$dl \approx z\sin d\theta \approx zd\theta. \tag{K.3}$$

Now the length of this element can be found by noting that:

$$\cos d\theta = \frac{dx}{l} \Rightarrow l \approx dx, \tag{K.4}$$

so that the strain on this element is given by:

$$\text{strain} = \frac{dl}{l} \approx z\frac{d\theta}{dx}. \tag{K.5}$$

Then the force, ΔF, being exerted by this element can be found from the strain by making use of the definition of Young's modulus, Y:

$$Y = \frac{\text{stress}}{\text{strain}} = \frac{\text{Force}/\text{Area}}{\text{strain}}, \tag{K.6}$$

so that:

$$\Delta F = \frac{YdldA}{l} \approx zYdA\frac{d\theta}{dx}. \tag{K.7}$$

It is also clear from elementary geometry that:

$$\tan\theta = \frac{dz_c}{dx} \Rightarrow \theta \approx \frac{dz_c}{dx}, \tag{K.8}$$

so that Equation K.7 becomes:

$$\Delta F \approx zYdA\frac{d^2z_c}{dx^2}. \tag{K.9}$$

This means that the moment, $\Delta\Gamma$, generated by this element about the unperturbed axis is given by:

$$\Delta\Gamma = z\Delta F \approx z^2YdA\frac{d^2z_c}{dx^2}, \tag{K.10}$$

so that the net moment contribution of the infinitesimal beam section at position x is given by:

$$\Gamma(x) = \int_{Area}\Delta\Gamma(x)dA_x \approx Y\frac{d^2z_c}{dx^2}\int_{Area}z^2dA_x. \tag{K.11}$$

Since the geometric area moment of inertia, $I(x)$, is defined by:

$$I(x) = \int_{Area}z^2dA_x, \tag{K.12}$$

Equation K.11 becomes:

$$\Gamma(x) = YI(x)\frac{d^2z_c}{dx^2}. \tag{K.13}$$

In the AFM situation, when the system is in equilibrium, it is clear that the net moment at the free end of the cantilever must be zero and the net moment at the anchor point of the cantilever must be given by:

$$\Gamma(0) = -F_{ext} l\,, \tag{K.14}$$

where l is the length of the cantilever. For a concentrated force at the end of a beam, the net moment as a function of the position along the beam is then given by:

$$\Gamma(x) = -F_{ext}(l - x), \tag{K.15}$$

so that:

$$-F_{ext}(l - x) = YI(x)\frac{d^2 z_c}{dx^2} \Rightarrow \frac{d^2 z_c}{dx^2} = -\frac{F_{ext}(l - x)}{YI(x)}. \tag{K.16}$$

From this, $z_c(x)$ can be determined and the compliance of the cantilever, k_c, is then given by:

$$k_c = -\left.\frac{dF_{ext}}{dz_c}\right|_{x=l}. \tag{K.17}$$

For simplicity only two cantilever cross-sectional geometries will be considered here. The first is that of a uniform rectangle. From symmetry it is clear that the unperturbed axis will lie along the central axis of the rectangle. For a rectangle of width, b, and thickness, t, the area moment of inertia is given by:

$$I_{rect}(x) = \frac{b(x)t^3(x)}{12}. \tag{K.18}$$

For a circle of radius, r, the area moment of inertia is given by:

$$I_{circle}(x) = \frac{\pi}{4} r^4(x). \tag{K.19}$$

Now the deflection for cantilevers of simple geometries can be calculated. The first geometry is illustrated in Figure K.3. This cantilever has a rectangular cross-section of uniform thickness, t. It consists of two sections: the first section is of width, b_1, and length, l_1; the second section of length, l_2, tapers uniformly from width, b_1, to width, b_2. It is useful to define the following quantities:

$$l_t = l_1 + l_2,\ b_f = \frac{b_1}{b_2}\ \lambda = \frac{b_1 - b_2}{l_2},\ \mu = b_1 + \lambda l_1. \tag{K.20}$$

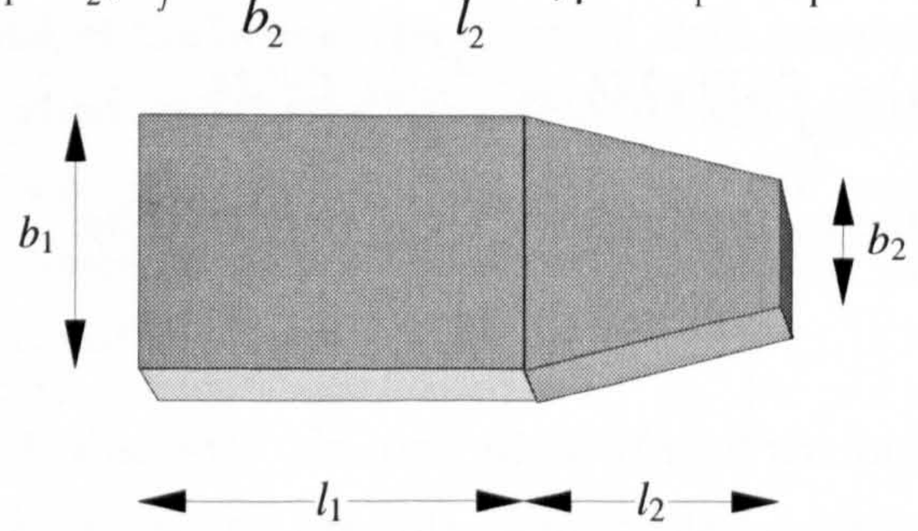


Figure K.3: Diagram of a simple cantilever beam with rectangular cross-section.

Then it is clear that:

$$b(x) = \begin{cases} b_1, & 0 \le x \le l_1 \\ -\lambda x + \mu, & l_1 \le x \le l_t \end{cases}, \quad t(x) = t. \tag{K.21}$$

From Equations K.16 and K.18:

$$\frac{d^2 z_c}{dx^2} = -\frac{12 F_{ext}}{Y t^3} \frac{(l_t - x)}{b(x)}, \tag{K.22}$$

which becomes:

$$\frac{d^2 z_c}{dx^2} = -\frac{12 F_{ext}}{Y t^3} \begin{cases} \dfrac{(l_t - x)}{b_1}, & 0 \le x \le l_1 \\ \dfrac{(l_t - x)}{-\lambda x + \mu}, & l_1 \le x \le l_t \end{cases}, \tag{K.23}$$

so that:

$$\left.\frac{dz_c}{dx}\right|_{x \le l_1} = -\frac{12 F_{ext}}{Y t^3} \int_0^x \frac{(l_t - x)}{b_1} dx \tag{K.24}$$

and:

$$\left.\frac{dz_c}{dx}\right|_{x \ge l_1} = -\frac{12 F_{ext}}{Y t^3} \left[\int_0^{l_1} \frac{(l_t - x)}{b_1} dx + \int_{l_1}^x \frac{(l_t - x)}{-\lambda x + \mu} dx \right], \tag{K.25}$$

which becomes:

$$\left.\frac{dz_c}{dx}\right|_{x \le l_1} = -\frac{12 F_{ext}}{Y t^3} \frac{1}{b_1} \left[l_t x - \frac{1}{2} x^2 \right]_0^x \tag{K.26}$$

and:

$$\left.\frac{dz_c}{dx}\right|_{x \ge l_1} = -\frac{12 F_{ext}}{Y t^3} \left[\frac{1}{b_1} \left[l_t x - \frac{1}{2} x^2 \right]_0^{l_1} + \frac{1}{\lambda} \left[x + \frac{(\mu - \lambda l_t)}{\lambda} \ln(\lambda x - \mu) \right]_{l_1}^x \right]. \tag{K.27}$$

Substituting in the limits then gives:

$$\left.\frac{dz_c}{dx}\right|_{x \le l_1} = -\frac{12 F_{ext}}{Y t^3} \frac{1}{b_1} \left[l_t x - \frac{1}{2} x^2 \right] \tag{K.28}$$

and:

$$\left.\frac{dz_c}{dx}\right|_{x \ge l_1} = -\frac{12 F_{ext}}{Y t^3} \left[\frac{l_1}{b_1} \left(l_t - \frac{1}{2} l_1 \right) + \frac{1}{\lambda} \left((x - l_1) + \frac{(\mu - \lambda l_t)}{\lambda} \ln\left(\frac{\lambda x - \mu}{\lambda l_1 - \mu} \right) \right) \right]. \tag{K.29}$$

So the cantilever deflection is given by:

$$z_c(x \ge l_1) = -\frac{12 F_{ext}}{Y t^3} \left[\begin{array}{l} \dfrac{1}{b_1} \displaystyle\int_0^{l_1} \left(l_t x - \frac{1}{2} x^2 \right) dx + \frac{l_1}{b_1} \left(l_t - \frac{1}{2} l_1 \right) \int_{l_1}^x 1 dx + \\ \dfrac{1}{\lambda} \displaystyle\int_{l_1}^x (x - l_1) dx + \frac{(\mu - \lambda l_t)}{\lambda^2} \int_{l_1}^x \ln\left(\frac{\lambda x - \mu}{\lambda l_1 - \mu} \right) dx \end{array} \right], \tag{K.30}$$

which becomes:

$$z_c\left(x \geq l_1\right) = -\frac{12F_{ext}}{Yt^3}\left[\begin{array}{l}\frac{1}{b_1}\left[\frac{1}{2}l_t x^2 - \frac{1}{6}x^3\right]_0^{l_1} + \frac{l_1}{b_1}\left(l_t - \frac{1}{2}l_1\right)\left[x\right]_{l_1}^{x} \\ \quad + \frac{1}{\lambda}\left[\frac{1}{2}x^2 - l_1 x\right]_{l_1}^{x} + \frac{\left(\mu - \lambda l_t\right)}{\lambda^2}\left[\frac{\lambda x - \mu}{\lambda}\ln\left(\frac{\lambda x - \mu}{\lambda l_1 - \mu}\right) - x\right]_{l_1}^{x}\end{array}\right], \tag{K.31}$$

and upon substituting the limits yields:

$$z_c\left(x \geq l_1\right) = -\frac{12F_{ext}}{Yt^3}\left[\begin{array}{l}\frac{l_1^2}{2b_1}\left(l_t - \frac{1}{3}l_1\right) + \frac{l_1}{b_1}\left(l_t - \frac{1}{2}l_1\right)\left(x - l_1\right) + \frac{1}{2\lambda}\left(x - l_1\right)^2 \\ \quad + \frac{\left(\mu - \lambda l_t\right)}{\lambda^2}\left[\frac{\lambda x - \mu}{\lambda}\ln\left(\frac{\lambda x - \mu}{\lambda l_1 - \mu}\right) - \left(x - l_1\right)\right]\end{array}\right]. \tag{K.32}$$

Since only the deflection of the end of the cantilever is of interest, Equation K.32 can be evaluated at $x = l_t$ to give:

$$z_c\left(l_t\right) = -\frac{12F_{ext}}{Yt^3}\left[\begin{array}{l}\frac{l_1^2}{2b_1}\left(l_t - \frac{1}{3}l_1\right) + \frac{l_1 l_2}{b_1}\left(l_t - \frac{1}{2}l_1\right) + \frac{l_2^2}{2\lambda} \\ \quad + \frac{\left(\mu - \lambda l_t\right)}{\lambda^2}\left[\frac{\lambda l_t - \mu}{\lambda}\ln\left(\frac{\lambda l_t - \mu}{\lambda l_1 - \mu}\right) - l_2\right]\end{array}\right], \tag{K.33}$$

which can be simplified to:

$$z_c\left(l_t\right) = -\frac{12F_{ext}}{Yb_1 t^3}\left[\frac{1}{3}\left(l_1 + l_2\right)^3 + \frac{b_1^2 - 5b_1 b_2 - 2b_2^2}{6\left(b_1 - b_2\right)^2}l_2^3 + \frac{b_1 b_2^2 \ln b_f}{\left(b_1 - b_2\right)^3}l_2^3\right]. \tag{K.34}$$

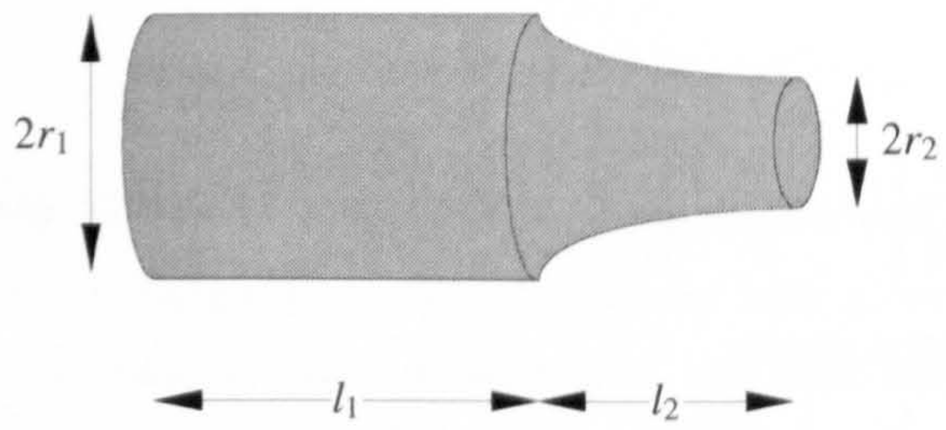


Figure K.4: Diagram of a simple tip with circular cross-section.

The second geometry is illustrated in Figure K.4. This cantilever has a circular cross-section and also consists of two sections: the first section of length, l_1, is of uniform radius, r_1; the second section of length, l_2, tapers exponentially from radius, r_1, to radius, r_2.

It is useful to define the following quantities:

$$l_t = l_1 + l_2, r_f = \frac{r_1}{r_2}, \lambda = \frac{\ln r_f}{l_2}, \mu = \lambda l_1. \tag{K.35}$$

Then it is clear that:

$$r(x) = \begin{cases} r_1, & 0 \leq x \leq l_1 \\ r_1 \exp\left(-\lambda x + \mu\right), & l_1 \leq x \leq l_t \end{cases}. \tag{K.36}$$

From Equations K.16 and K.19:

$$\frac{d^2 z_c}{dx^2} = -\frac{4F_{ext}}{\pi Y}\frac{(l_t - x)}{r^4(x)}, \tag{K.37}$$

which becomes:

$$\frac{d^2 z_c}{dx^2} = -\frac{4F_{ext}}{\pi Y}\begin{cases} \dfrac{(l_t - x)}{r_1^4}, & 0 \le x \le l_1 \\ \dfrac{(l_t - x)}{r_1^4 e^{-4(\lambda x - \mu)}}, & l_1 \le x \le l_t \end{cases}, \tag{K.38}$$

so that:

$$\left.\frac{dz_c}{dx}\right|_{x \le l_1} = -\frac{4F_{ext}}{\pi Y}\int_0^x \frac{(l_t - x)}{r_1^4}dx \tag{K.39}$$

and:

$$\left.\frac{dz_c}{dx}\right|_{x \ge l_1} = -\frac{4F_{ext}}{\pi Y}\left[\int_0^{l_1}\frac{(l_t - x)}{r_1^4}dx + \int_{l_1}^x \frac{(l_t - x)}{r_1^4 e^{-4(\lambda x - \mu)}}dx\right], \tag{K.40}$$

which becomes:

$$\left.\frac{dz_c}{dx}\right|_{x \le l_1} = -\frac{4F_{ext}}{\pi Y}\frac{1}{r_1^4}\left[l_t x - \frac{1}{2}x^2\right]_0^x \tag{K.41}$$

and:

$$\left.\frac{dz_c}{dx}\right|_{x \ge l_1} = -\frac{4F_{ext}}{\pi Y}\left[\frac{1}{r_1^4}\left[l_t x - \frac{1}{2}x^2\right]_0^{l_1} + \frac{1}{4\lambda r_1^4 e^{4\mu}}\left[(l_t - x)e^{4\lambda x} + \frac{1}{4\lambda}e^{4\lambda x}\right]_{l_1}^x\right]. \tag{K.42}$$

Substituting the limits then gives:

$$\left.\frac{dz_c}{dx}\right|_{x \le l_1} = -\frac{4F_{ext}}{\pi Y r_1^4}\left[l_t x - \frac{1}{2}x^2\right] \tag{K.43}$$

and:

$$\left.\frac{dz_c}{dx}\right|_{x \ge l_1} = -\frac{4F_{ext}}{\pi Y r_1^4}\left[l_1\left(l_t - \frac{1}{2}l_1\right) + \frac{1}{4\lambda}\left(\left(l_t - x + \frac{1}{4\lambda}\right)e^{4\lambda(x - l_1)} - \left(l_t - l_1 + \frac{1}{4\lambda}\right)\right)\right]. \tag{K.44}$$

Then the cantilever deflection is given by:

$$z_c(x \ge l_1) = -\frac{4F_{ext}}{\pi Y r_1^4}\left[\begin{array}{l} \displaystyle\int_0^{l_1}\left(l_t x - \frac{1}{2}x^2\right)dx + l_1\left(l_t - \frac{1}{2}l_1\right)\int_{l_1}^x 1dx + \\ \displaystyle\frac{e^{-4\mu}}{4\lambda}\int_{l_1}^x\left(l_t - x + \frac{1}{4\lambda}\right)e^{4\lambda x}dx - \frac{1}{4\lambda}\left(l_t - l_1 + \frac{1}{4\lambda}\right)\int_{l_1}^x 1dx \end{array}\right], \tag{K.45}$$

which becomes:

$$z_c(x \ge l_1) = -\frac{4F_{ext}}{\pi Y r_1^4}\left[\begin{array}{l} \left[\frac{1}{2}l_t x^2 - \frac{1}{6}x^3\right]_0^{l_1} + l_1\left(l_t - \frac{1}{2}l_1\right)[x]_{l_1}^x + \\ \frac{e^{-4\mu}}{16\lambda^2}\left[\left(l_t - x + \frac{1}{2\lambda}\right)e^{4\lambda x}\right]_{l_1}^x - \frac{1}{4\lambda}\left(l_t - l_1 + \frac{1}{4\lambda}\right)[x]_{l_1}^x \end{array}\right], \tag{K.46}$$

and upon substituting the limits yields:

$$z_c(x \geq l_1) = -\frac{4F_{ext}}{\pi Y r_1^4}\left[\begin{array}{l}\frac{1}{2}l_1^2\left(l_t-\frac{1}{3}l_1\right)+l_1\left(l_t-\frac{1}{2}l_1\right)(x-l_1)-\frac{1}{4\lambda}\left(l_t-l_1+\frac{1}{4\lambda}\right)(x-l_1)\\ \quad+\frac{1}{16\lambda^2}\left[\left(l_t-x+\frac{1}{2\lambda}\right)e^{4\lambda(x-l_1)}-\left(l_t-l_1+\frac{1}{2\lambda}\right)\right]\end{array}\right]. \quad \text{(K.47)}$$

Again, since only the deflection of the end of the cantilever is of interest, Equation K.47 can be evaluated at $x = l_t$ to give:

$$z_c(l_t) = -\frac{4F_{ext}}{\pi Y r_1^4}\left[\begin{array}{l}\frac{1}{2}l_1^2\left(l_t-\frac{1}{3}l_1\right)+l_1l_2\left(l_t-\frac{1}{2}l_1\right)-\frac{l_2}{4\lambda}\left(l_2+\frac{1}{4\lambda}\right)\\ \quad+\frac{1}{16\lambda^2}\left[\frac{1}{2\lambda}e^{4\lambda l_2}-\left(l_2+\frac{1}{2\lambda}\right)\right]\end{array}\right], \quad \text{(K.48)}$$

which can be simplified to:

$$z_c(l_t) = -\frac{4F_{ext}}{\pi Y r_1^4}\left[\frac{1}{3}(l_1+l_2)^3+\left(\frac{1}{32\ln^3 r_f}\left(r_f^4-1\right)-\frac{1}{8\ln^2 r_f}-\frac{1}{4\ln r_f}-\frac{1}{3}\right)l_2^3\right]. \quad \text{(K.49)}$$

From Equations K.34 and K.49 the compliance of both cantilever geometries can be calculated by using Equation K.17 to give:

$$\begin{aligned} k_c^{\text{rectangular}} &= \frac{Yb_1t^3}{12}\left[\frac{1}{3}(l_1+l_2)^3+\frac{b_1^2-5b_1b_2-2b_2^2}{6(b_1-b_2)^2}l_2^3+\frac{b_1b_2^2\ln b_f}{(b_1-b_2)^3}l_2^3\right]^{-1} \\ k_c^{\text{circular}} &= \frac{\pi Y r_1^4}{4}\left[\frac{1}{3}(l_1+l_2)^3+\left(\frac{1}{32\ln^3 r_f}\left(r_f^4-1\right)-\frac{1}{8\ln^2 r_f}-\frac{1}{4\ln r_f}-\frac{1}{3}\right)l_2^3\right]^{-1}. \end{aligned} \quad \text{(K.50)}$$

Some special cases can now be considered. In the case where the cantilever is of uniform cross-section, *i.e.* $l_2 = 0$, the compliance is given by:

$$k_c^{\text{rectangular}} = \frac{Ybt^3}{4l^3}, \quad k_c^{\text{circular}} = \frac{3\pi Y r^4}{4l^3}. \quad \text{(K.51)}$$

In the case where the cantilever is fully tapered, *i.e.* $l_1 = 0$, the compliance is given by:

$$k_c^{\text{rectangular}} = \frac{Yt^3}{6l^3}\left[\frac{(b_1-b_2)^3}{(b_1-b_2)^2-2b_2(b_1-b_2\ln b_f-b_2)}\right]. \quad \text{(K.52)}$$

$$k_c^{\text{circular}} = \frac{8\pi Y r_1^4}{l^3}\left[\frac{\ln^3 r_f}{r_f^4-8\ln^2 r_f-4\ln r_f-1}\right]$$

K.2 CONTACT MODE CANTILEVERS

From Figure 4.4.1 in §4.4.2, the cantilever dimensions used for contact mode cantilevers (as measured to the apex of the tip, not to the end of the cantilever) are given by:

$$l_1 = l_2 = 0.5l_t,\ b_1 = 0.5l_t + 45\mu\text{m},\ b_2 = 45\mu\text{m}. \tag{K.53}$$

The mass of these cantilevers, m_c, in picograms is then given by:

$$m_c = \rho_c t\left(b_1 l_t - 0.5\Delta b l_2\right) = 0.375\rho_c t l_t\left(l_t + 120\mu\text{m}\right), \tag{K.54}$$

where ρ_c is the density of the cantilever in grams per cubic centimetre and all cantilever dimensions are entered in microns. Substituting Equation K.53 into Equation K.50 gives:

$$k_c = \frac{\left(0.5l_t + 45\right)Yt^3}{12\times10^6}\left[\frac{17l_t^2 - 6\times45l_t - 24\times45^2}{48}l_t + \left(\frac{45^2 l_t}{2} + 45^3\right)\ln\left(\frac{l_t}{90}+1\right)\right]^{-1} \tag{K.55}$$

where all cantilever dimensions are entered in microns. The resonant frequency, f_c, of these cantilevers can be approximated by using [Petersen[1] (1978)]:

$$f_c \approx \frac{1}{\pi}\sqrt{\frac{k_c}{m_c}}, \tag{K.56}$$

which is strictly only valid for cantilevers of uniform rectangular cross-section. However, this formula will be used here as an order of magnitude estimator of the resonant frequency of these cantilevers.

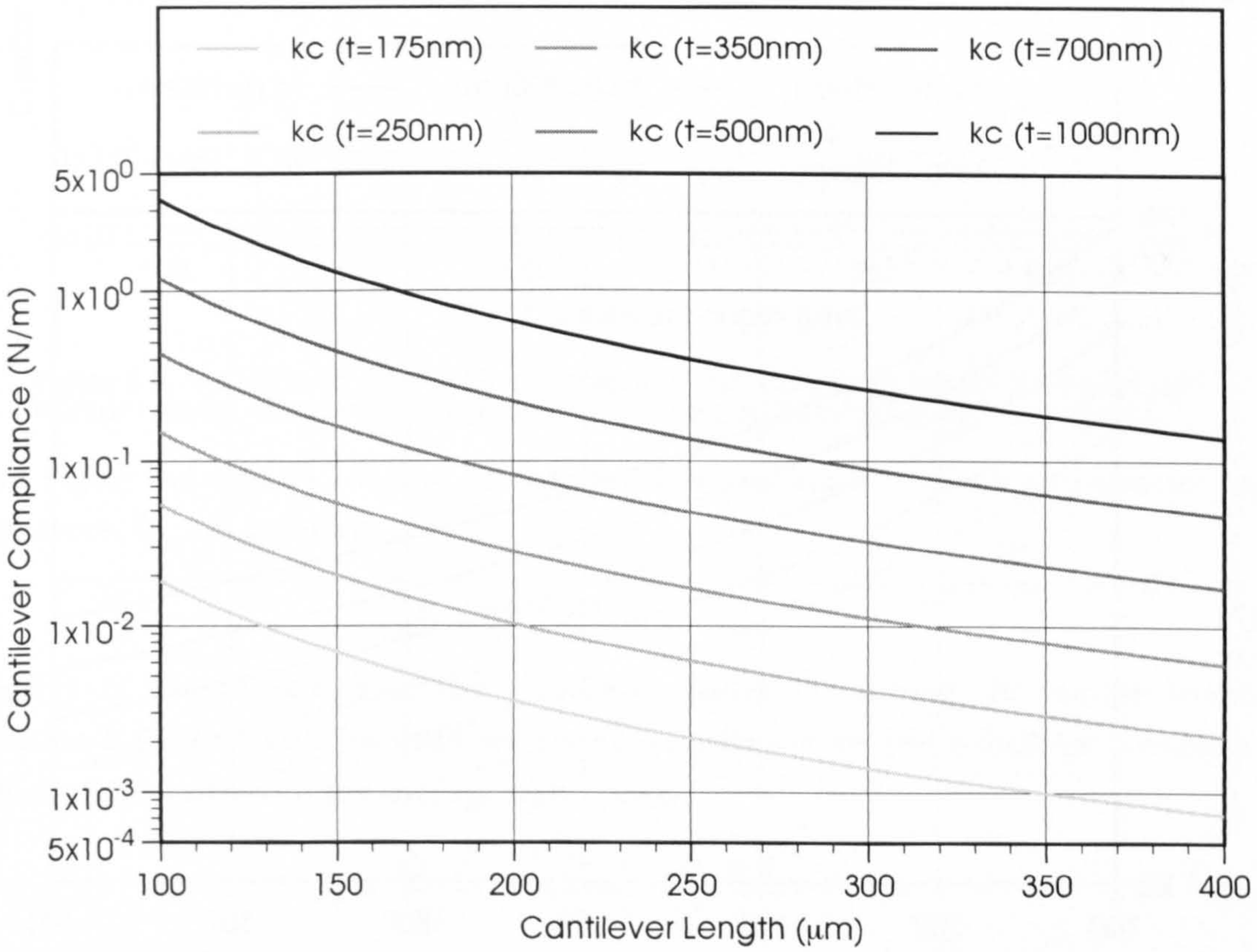


Figure K.5: Variation of cantilever compliance with cantilever length for silicon nitride contact mode AFM cantilevers conforming to the geometry of Figure 4.4.1 in §4.4.2.

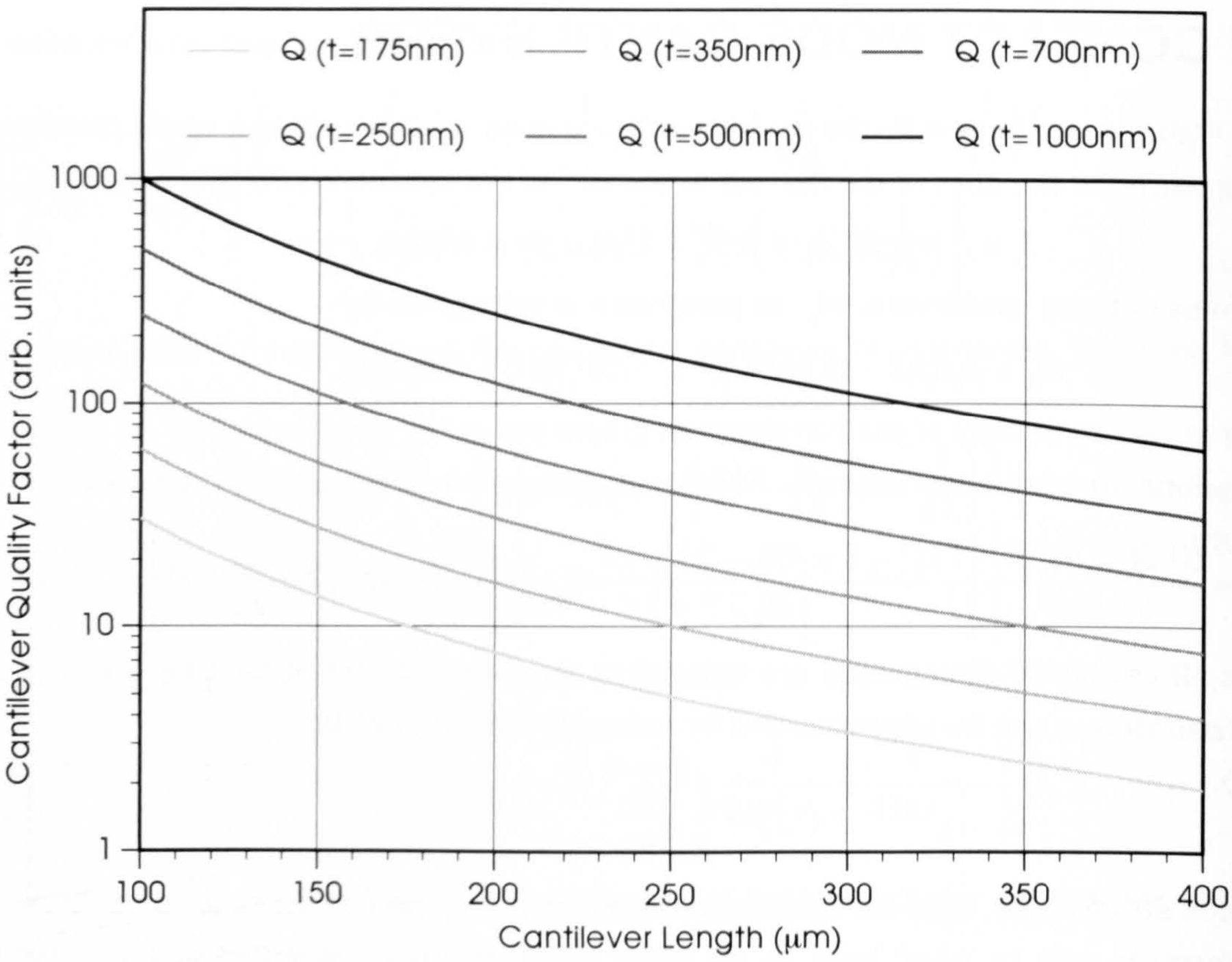


Figure K.6: Relative variation of cantilever quality factor with cantilever length for silicon nitride contact mode AFM cantilevers conforming to the geometry of Figure 4.4.1 in §4.4.2.

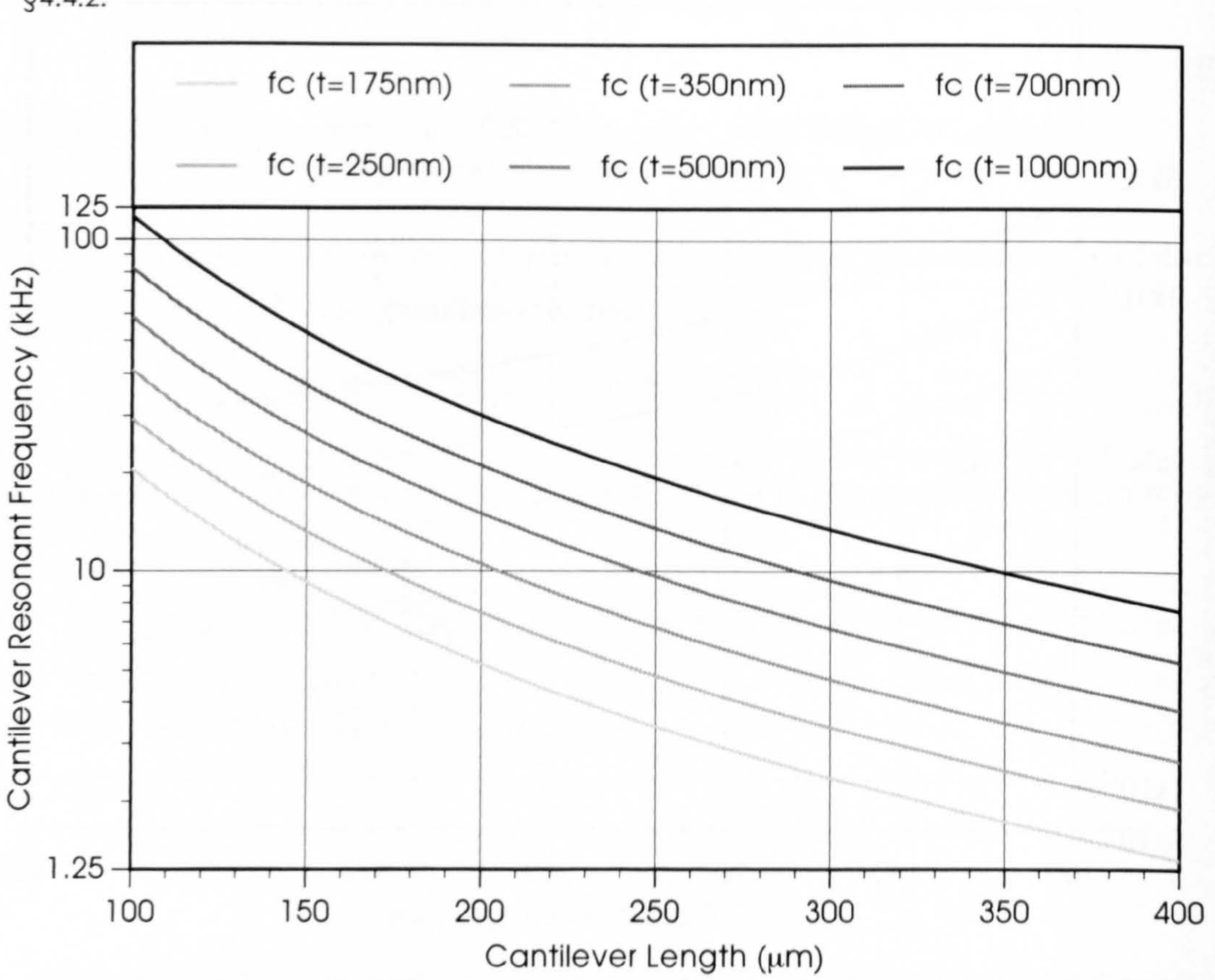


Figure K.7: Variation of cantilever resonant frequency with cantilever length for silicon nitride contact mode AFM cantilevers conforming to the geometry of Figure 4.4.1 in §4.4.2.

Figure K.5 plots the cantilever compliance as a function of cantilever length for six values of cantilever thickness using data given in Table 4.4.1 in §4.4.1. Figure K.6 plots the relative quality factor (with the highest quality factor normalised to 1000) and Figure K.7 plots the approximate resonant frequency as a function of cantilever length for six values of cantilever thickness using the data given in Table 4.4.1.

K.3 NON-CONTACT MODE CANTILEVERS

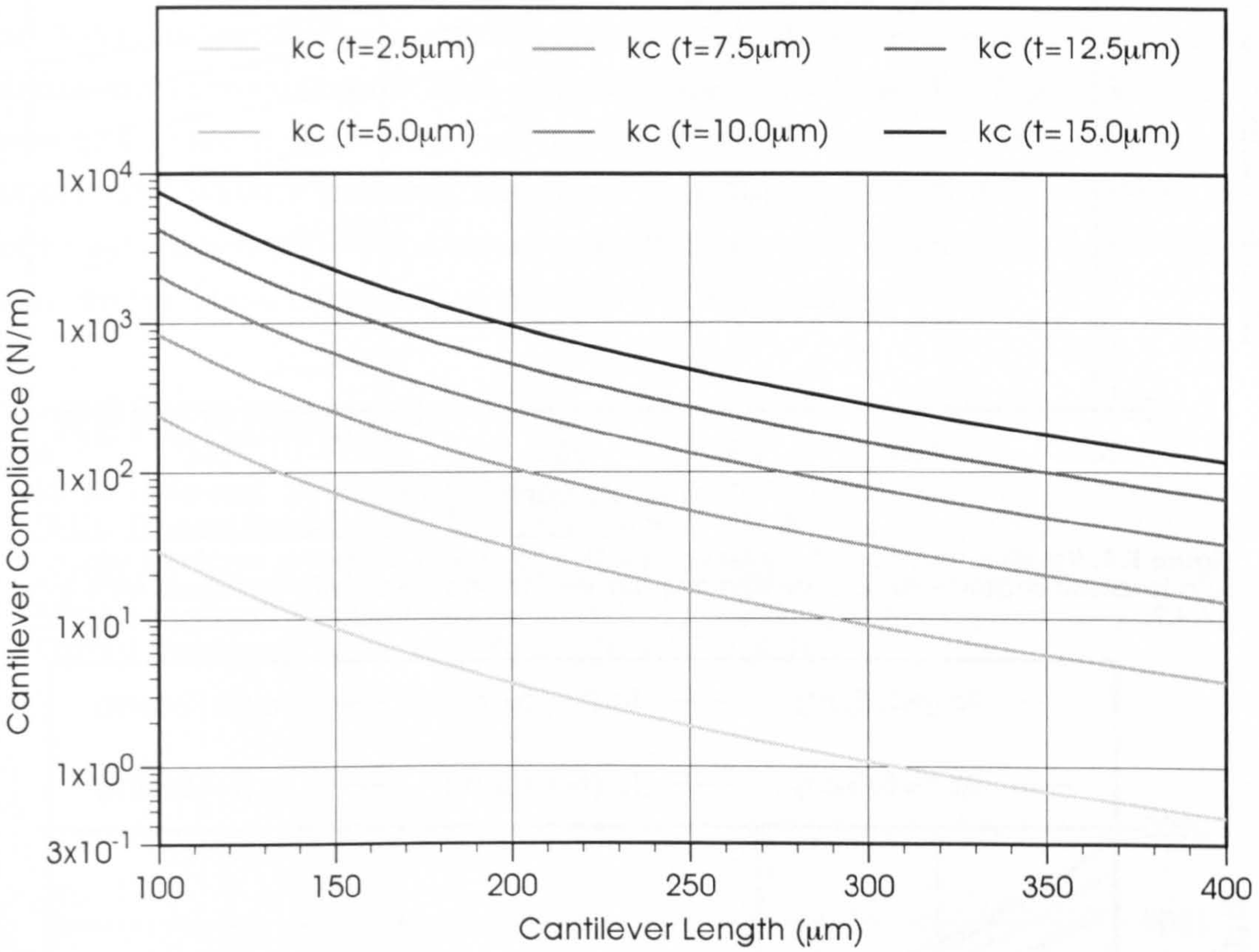


Figure K.8: Variation of cantilever compliance with cantilever length for silicon non-contact mode AFM cantilevers conforming to the geometry of Figure 4.4.4 in §4.4.3.

From Figure 4.4.4 in §4.4.3, the average width of the beam used for non-contact mode cantilevers, b_{ave}, is given by:

$$b_{ave} = 40\mu\text{m} + \frac{t}{\sqrt{2}}, \tag{K.57}$$

where t is the thickness of the cantilever beam. Assuming, for simplicity, that Equation K.51 can still be used as an approximation to the cantilever stiffness by replacing the width with the average width, then:

$$k_c = \frac{Yb_{ave}t^3}{4l^3} = \frac{Y\left(t + 40\sqrt{2}\mu\text{m}\right)t^3}{4\sqrt{2}l^3}. \tag{K.58}$$

The mass of the cantilever, m_c, excluding the tip is given by:

$$m_c = \rho_c t b_{ave} l = \frac{1}{\sqrt{2}}\rho_c t\left(t + 40\sqrt{2}\mu\text{m}\right)l. \tag{K.59}$$

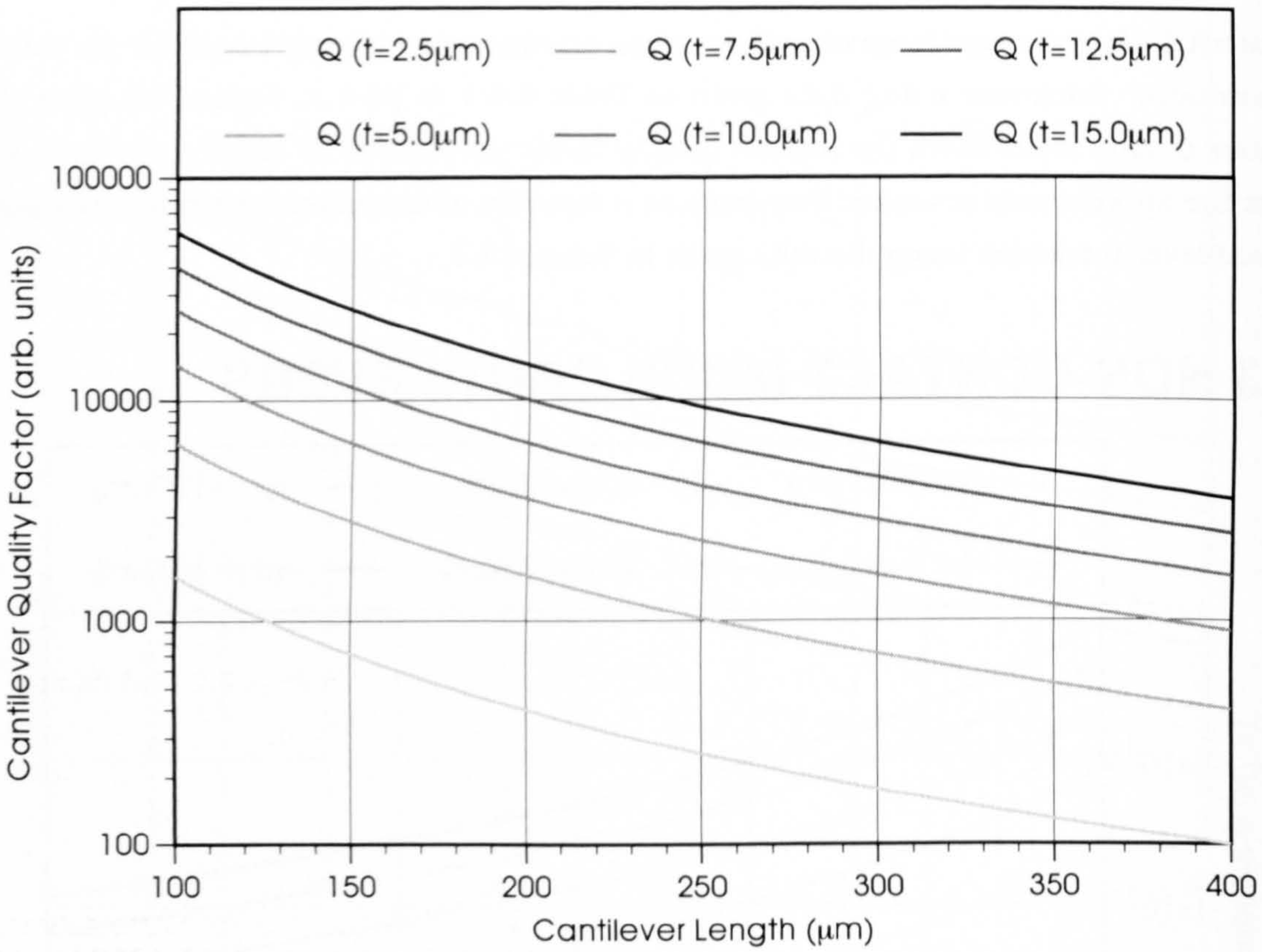


Figure K.9: Relative variation of cantilever quality factor with cantilever length for silicon non-contact mode AFM cantilevers conforming to the geometry of Figure 4.4.4 in §4.4.3.

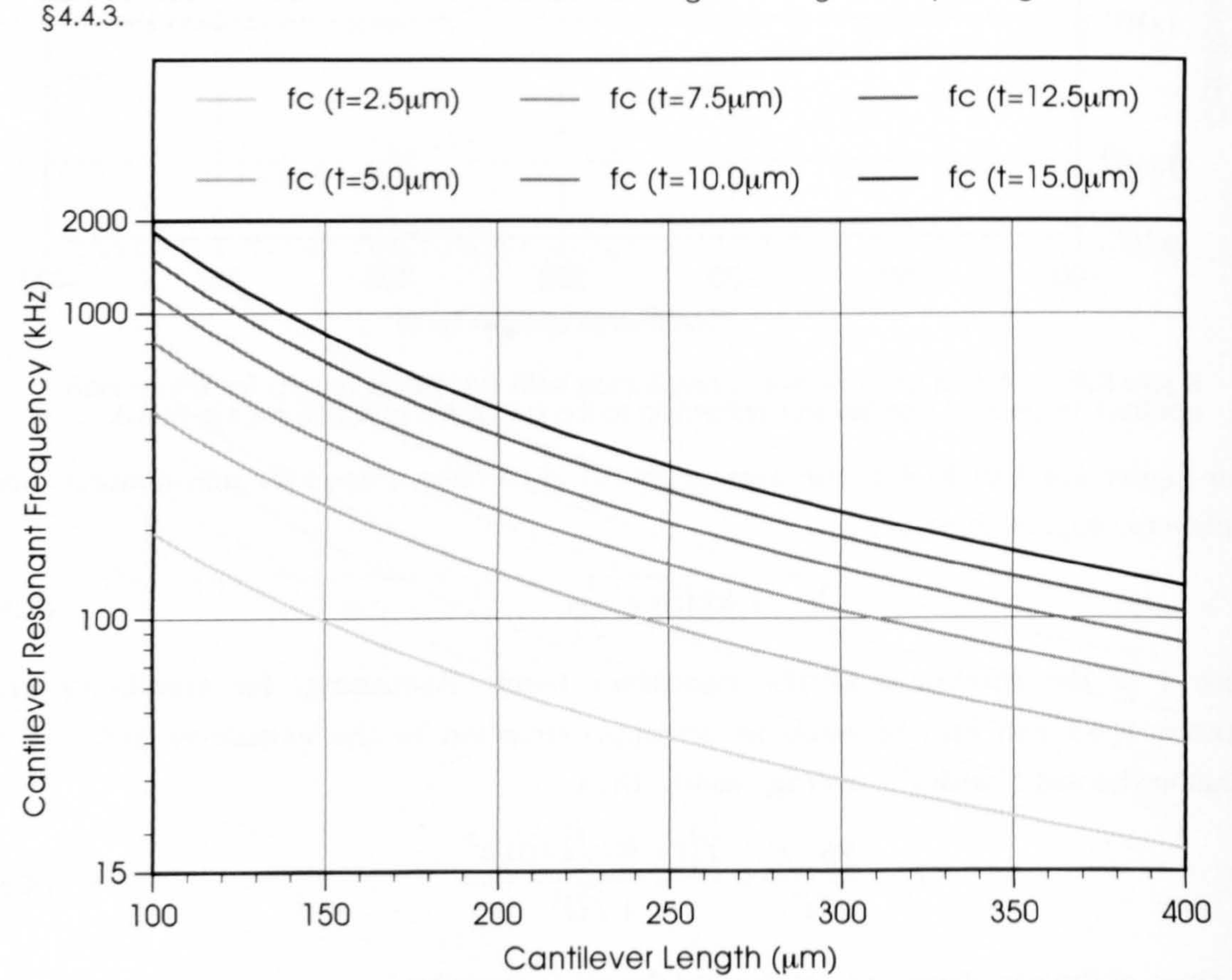


Figure K.10: Variation of cantilever resonant frequency with cantilever length for silicon non-contact mode AFM cantilevers conforming to the geometry of Figure 4.4.4 in §4.4.3.

Approximating the tip as a square based pyramid, the mass of the tip, m_t, is given by:

$$m_t = \frac{1}{3}\rho_c b_p{}^2 h, \tag{K.60}$$

where ρ_c is the density of the cantilever (silicon) in kg/m^3. Then from Sarid[2] (1991) the resonant frequency of the cantilever, f_c, is approximately given by:

$$f_c \approx \frac{1}{2\pi}\sqrt{\frac{k}{m_t + 0.24m_c}} = \frac{1}{4\pi}\sqrt{\frac{3Yb_{ave}t^3}{\rho_c l^3\left(b_p{}^2 h + 0.72tb_{ave}l\right)}}. \tag{K.61}$$

Figure K.8 plots the cantilever stiffness as a function of cantilever length for six values of cantilever thickness using the data given in Table 4.4.1 in §4.4.1. Figure K.9 plots the relative quality factor (with the lowest quality factor normalised to 100) and Figure K.10 plots the approximate resonant frequency as a function of cantilever length for six different values of cantilever thickness using the data given in Table 4.4.1 and assuming that the tip is a 20μm tall square pyramid with a base width of 30μm.

K.4 REFERENCES

1. K. E. Petersen, *DYNAMIC MICROMECHANICS ON SILICON: TECHNIQUES AND DEVICES*, IEEE Transactions On Electron Devices **Vol. ED-25,** No. 10, pp. 1241-1250 (1978).

2. D. Sarid, *Chapter 1: MECHANICAL PROPERTIES OF LEVERS* in Scanning Force Microscopy with Applications to Electric, Magnetic and Atomic Forces, published by Oxford University Press, New York, pp. 1-17 (1991).

APPENDIX L

CANTILEVER DEVELOPMENT

L.1 CONTACT MODE CANTILEVERS

The initial attempts to fabricate contact mode AFM cantilevers consisted of depositing 500nm of low stress silicon nitride by PECVD onto the substrates followed by 100nm of nichrome (60% nickel/40% chromium) in the Plassys MEB 450. As noted in §4.4.2, the coating of these substrates with PMMA was not carried out simply by using the spin coating technique, as it was found that spin coating would not adequately coat the apices of the pyramidal tips. Instead it was combined with an alternative coating technique, which has been named 'float coating'. The development of this technique is discussed in detail in §5.2. Suffice to say that with this technique the resist coating is highly stressed so that the resist close to the base of the pyramidal tips is sometimes discontinuous. To overcome this, both spin coating and float coating techniques were used in a complementary manner. First the substrate was spin coated using 4% BDH spun at 2500 rpm for 60 seconds and then baked overnight at 180°C. Then the substrate was float coated with 4% BDH and baked for 30 minutes at 80°C.

The resist was then exposed by the beamwriter with a dose of 350μC/cm^2 and a spot size of 300nm. The pattern simply consisted of horizontal strips [see Figures 4.2.3 and 4.4.3] with the cantilever geometries of Figure 4.4.2 subtracted from the exposure area. As the alignment accuracy of this level was not critical, only a single alignment was carried out using markers at four corners of the substrate and the pattern was written as a single cell. The resist was then developed in MIBK:IPA 1:1 for 35 seconds.

The nichrome layer was etched in chrome etch for 6 minutes (again the level of resist undercut was not a problem for this level). The substrate was then ready for a 15 minute dry etch in C_2F_6 in the Plasma Technology BP80. After this the nichrome could be stripped off in a 10 minute chrome etch.

Since the release etch would only have to etch through about 45μm of silicon, it was decided that the release etch be carried out using a slower etch. The wafers were etched in 7 molar (+5%) potassium hydroxide solution at 65°C. This gives a one-sided etch rate of about 28μm/hr. However, when the etch has penetrated the wafer, large areas of the wafer become transparent, offering a visual indicator that the etch has reached the halfway point. Since the height of the pyramidal tips and the etch depth of the membrane definition windows can be measured prior to etching (by taking precautions to avoid smashing the pyramidal tips in the process) and knowing the initial wafer thickness, the thickness of silicon to be etched can also be determined. When the substrate becomes transparent, an estimate can then be made as to how much longer the etch should be allowed to continue. In this way the etch can be made self-calibrating. After etching, the substrate would then be neutralised in acid and given an organic clean [Appendix D] before being gently blown dry from IPA.

Figure L.1 shows the first low stress silicon nitride cantilevers produced by this process. The pattern in Figure L.1b has been badly misaligned due to an error on the part of the author. However, both micrographs clearly show the lack of any curling of the cantilever indicating that the silicon nitride indeed has a low level of residual stress. Unfortunately, both micrographs highlight a problem with the pattern used. Since convex corner undercutting has been used to taper the silicon substrate approaching the cantilever [see Figure 4.2.3], the simple pattern used above results in the formation of wide silicon nitride flaps. These flaps can break in an untidy way, as in the leftmost micrograph, and endanger the cantilevers of adjacent devices. Therefore to avoid this problem, the pattern was modified so as to remove the silicon nitride at an angle of 45° from the base of the cantilever to the cleave line between devices during the dry etch step.

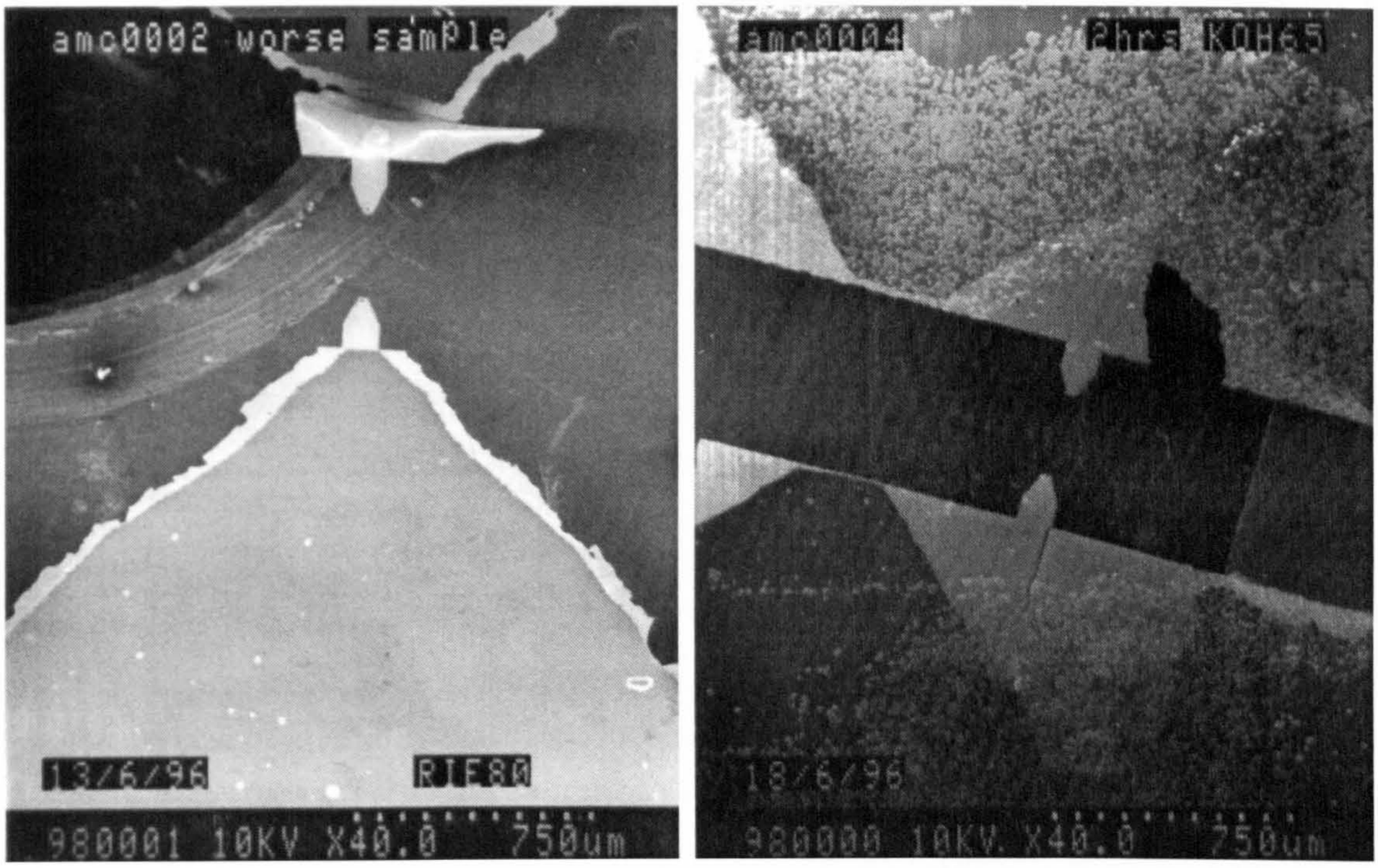


Figure L.1: Micrographs showing some of the problems associated with releasing cantilevers for contact mode AFM. The images are of 250μm long, 500nm thick low stress PECVD silicon nitride cantilevers. (a) Left Image: the untidy edges are due to the silicon nitride fracturing rather than being patterned; (b) Right Image: this pattern was misaligned, but the flatness of the cantilever and the silicon nitride 'flaps' can be clearly seen.

Figure L.2 illustrates the effect of this change. It is clear that while the edges of the cantilevers are now much tidier and pose no threat to other devices, the 45° angle has allowed the cantilever to become undercut by the etch due to the exposure of an edge not along an [011] axis. This means that the base of the cantilever is attached to silicon nitride rather than being firmly anchored on the silicon substrate. The problem with this is that if the cantilever is placed under stress, as in the course of normal use, cracks may form where the base of the cantilever meets the rest of the silicon nitride. To overcome this, the pattern was once again modified to introduce 'shoulders' around 100μm wide along the [011] axes adjacent to the base of the cantilever. These shoulders would also be undercut by the etch, but by the time the etch front reaches the cantilever, the etch will

have been completed. This compensation can be adjusted by simply varying the width of the shoulder. Figure L.3 illustrates the success of this scheme. Figure 4.4.3 in §4.4.2 illustrates the final pattern used.

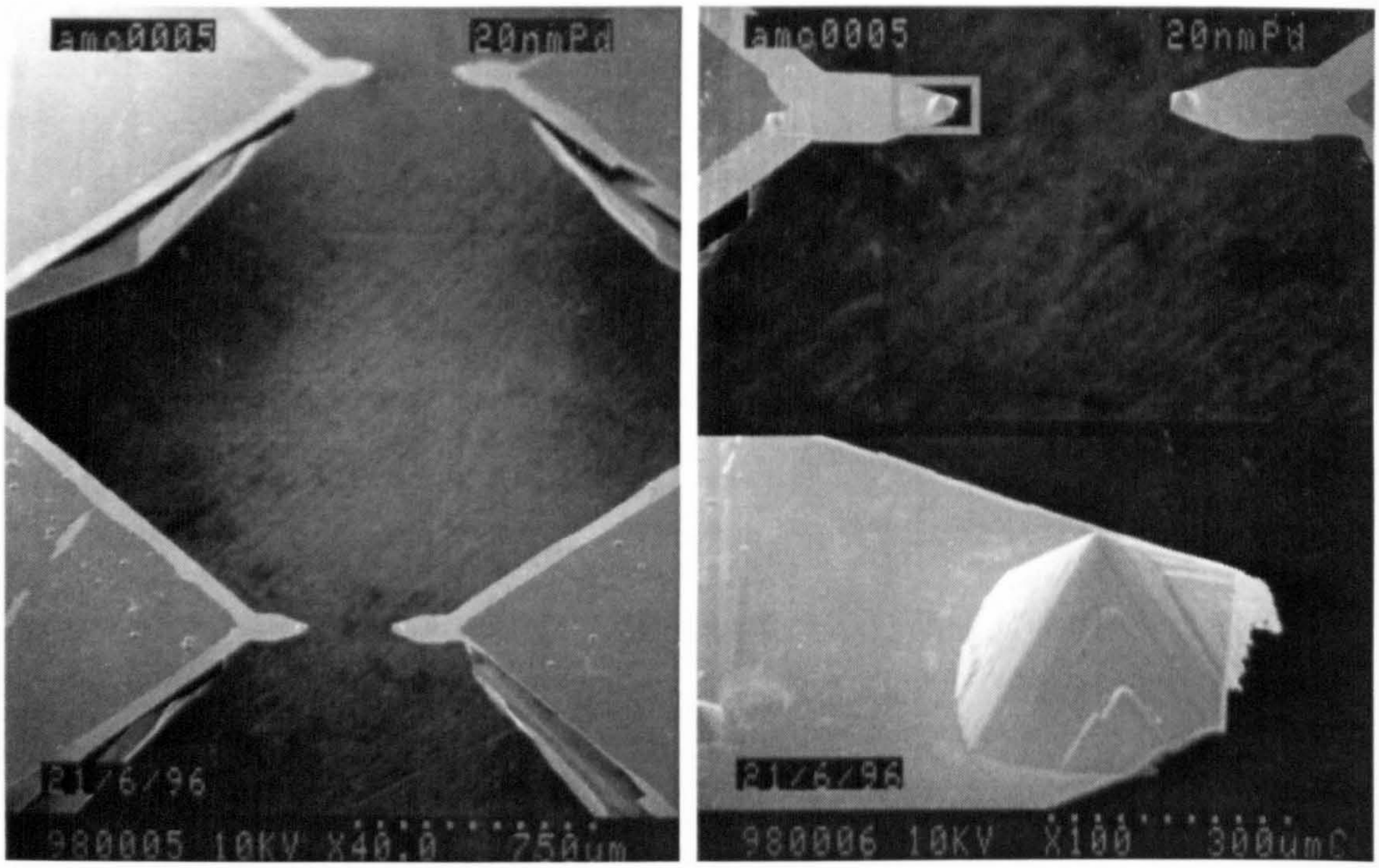


Figure L.2: Micrographs showing some of the problems associated with releasing cantilevers for contact mode AFM. The images are of 250μm long, 500nm thick low stress PECVD silicon nitride cantilevers. The patterning of the silicon nitride below the base of the cantilever has eliminated the 'flaps' of Figure L.1, but introduced a cantilever base undercutting problem. The bottom half of the right image is a 10x magnification over the displayed scale which applies to the top half.

Unfortunately, during the development of this process a transition had to be made from BDH to Aldrich [see §3.5.3], which did not go as smoothly as hoped. It was discovered that using a PMMA scheme of 8% equivalent Aldrich spun at 2500 rpm for 60 seconds and 4% equivalent Aldrich float coated resulted in the nichrome at the top of the pyramidal tips being attacked by the chrome etch. The reason for this is not clear, but one theory is that because the nature of the float coating technique is such that the PMMA is stretched over the tips [see §5.2], the higher molecular weight of the Aldrich in comparison with BDH means that the resulting layer is more porous. This increase in porosity may be sufficient to allow the active components of the etching solution to penetrate the resist layer where the resist is most stressed. The only evidence to support this theory is that on the one occasion upon which the float coated layer consisted of 4% Elvacite, which is of an even higher molecular weight, the nichrome was also attacked.

Working around this problem unfortunately meant the addition of an extra EBL level prior to dry etching the silicon nitride but after wet etching the nichrome layer and stripping off the PMMA. The purpose of this level was to lift-off a 10μm square of 50nm thick nichrome over the apices of the tips as a protection layer.

A different problem later arose with the nichrome being used. For some mysterious reason the layers of nichrome being deposited began to crack, peel and curl up at a thickness of 100nm. This was attributed to an increase in the residual stress of the film. The cause of this is not known, although it may be associated with the fact that the nichrome used is not eutectic and is being evaporated from a single source in an electron-beam evaporator, rather than using a dual-gun chromium-nickel system as is more conventional [Gourlay[1] (1993-1996)]. In any case, the problem seemed to affect isolated and small scale structures more than large structures, and thicker layers more than thinner layers.

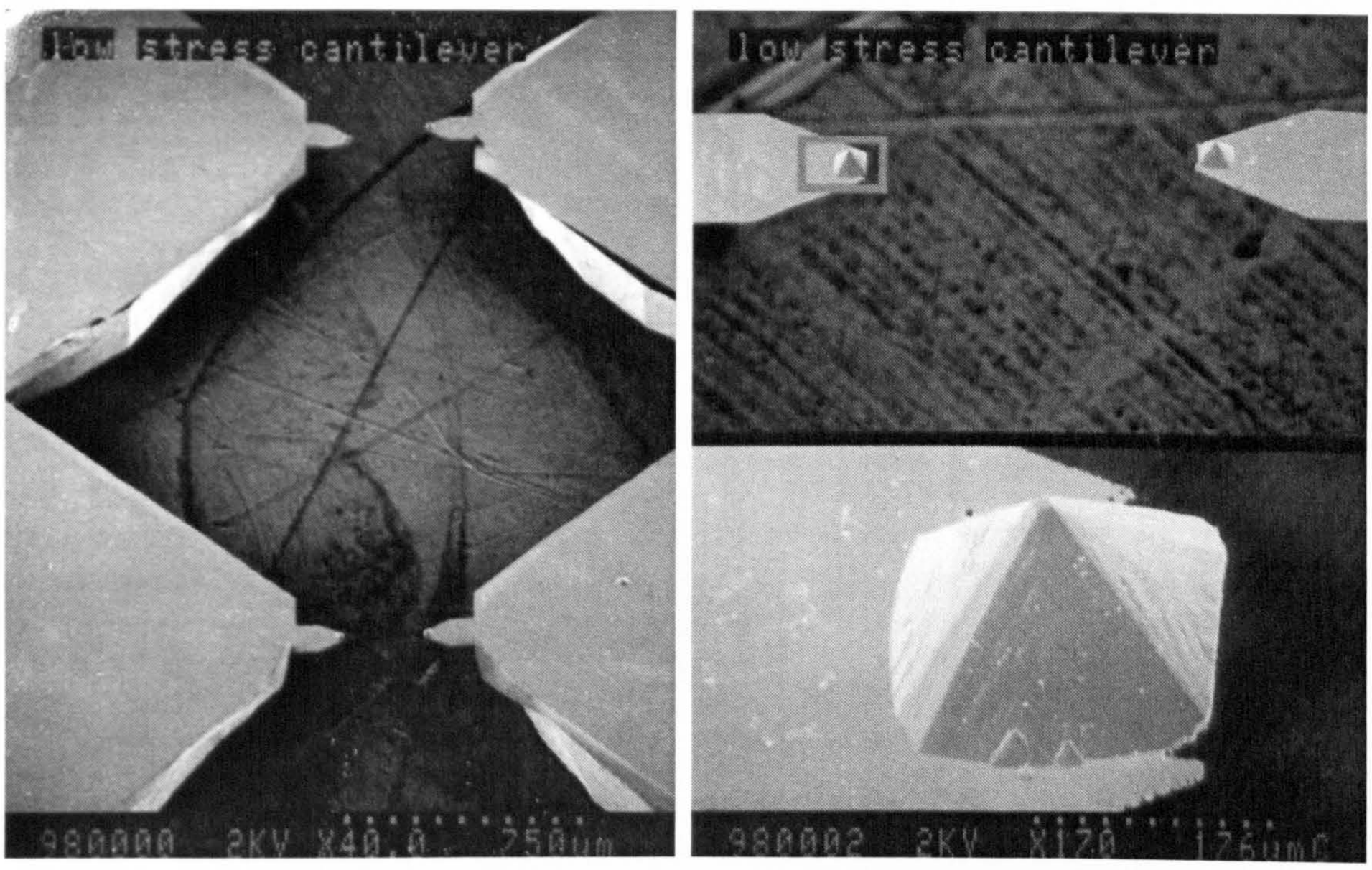


Figure L.3: Micrographs showing some of the problems associated with releasing cantilevers for contact mode AFM. The images are of 250μm long, 500nm thick low stress PECVD silicon nitride cantilevers. The introduction of 'shoulders' at the base of the cantilever has eliminated the undercutting of the anchoring point of the cantilevers of Figure L.2. The bottom half of the right image is a 10x magnification over the displayed scale which applies to the top half.

Another symptom seemed to be the difficulty and variability with which the nichrome was being etched using chrome etch, with some stubborn deposits, probably nickel rich, often remaining. After an initial trial with 50nm thick nichrome, it was decided that the best way forward for this level would be to have a 100nm layer of chromium sputtered onto the substrate immediately after silicon nitride deposition. This would have the advantage of conformally covering the surface. Therefore a 4 inch chromium sputter target was obtained with a view to coating whole 3 inch wafers. This seemed to overcome all the remaining problems with this level. It should be noted that the protection layer lift-off level was not affected by this problem and remained as a nichrome layer.

One further technical detail was that of the positioning of the base of the cantilever with respect to the expected edge of the 'cliff base' being etched from the other side of the

substrate [see Figures 4.2.2 and 4.2.3 in §4.2, Figure 4.2.2 can be interpreted as a cross-sectional view along a line joining the centre of the 'cliff bases' of devices 'A' and 'B' in Figure 4.2.3]. To accommodate feature growth and some misalignment, the cantilevers were deliberately recessed into the silicon substrate by 5μm [see Figure 4.4.3 in §4.4.2]. This allowed for a 5μm error in the position at which the 'cliff base' would appear after the release etch. This was done to ensure that the cantilever would always be anchored onto the silicon substrate and avoid the situation where the 'cliff base' appeared behind the cantilever.

The detailed processing steps for this level are listed in Appendix G (§G.3).

L.2 NON-CONTACT MODE CANTILEVERS

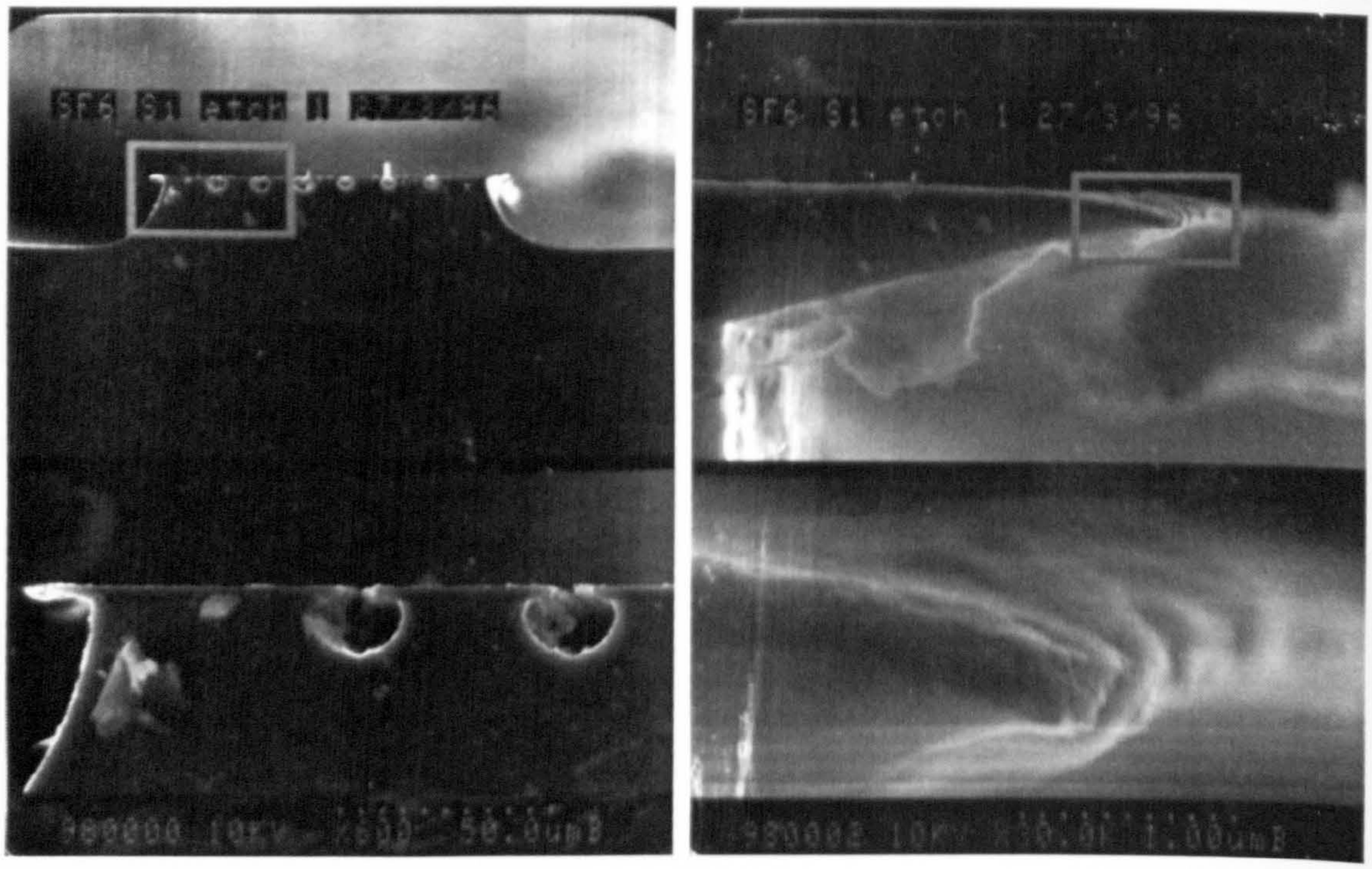


Figure L.4: Micrographs showing a possible deep silicon dry etch for defining non-contact mode AFM cantilevers. (a) Left Image: the etch is faster in open areas than closed areas; (b) Right Image: the etch undercuts the mask significantly though the mask remains intact. Some polymer deposition has also taken place. The bottom images are a 5x magnification over the displayed scales which apply to the top images.

Initial thoughts on fabrication focused on the idea of simply vertically dry etching through the silicon membrane using a nichrome metal mask. In order to end up with a beam of between 10μm and 20μm thick, it would be necessary to make sure that the cantilever membrane was sufficiently thin after the tip definition etch. This could be done before the tip definition level and immediately after the second photolithography level wet etch by introducing another slower wet etch between these two levels. After the second level wet etch, the cantilever membrane etch depth could be measured and so the membrane thickness could be found by subtracting this from the wafer thickness. As the tip definition etch would reduce this thickness by approximately 20μm from both sides,

the membrane thickness at this point should be between 50μm and 60μm. This could be obtained by carrying out a slow etch in 7 molar (+5%) potassium hydroxide solution at 65°C (as used for the release etch in §L.1).

Alternatively, the second level wet etch could be increased by approximately 5 minutes to a value of 75 minutes giving an expected etch depth of about 320μm and leaving a membrane about 60μm thick. However, this would be more difficult to control accurately so that the first option was preferred.

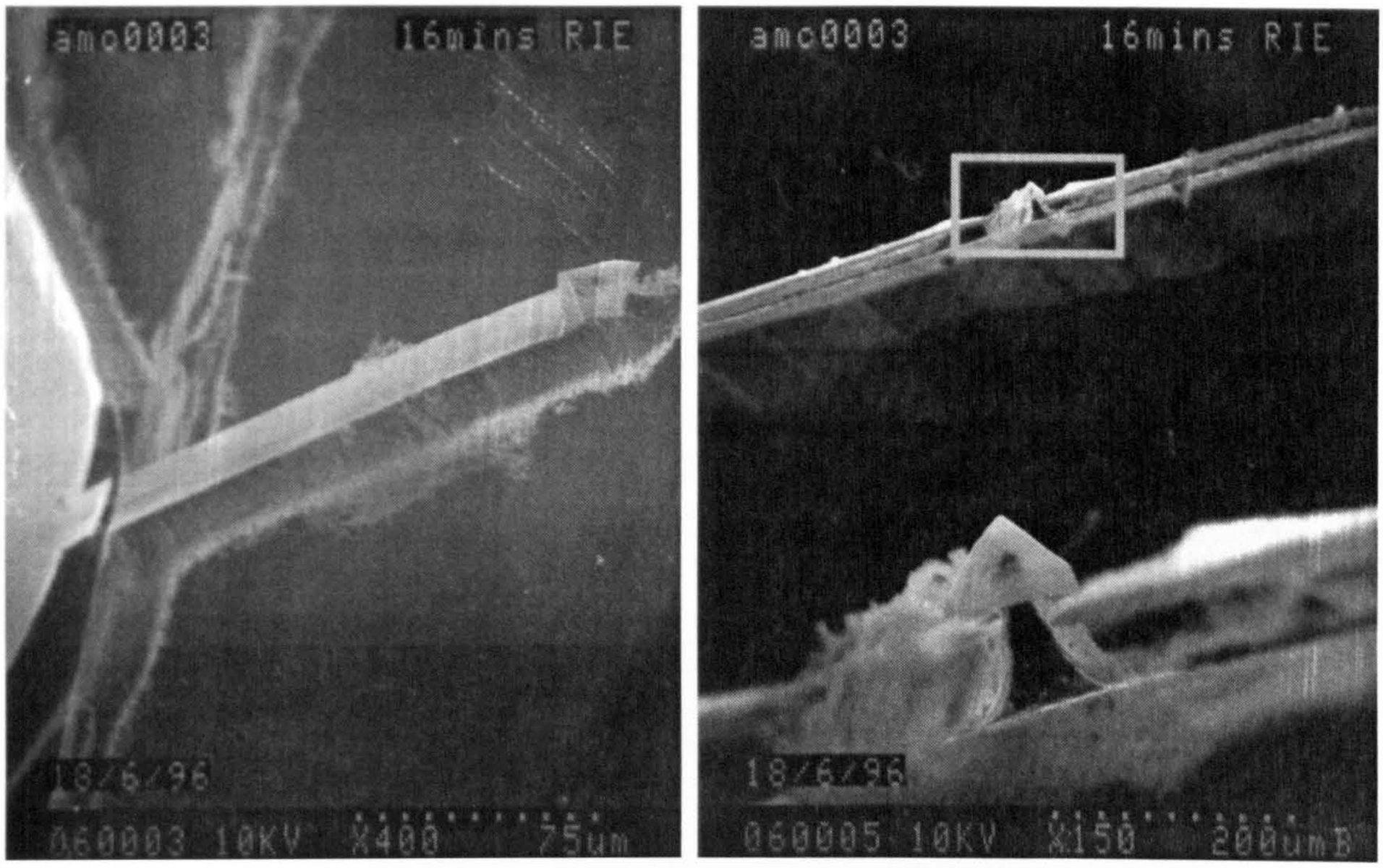


Figure L.5: Micrographs showing an attempted dry etch method for defining non-contact mode AFM cantilevers. The images are of a 200μm long, 20μm wide, silicon cantilever beam. The mask undercutting and polymer deposition seen in Figure L.4 is very much in evidence. The bottom half of the right image is a 5x magnification over the displayed scale which applies to the top half.

Much depended on the development of a suitable dry etch. An early test carried out using a test substrate meant for a different purpose consisted of 1.5μm wide openings in a 50nm thick nichrome layer on a 500nm thick PECVD silicon nitride layer and suggested that such an etch might be possible. Figure L.4 shows the result of this test carried out on the Plasma Technology RIE80 using SF_6. This etch was devised by Dr. S. K. Murad. From Figure L.4a, it is clear that the etch rate is affected by geometry with confined openings etching much more slowly than wider areas. The etch depth in the wide open area is about 15μm. Also noticeable is a significant degree of polymer deposition. This could probably be cleaned up by introducing some oxygen into the etch [Murad[2] (1993-1996)]. From Figure L.4b it can be seen that the nichrome layer is virtually untouched and still appears to be 50nm thick. However, it can also be seen that the etch results in a significant degree of mask undercutting.

The first trial with a real substrate showed the effect of these problems on the result. This trial was carried out with a 20μm wide beam and with a membrane thickness of about 20μm. Unfortunately the membrane was slightly cracked in places prior to etching. Figure L.5 shows that the etch did penetrate through the membrane. It is also clear that polymer deposition and undercutting of the mask are the two main problems that had to be solved. However, it also appeared that to control such a violent etch would require some effort and as time was short an alternative technique was sought.

This alternative was based on wet etching. The same thickness of silicon membrane as used above but with a silicon nitride mask could simply be wet etched. The advantage of this would be that the resulting cantilever beam would be half the initial thickness of the silicon membrane, as etching takes place from both sides, allowing a factor of two improvement in controllability.

The pattern used for such a beam is illustrated in Figure 4.4.4 in §4.4.3. Extra width was given around the tip region to offset the effects of convex corner undercutting during the cantilever release etch. The beam was 40μm wide, but the etch stopping {1 1 1} planes meant that the sides of the beam were not vertical, but sloped at an angle of about 54.74° to the horizontal [see §3.7.3]. This had the effect of making the cantilever stiffer than a vertical-sided beam.

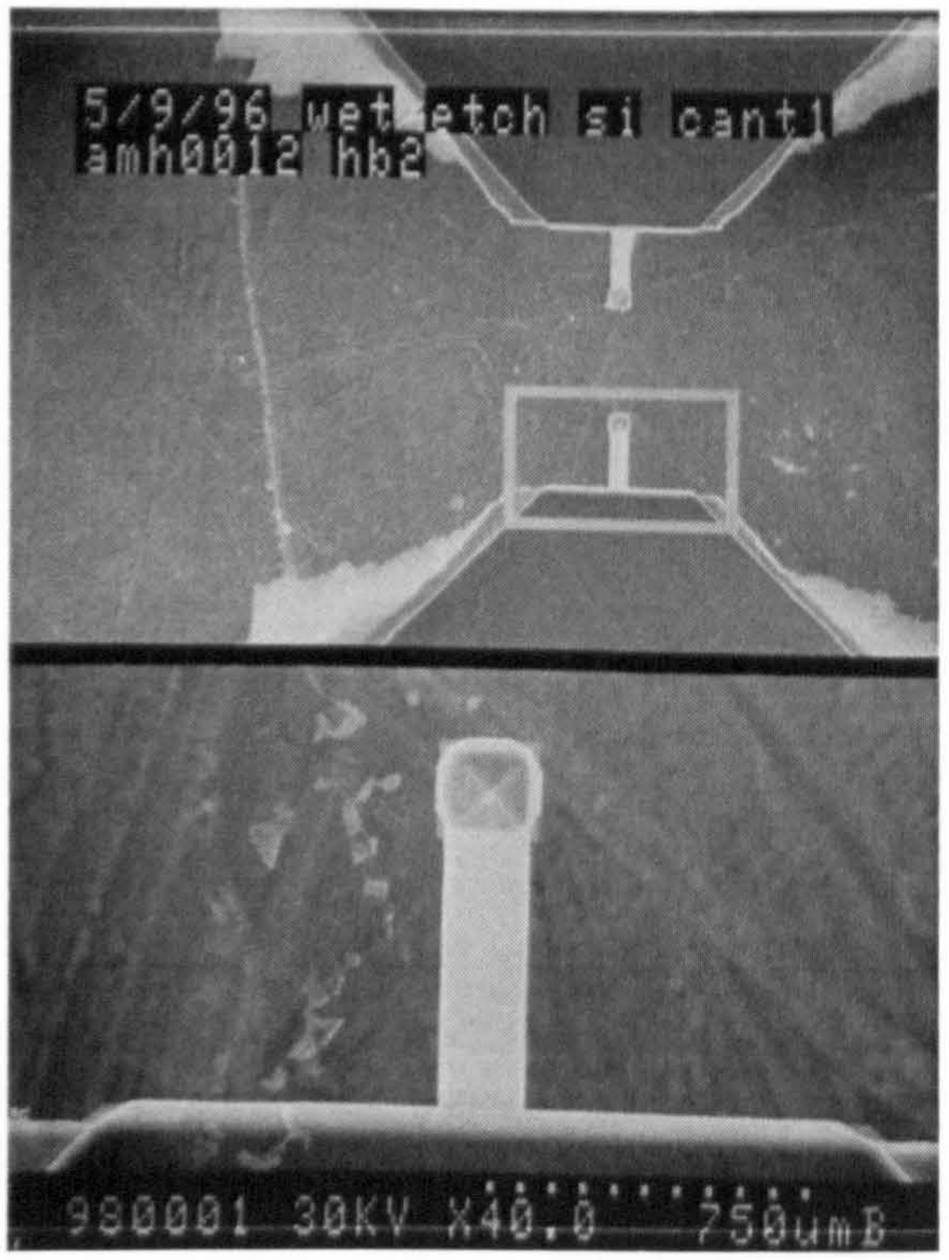


Figure L.6: Micrograph showing a wet etch defined non-contact mode AFM cantilever. The image is of a 200μm long, 40μm wide, silicon cantilever beam. The bottom image is a 5x magnification over the displayed scale which applies to the top image.

From some rudimentary estimates of the compliance of such a cantilever [see Appendix K (§K.3)], it was decided that a cantilever length of 250μm and thickness of 5μm, corresponding to a compliance of about 16N/m and a resonant frequency of about

94kHz, be aimed for initially with a view to increasing the *length to 350μm*, corresponding to a compliance of around 6N/m and a resonant frequency of about 50kHz, in the long term. Since a membrane of down to 10μm in thickness can be produced relatively easily using the extra wet etching step detailed above, a cantilever beam of 5μm in thickness did not seem unreasonable. In any case, the same trick as used for contact mode cantilevers could also be applied here, *i.e.* knowing the thickness of the cantilever membrane prior to release etching and the time taken for the etch to penetrate the substrate by etching half way through this membrane (as indicated by large areas of the substrate becoming transparent) the etch rate could be calibrated in-situ. Then the etch would simply be allowed to continue until the beam had reached the required thickness.

The fabrication scheme therefore becomes very similar to that used for contact mode AFM cantilevers. Assuming that the cantilever membrane has been thinned to an appropriate level by the additional silicon wet etch prior to tip definition, 500nm of PECVD silicon nitride is deposited on the substrate as an isolation layer (this is thicker than required but has the advantage of being the same as for contact mode AFM substrates meaning that deposition can be carried out on both sets of wafers together). Again in the interest of process reusablity between the contact and non-contact mode AFM cantilevers, the same 100nm of sputtered chromium is patterned as before and used to mask the dry etching of the silicon nitride. Then the substrate undergoes the same release etch as used for contact mode AFM cantilevers.

The first attempt to implement this made use of 5μm thick cantilever membranes and successfully resulted in the fabrication of cantilevers approximately 2.5μm thick as shown in Figure L.6. This thickness was attempted to test the degree of control possible. The conclusion was that 5μm thick cantilevers would be readily achievable. As with contact mode cantilevers, the pattern used to etch the chromium also had a 45° taper and shoulders and are recessed into the substrate by 5μm [see §L.1]. Thus, it was possible to devise two similar processes to result in different kinds of cantilever for contact and non-contact mode AFM applications. The detailed processing steps for this level are listed in Appendix G (§G.3).

L.3 REFERENCES

1. D. Gourlay, *Private Communication*, (1993-1996).
2. S. K. Murad, *Private Communication*, (1993-1996).

APPENDIX M

FINAL SUBSTRATE DESIGN INFORMATION

AMDEV7.GDS — Cell Organisation

BASIC STUFF

coarse	coarse photolithography alignment structure
align	fine photolithography alignment structure
3inwafer	outline of a 3" wafer
3ingrat10	20μm period 1:1 grating used to rotationally align to wafer flats
thru380	through-wafer alignment markers for 1st level e-beam
mask7.thru380	set of three "thru380"s used in mask pattern
bwalign1	topo markers written during 1st level e-beam
bound380	device boundary and topo marker positions (includes 2 x "bwalign1")
dev380	single device for 380μm thick wafer (includes "bound380", "dev380.back")
pair380	pair of devices for 380μm thick wafer (includes 2 x "dev380")
313pyr4	
313pyr5	
...	
313pyr18	313 orientation squares for pyramids (4 to 18μm)
313lines4	
313lines5	
...	
313lines12	313 orientation gratings for pyramid calibration (4 to 12μm)

MASK & FIRST LEVEL E-BEAM STUFF

pyrcal7.380	cell containing "313lines7"..."313lines12" for pyramid calibration
bwcross1	cross written during 1st level e-beam
3incross1	through-wafer cross or "window" (includes "thru380")
logo	Glasgow University logo in squares (layer 0 - usually off)
logo1	Glasgow University logo in stripes (layer 0 - usually off)
0.9logo1	90% version of "logo1" (includes "logo1")
logo2	"Nanoelectronics Research Centre" in stripes (layer 0 - usually off)
logo3	"University of Glasgow" in stripes (layer 0 - usually off)
textlogo1	Vertically arranged "University of Glasgow" in stripes (layer 0 - usually off)
kmount1	grooves for kinematic mounts on back of device (layer 0 - usually off)
dev380.back	back of a device with logo and kinematic mount (includes "textlogo1", "kmount1")
10uprow380	a row of ten devices facing 'up' (includes 10 x "dev380", 2 x "align")
10row380	a row of ten pair of devices facing each other (includes 2 x "10uprow380")
mask7.align	positioning of global topo markers and e-beam crosses for 1/4 wafer (includes 2 x 2 x "bwalign1", 2 x 10 x "bwcross1")
weaken	extra length of cleave line used to make device cleaving easier within 1/4 wafer
mask7.380	1/4 wafer cell containing 10 x 6 devices, logos, alignment and cleave lines (includes 3 x "10row380", 2 x "logo2"+ "logo3", "mask7.align", 20 x "weaken")
3inholder	beamwriter clamps and height reference positions
3inwindow	3" wafer border for very coarse chrome plate alignment
msk7.3in.back	3" wafer cleave lines (includes "3inholder", 8 x "3incross1", 2 x "pyrcal7.380", 8 x "mask7.thru380")
msk7.3in.base	2 x 2 x 6 x 10 devices on a 3" wafer (includes "msk7.3in.back", 4 x "mask7.380")
msk7a.3in.380	1st level photolith **ferric** mask = **flipped** version of "msk7b.3in.380"
msk7b.3in.380	2nd level photolith **chrome** mask (includes "msk7.3in.back", "3inwindow", 4 x "mask7.380", "3ingrat10", 8 x "coarse", 8 x "logo1")

MISCELLANEOUS MEMBRANE STUFF

mem100.380	openings used to form 100μm membranes for 380μm wafers
mem150.380	openings used to form 150μm membranes for 380μm wafers
memap380	a single device for forming membranes to be used in aperture tests (includes 8 x "mem150.380" - this cell is not used in the mask)
memtem380	a single device for forming membranes to be used as TEM substrates (includes "mem100.380")
mini7.align	positioning of global topo and align-through markers for corner piece (includes 8 x "align", 4 x "mask7.thru380")
memap7.380	corner piece cell containing 4 x 2 aperture test membrane devices (includes "mini7.align", 4 x 2 x "memap380", 3 x 5 x "bwalign1")
memtem7.380	corner piece cell containing 5 x 5 TEM substrate devices (includes "mini7.align", 5 x 5 x "memtem380", 6 x 6 x "bwalign1")

OTHER E-BEAM LEVEL STUFF

313rpcant100	
313rpcant150	
...	
313rpcant350	313 edged contact-mode cantilevers 100 to 350μm long
313rpdev100	
313rpdev150	
...	pairs of devices with contact-mode cantilevers 100 to 350μm long
313rpdev350	(includes "pair380", 2 x "313rpcant100"..."313rpcant350")
313rp10dev100	
313rp10dev150	pairs of devices with contact-mode cantilevers 100 to 350μm long and
...	10μm squares for pyramids
313rp10dev350	(includes 2 x "313pyr10", "313rpdev100"..."313rpdev350")
111atcant100	
111atcant150	
...	
111atcant350	111 edged non-contact-mode cantilevers 100 to 350μm long
111atdev100	
111atdev150	
...	pairs of devices with non-contact-mode cantilevers 100 to 350μm long
111atdev350	(includes "pair380", 2 x "111atcant100"..."111atcant350")
111at10dev100	
111at10dev150	pairs of devices with non-contact-mode cantilevers 100 to 350μm long
...	and 10μm squares for pyramids
111at10dev350	(includes 2 x "313pyr10", "111atdev100"..."111atdev350")
rpdevs100	
rpdevs150	1/4 wafer of 10 x 6 contact-mode devices with cantilevers 100 to 350μm
...	long
rpdevs350	(includes "mask7.align", 10 x 3 x "313rp10dev100"..."313rp10dev350")
3inrpdevs100	
3inrpdevs150	3" wafer of 2 x 2 x 60 contact-mode devices with cantilevers 100 to 350
...	μm long
3inrpdevs350	(includes "msk7.3in.back", 2 x 2 x "rpdevs100"..."rpdevs350")
atdevs100	
atdevs150	1/4 wafer of 10 x 6 non-contact-mode devices with cantilevers 100 to
...	350μm long
atdevs350	(includes "mask7.align", 10 x 3 x "111at10dev100"..."111at10dev350")
3inatdevs100	
3inatdevs150	3" wafer of 2 x 2 x 60 non-contact-mode devices with cantilevers 100 to
...	350 μm long
3inatdevs350	(includes "msk7.3in.back", 2 x 2 x "atdevs100"..."atdevs350")

DEVICE-SPECIFIC PATTERNS

Your device-specific pattern needs to be in 3 kinds of cells:

(i) the pair version (local alignment) should contain one of the "313rp10dev???" or "111at10dev???" cells

(ii) the 1/4 wafer version (global alignment) should contain a 'purged' version (just one level) of one of the "atdevs???' or "rpdevs???" cells with the pair cells 'recelled' to your pair version in (i)

(iii) the 3" wafer version (3" alignment" should contain a 'purged' version (just one level) of "**msk7.3in.base**" (not "msk7.3in.back") with the "mask7.380"s 'recelled' to your 1/4 wafer version in (ii)

NB: Always rename your cells differently and use different file names for each kind of device.

AFM 3" Wafer: AMDEV7 mask: EBL registration

Figure 1 shows the clamping arrangements for the beamwriter (polished side view). The large flat must always be on the left with regard to pattern writing (i.e. the x-axis runs from the large flat to the small flat). The pattern should take account of the space lost due to holder clamps and height references. The eight windows (W1–W8) each contain four start points (S1–S4) as shown in figure 3. These start points are the equivalent of "crosses" for the purpose of executing jobs, although in a worst case scenario, the "crosses" may be used as alignment markers as they are of the correct size (50-60μm).

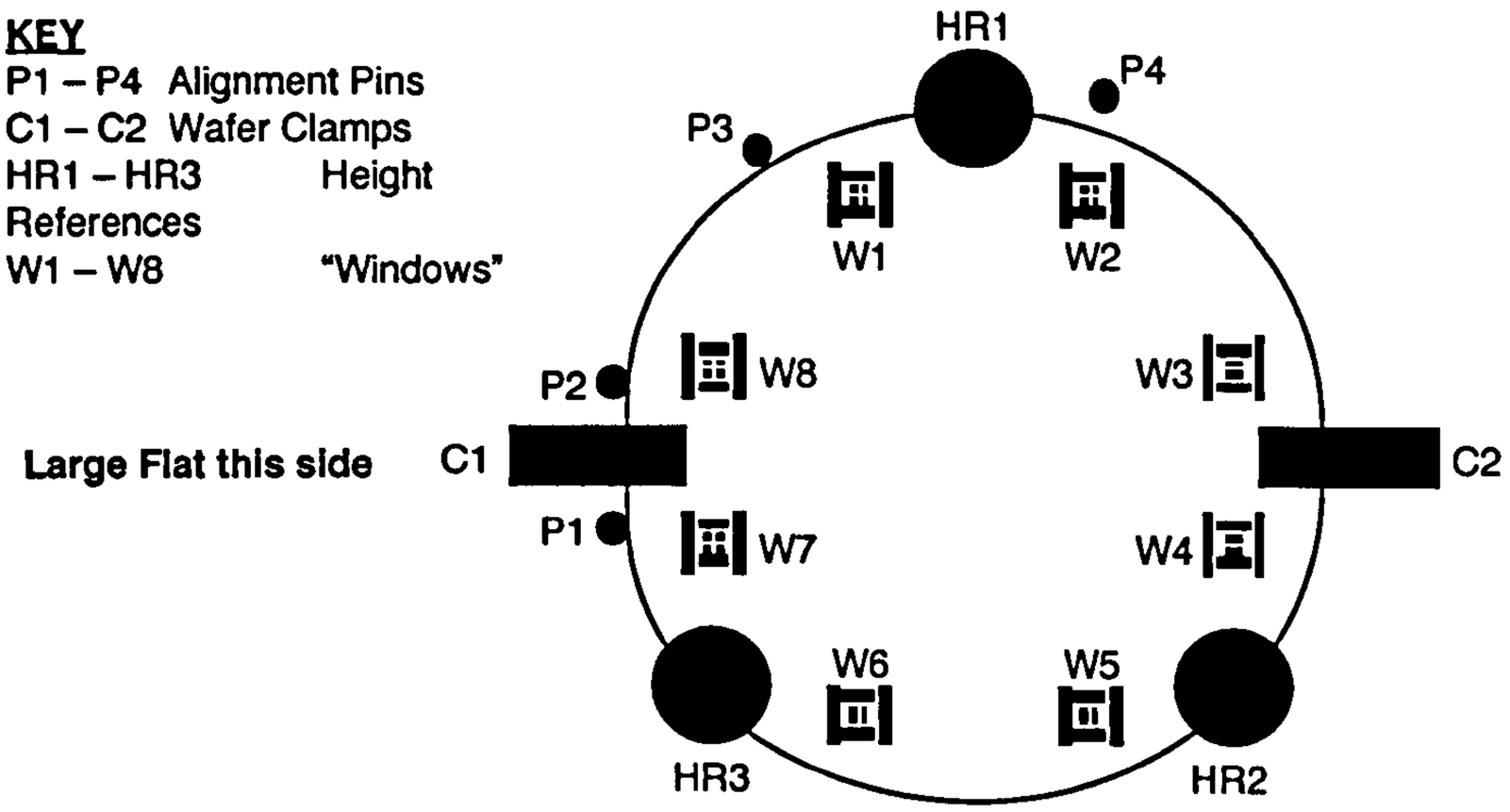


Figure 1: Wafer Loading - Note that alignment pin P4 is not in contact with the edge of the wafer.

Figure 2 shows the arrangement of the actual alignment markers. There are 8 sets of 3 markers in total, 4 sets (ABCD) for the main pattern and 4 sets (EFGH) for a pyramid calibration pattern (if desired). The markers should be between 50 and 60μm square. Taking account of the fact that markers must be in a rectangular arrangement, the three markers per set allows 7 possible choices of markers per pattern.

Figure 3 shows a close-up of a window with the four start point positions (S1–S4). When filling out the **AFM job sheet** the information required is the distance from the centre of the wafer to the window being used, the corner of the window to move to, and the distance from the window corner to the starting square being used; the remaining operator information for finding the start point is included on the job sheet. In addition a wait (to allow the holder to settle after it has been loaded) **should** be specified on the job sheet.

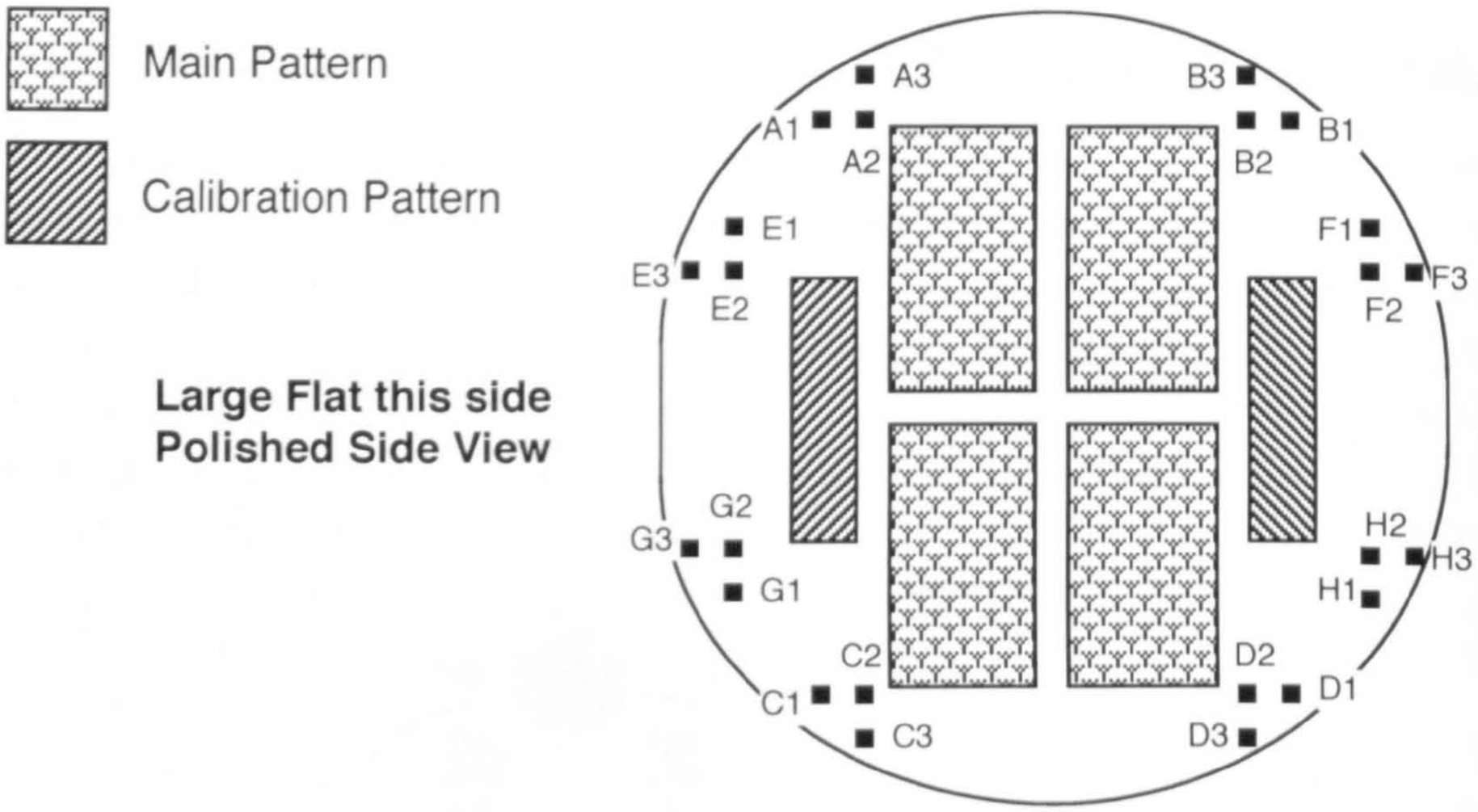


Figure 2: Marker Arrangements - Note that there are two sets (ABCD) and (EFGH).

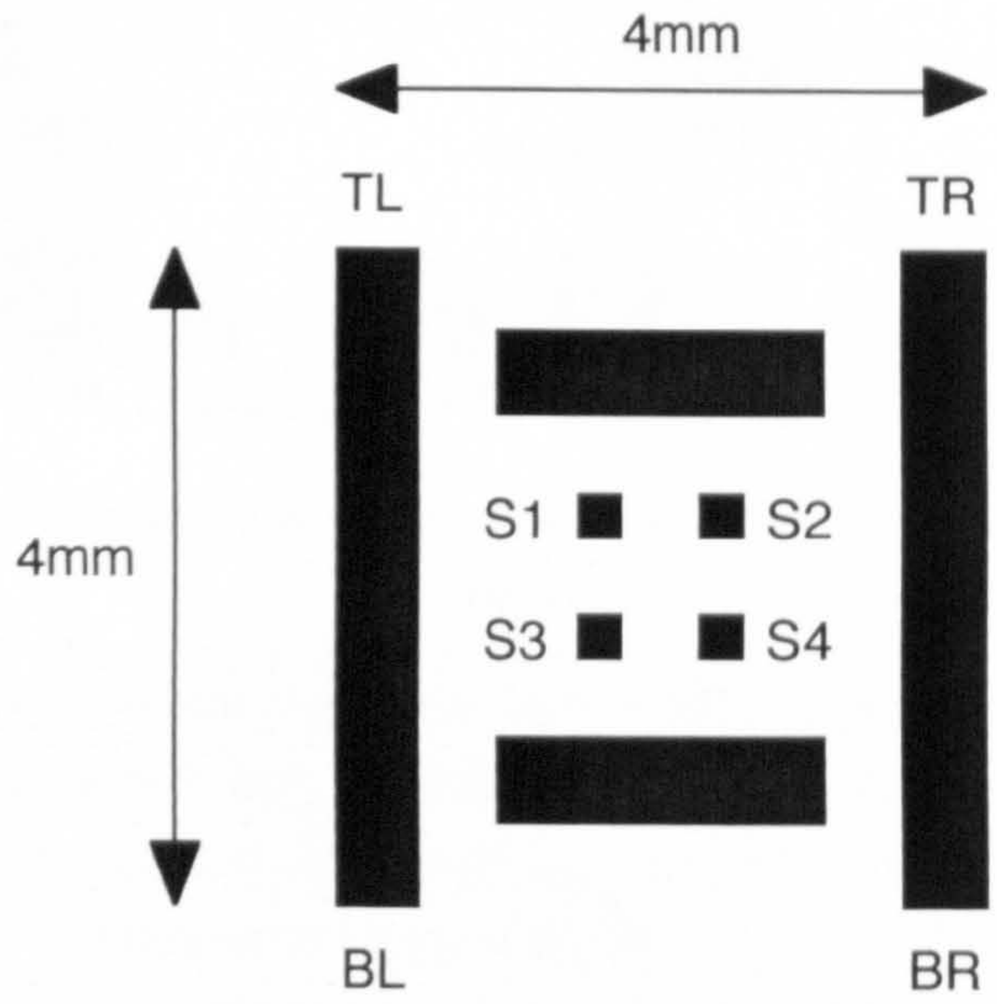


Figure 3: Window Close-Up - There are four 60μm square holes (S1 – S4). The squares (like the markers) are separated from each other by 650μm in x and y.

For the first e-beam level there are a number of pattern files:

attractive mode devices: marker outlines and pyramid definition level:-

AM3INATDEVS100_07	3 inch wafer, attractive mode device, 100μm long cantilever, layer 7
AM3INATDEVS150_07	3 inch wafer, attractive mode device, 150μm long cantilever, layer 7
AM3INATDEVS200_07	3 inch wafer, attractive mode device, 200μm long cantilever, layer 7
AM3INATDEVS250_07	3 inch wafer, attractive mode device, 250μm long cantilever, layer 7
AM3INATDEVS300_07	3 inch wafer, attractive mode device, 300μm long cantilever, layer 7
AM3INATDEVS350_07	3 inch wafer, attractive mode device, 350μm long cantilever, layer 7

repulsive mode devices: marker outlines and pyramid definition level:-

AM3INRPDEVS100_07	3 inch wafer, repulsive mode device, 100μm long cantilever, layer 7
AM3INRPDEVS150_07	3 inch wafer, repulsive mode device, 150μm long cantilever, layer 7
AM3INRPDEVS200_07	3 inch wafer, repulsive mode device, 200μm long cantilever, layer 7
AM3INRPDEVS250_07	3 inch wafer, repulsive mode device, 250μm long cantilever, layer 7
AM3INRPDEVS300_07	3 inch wafer, repulsive mode device, 300μm long cantilever, layer 7
AM3INRPDEVS350_07	3 inch wafer, repulsive mode device, 350μm long cantilever, layer 7

marker fill pattern:-

AM3INMARKERS_15	3 inch wafer, marker fill pattern, layer 15

pyramid calibration pattern:-

AM3INPYRCAL_18	3 inch wafer, pyramid calibration pattern

For the job, one of the attractive mode cantilever/repulsive mode device patterns, the marker fill pattern and optionally the pyramid calibration pattern is required giving 2 or optionally 3 matrices. The first of these is the most critical as it determines the accuracy of the marker positions with respect to the pyramids and is therefore a quick writing pattern (to minimise drift effects) utilising a 112nm spot (0.05μm resolution) with a dose of 300 to 350μC/square cm. The second pattern fills out the large border area of the markers and writes the crack arresters. There is a 2μm overlap for the marker borders and the marker outlines to accommodate drift. This pattern uses a 400nm spot (0.2μm resolution) with a dose of 300 to 350μC/square cm. The final pattern, if used, consists of gratings of different widths (7μm to 12μm) aligned with the <313> directions but with the same fill factor, which may be used as an in-situ indicator that the pyramid etch has been completed (should use 10nm NiCr + 20nm Au + 100nm Ti metallisation, and etch without removing the metal, for extra optical reflectance). This is also written with a 400nm spot (0.2μm resolution) and with a dose between 200 and 350μC/square cm (as yet this technique has not been tested or calibrated, the dose is therefore not known).

NB: AM3INPYRCAL_18.IWFL takes up 30MB of disk space and requires 1-2hrs to fractionate.

Pyramid/Topo Marker Definition Level

1) Decide which window (W1–W8), which edge (BL, BR, TL, TR) and which start position (S1–S4) is to be used - the appropriate distances for filling out the job sheet are given in Tables 1 and 2 below. Also decide the length of wait (to allow the holder to settle down just after it has been loaded) – this number should also be entered on the AFM job sheet. A reasonable value for this level is 15 minutes.

2) Decide which marker set is to be used for the main pattern – the choices are:-

bottom left = C1 => (A1 B1 C1 D1) or (A1 B2 C1 D2)

bottom left = C2 => (A2 B2 C2 D2) or (A2 B1 C2 D1) or (A3 B3 C2 D2)

bottom left = C3 => (A3 B3 C3 D3) or (A2 B2 C3 D3)

3) If the pyramid calibration pattern is also to be written, decide which marker set is to be used - the choices are:-

bottom left = G1 => (E1 F1 G1 H1) or (E2 F2 G1 H1)

bottom left = G2 => (E1 F1 G2 H2) or (E2 F2 G2 H2) or (E2 F3 G2 H3)

bottom left = G3 => (E3 F3 G3 H3) or (E3 F2 G3 H2)

4) Decide the marker size (add 10μm to the optical size and round to nearest 10μm), i.e. 50, 60 or 70μm. The step and repeat and centre can be found from table 3. All distances are in **microns**. Note that the centre position varies with the marker size. The distance from the "cross" or starting point (W1S1,...,W1S4;W2S1,...,W2S4;...;W8S1,...,W8S4) to the first marker (C1–C3 or G1–G3) can be found from tables 4 and 5. Again all distances are in **microns**.

Window	X-SHIFT	Y-SHIFT	Window	X-SHIFT	Y-SHIFT
W1	-10400 μm	32000 μm	W5	9600 μm	-32000 μm
W2	9600 μm	32000 μm	W6	-10400 μm	-32000 μm
W3	32100 μm	7500 μm	W7	-32900 μm	-7500 μm
W4	32100 μm	-7500 μm	W8	-32900 μm	-7500 μm

Table 1: Distances from the centre of the wafer to the centre of a window

BWL NOTES: Calibration ("calib") should always be set to "block" for all registration jobs (ignore any warnings bwl might throw up in the case of jobs where the repeats are more than "1,1"). For each matrix, (and whenever through-wafer alignment markers are used),

go to "screen extra" and set "mq_contra" to "high", "mq_stol" to "poor" and "mq_ptol" to "poor". Also set "cal_drift" to "1000". **After BWL** use eve to **insert** the following line into matrix 1 after the other ":==" parameters: "**$ bwl_mq_radius := = 120**".

Square	BL-SHIFT	TL-SHIFT	BR-SHIFT	TR-SHIFT
S1	(1675, 2550) µm	(1675, -1900) µm	(-2325, 2550) µm	(-2325, -1900) µm
S2	(2325, 2550) µm	(2325, -1900) µm	(-1675, 2550) µm	(-1675, -1900) µm
S3	(1675, 1900) µm	(1675, -2550) µm	(-2325, 1900) µm	(-2325, -2550) µm
S4	(2325, 1900) µm	(2325, -2550) µm	(-1675, 1900) µm	(-1675, -2550) µm

Table 2: Distances from the corners of the window to the centre of a starting position

(BL = Bottom Left Corner, TL = Top Left Corner, BR = Bottom Right Corner, TR = Top Right Corner)

Marker Set	Step & Repeat		Centre (50µm)		Centre (60µm)		Centre (70µm)	
A1 B1 C1 D1	48400	55400	-24225	-27725	-24230	-27730	-24235	-27735
A1 B2 C1 D2	47750	55400	-24225	-27725	-24230	-27730	-24235	-27735
A2 B2 C2 D2	47100	55400	-23575	-27725	-23580	-27730	-23585	-27735
A2 B1 C2 D1	47750	55400	-23575	-27725	-23580	-27730	-23585	-27735
A2 B2 C3 D3	47100	56050	-23575	-28375	-23580	-28380	-23585	-28385
A3 B3 C3 D3	47100	56700	-23575	-28375	-23580	-28380	-23585	-28385
A3 B3 C2 D2	47100	56050	-23575	-27725	-23580	-27730	-23585	-27735
E1 F1 G1 H1	56500	27800	-28275	-13925	-28280	-13930	-28285	-13935
E1 F1 G2 H2	56500	27150	-28275	-13275	-28280	-13280	-28285	-13285
E2 F2 G2 H2	56500	26500	-28275	-13275	-28280	-13280	-28285	-13285
E2 F3 G2 H3	57150	26500	-28275	-13275	-28280	-13280	-28285	-13285
E2 F2 G1 H1	56500	27150	-28275	-13925	-28280	-13930	-28285	-13935
E3 F3 G3 H3	57800	26500	-28925	-13275	-28930	-13280	-28935	-13285
E3 F2 G3 H2	57150	26500	-28925	-13275	-28930	-13280	-28935	-13285

Table 3: Step & Repeat and Centre position for different marker choices.

Matrix 1

Pattern:	as indicated earlier - according to length and type of cantilever required
Dose:	300μC/square cm (could be 350μC/square cm depending on resist condition)
Repeats:	1, 1
Spot Size:	112nm
(Resolution:	0.05μm)
Step & Repeat:	as from table 1 (for marker sets ABCD)
Centre:	as from table 1 (for marker sets ABCD and marker size 50/60/70μm)
Cross:	as from table 2 (for marker sets ABCD and appropriate cross)
Marker Type:	Rect
Marker Contra:	Negative
Marker Size:	50μm, 60μm or 70μm (= optical size + 10μm, rounded to nearest 10μm)

Matrix 2

Pattern:	AM3INMARKERS_15
Dose:	as matrix 1
Repeats:	as matrix 1
Spot Size:	400nm
(Resolution:	0.2μm)
Step & Repeat:	as matrix 1
Centre:	as matrix 1
Cross:	as matrix 1
Marker Type:	as matrix 1
Marker Contra:	as matrix 1
Marker Size:	as matrix 1

Optional Matrix 3

Pattern:	AM3INPYRCAL_18
Dose:	(between 200 and 350μC/square cm - to be determined by trial and error)
Repeats:	as matrix 1, 2
Spot Size:	400nm
(Resolution:	0.2μm)
Step & Repeat:	as from table 1 (for marker sets EFGH)
Centre:	as from table 1 (for marker sets EFGH and marker size as matrix 1, 2)
Cross:	as from table 2 (for marker sets EFGH and same cross as matrix 1, 2)
Marker Type:	as matrix 1, 2
Marker Contra:	as matrix 1, 2
Marker Size:	as matrix 1, 2

Cross	Cross->C1		Cross->C2		Cross->C3	
W1 S1	-13875	-60025	-13225	-60025	-13225	-60675
W1 S2	-14525	-60025	-13875	-60025	-13875	-60675
W1 S3	-13875	-59375	-13225	-59375	-13225	-60025
W1 S4	-14525	-59375	-13875	-59375	-13875	-60025
W2 S1	-33875	-60025	-33225	-60025	-33225	-60675
W2 S2	-34525	-60025	-33875	-60025	-33875	-60675
W2 S3	-33875	-59375	-33225	-59375	-33225	-60025
W2 S4	-34525	-59375	-33875	-59375	-33875	-60025
W3 S1	-56375	-35525	-55725	-35525	-55725	-36175
W3 S2	-57025	-35525	-56375	-35525	-56375	-36175
W3 S3	-56375	-34875	-55725	-34875	-55725	-35525
W3 S4	-57025	-34875	-56375	-34875	-56375	-35525
W4 S1	-56375	-20525	-55725	-20525	-55725	-21175
W4 S2	-57025	-20525	-56375	-20525	-56375	-21175
W4 S3	-56375	-19875	-55725	-19875	-55725	-20525
W4 S4	-57025	-19875	-56375	-19875	-56375	-20525
W5 S1	-33875	3975	-33225	3975	-33225	3325
W5 S2	-34525	3975	-33875	3975	-33875	3325
W5 S3	-33875	4625	-33225	4625	-33225	3975
W5 S4	-34525	4625	-33875	4625	-33875	3975
W6 S1	-13875	3975	-13225	3975	-13225	3325
W6 S2	-14525	3975	-13875	3975	-13875	3325
W6 S3	-13875	4625	-13225	4625	-13225	3975
W6 S4	-14525	4625	-13875	4625	-13875	3975
W7 S1	8625	-20525	9275	-20525	9275	-21175
W7 S2	7975	-20525	8625	-20525	8625	-21175
W7 S3	8625	-19875	9275	-19875	9275	-20525
W7 S4	7975	-19875	8625	-19875	8625	-20525
W8 S1	8625	-35525	9275	-35525	9275	-36175
W8 S2	7975	-35525	8625	-35525	8625	-36175
W8 S3	8625	-34875	9275	-34875	9275	-35525
W8 S4	7975	-34875	8625	-34875	8625	-35525

Table 4: Cross distance from various crosses to markers C1, C2 and C3.

Additional Information

Pattern files AM3INRPDEVS100_07 to AM3INRPDEVS350_07 are generated from AMDEV7.GDS (AMDEV7_GDS.CLIB). The cells (structs) are 3INRPDEVS100 to 3INRPDEVS350 using only layer 7. The extent (extent all) is (-50000, -50000), (50000, 50000) and the resolution is 0.05μm. Default block size and format are used with no overlaps.

Pattern files AM3INATDEVS100_07 to AM3INATDEVS350_07 are generated from AMDEV7.GDS (AMDEV7_GDS.CLIB). The cells (structs) are 3INATDEVS100 to 3INATDEVS350 using only layer 7. The extent (extent all) is (-50000, -50000), (50000, 50000) and the resolution is 0.05μm. Default block size and format are used with no overlaps.

Pattern file AM3INMARKERS_15 is generated from AMDEV7.GDS (AMDEV7_GDS.CLIB). The cell (struct) is MSK7B.3IN.380 using only layer 15. The extent (extent all) is (-50000, -50000), (50000, 50000) and the resolution is 0.2μm. Default block size and format are used with no overlaps.

Pattern file AM3INPYRCAL_18 is generated from AMDEV7.GDS (AMDEV7_GDS.CLIB). The cell (struct) is MSK7B.3IN.380 using only layer 18. The extent (extent all) is (-50000, -50000), (50000, 50000) and the resolution is 0.2μm. Default block size and format are used with no overlaps.

Cross	Cross->G1		Cross->G2		Cross->G3	
W1 S1	-17925	-46225	-17275	-45575	-18575	-45575
W1 S2	-18575	-46225	-18575	-45575	-19225	-45575
W1 S3	-17925	-45575	-17925	-44925	-18575	-44925
W1 S4	-18575	-45575	-18575	-44925	-19225	-44925
W2 S1	-37925	-46225	-37275	-45575	-38575	-45575
W2 S2	-38575	-46225	-38575	-45575	-39225	-45575
W2 S3	-37925	-45575	-37925	-44925	-38575	-44925
W2 S4	-38575	-45575	-38575	-44925	-39225	-44925
W3 S1	-60425	-21725	-59775	-21075	-61075	-21075
W3 S2	-61075	-21725	-61075	-21075	-61725	-21075
W3 S3	-60425	-21075	-60425	-20425	-61075	-20425
W3 S4	-61075	-21075	-61075	-20425	-61725	-20425
W4 S1	-60425	-6725	-59775	-6075	-61075	-6075
W4 S2	-61075	-6725	-61075	-6075	-61725	-6075
W4 S3	-60425	-6075	-60425	-5425	-61075	-5425
W4 S4	-61075	-6075	-61075	-5425	-61725	-5425
W5 S1	-37925	17775	-37275	18425	-38575	18425
W5 S2	-38575	17775	-38575	18425	-39225	18425
W5 S3	-37925	18425	-37925	19075	-38575	19075
W5 S4	-38575	18425	-38575	19075	-39225	19075
W6 S1	-17925	17775	-17275	18425	-18575	18425
W6 S2	-18575	17775	-18575	18425	-19225	18425
W6 S3	-17925	18425	-17925	19075	-18575	19075
W6 S4	-18575	18425	-18575	19075	-19225	19075
W7 S1	4575	-6725	5225	-6075	3925	-6075
W7 S2	3925	-6725	3925	-6075	3275	-6075
W7 S3	4575	-6075	4575	-5425	3925	-5425
W7 S4	3925	-6075	3925	-5425	3275	-5425
W8 S1	4575	-21725	5225	-21075	3925	-21075
W8 S2	3925	-21725	3925	-21075	3275	-21075
W8 S3	4575	-21075	4575	-20425	3925	-20425
W8 S4	3925	-21075	3925	-20425	3275	-20425

Table 5: Cross distance from various crosses to markers G1, G2 and G3.

AFM 3" Wafer Jobs – Operator Instructions

Loading

Refer to Figure 1 below. Align the wafer in the holder with the **large flat** against alignment pins P1, P2 and P3. Note that alignment pin **P4 will not be in contact** with the wafer edge. Check the wafer rotation by aligning the square "windows" (W1-W8) vertically and horizontally with the axes of the holder.

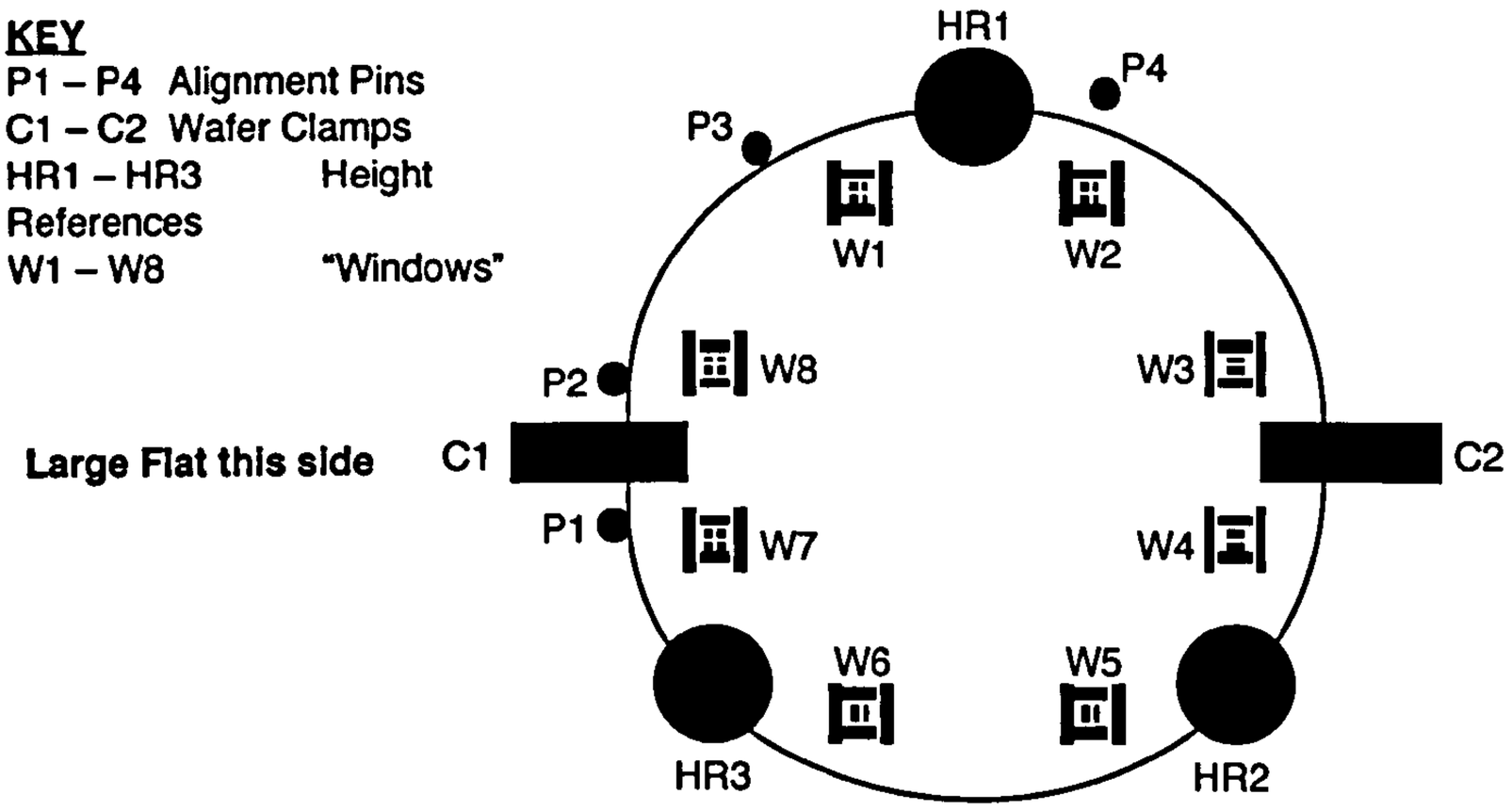


Figure 1: Wafer Loading - Notice that alignment pin P4 is not in contact with the edge of wafer.

Instructions on finding the start point for the job are overleaf.

Finding the Start Point

Refer to Figure 2 below. This is a two stage procedure. In the first stage you will have to find the top or bottom of one of the vertical "window" edges (BL, TL, BR, TR). In the second stage you will move a fixed distance to one of the 60μm square openings (S1 – S4). This will be the starting point for the job.

(i) Select the 20nm spot.

(ii) Move to the centre of the wafer using the "sub_centre" command.

(iii) Move to the centre of the window specified on the job sheet (W1 – W8) using the "mpos/rel" command with the figures given in the table below in **microns.**

Window	X-SHIFT	Y-SHIFT	Window	X-SHIFT	Y-SHIFT
W1	-10400 μm	32000 μm	W5	9600 μm	-32000 μm
W2	9600 μm	32000 μm	W6	-10400 μm	-32000 μm
W3	32100 μm	7500 μm	W7	-32900 μm	-7500 μm
W4	32100 μm	-7500 μm	W8	-32900 μm	-7500 μm

(iv) Turn the monitor on.

(v) If the job sheet specifies that the top left (TL) or the bottom left (BL) of the "window" is to be used, move leftwards in steps of 250μm until you see the black line that marks the "window" edge (it is over 150μm wide).

If the job sheet specifies that the top right (TR) or bottom right (BR) of the "window" is to be used, move rightwards in steps of 250μm until you see the black line that marks the "window" edge (it is over 150μm wide).

(vi) If the job sheet specifies that the top left (TL) or the top right (TR) of the "window" is to be used, move upwards in steps of 250μm until you hit the top edge of the frame. Then centre on this edge.

If the job sheet specifies that the bottom left (BL) or the bottom right (BR) of the "window" is to be used, move downwards in steps of 250μm until you hit the bottom edge of the frame. Then centre on this edge.

(vii) Now move to the centre of the square specified on the job sheet (S1 – S4) using the "mpos/rel" command with the figures given in the table below in **microns.**

Square	BL-SHIFT	TL-SHIFT	BR-SHIFT	TR-SHIFT
S1	(1675, 2550) μm	(1675, -1900) μm	(-2325, 2550) μm	(-2325, -1900) μm
S2	(2325, 2550) μm	(2325, -1900) μm	(-1675, 2550) μm	(-1675, -1900) μm
S3	(1675, 1900) μm	(1675, -2550) μm	(-2325, 1900) μm	(-2325, -2550) μm
S4	(2325, 1900) μm	(2325, -2550) μm	(-1675, 1900) μm	(-1675, -2550) μm

(viii) This is the starting point for the job. The job can now be tested as usual and added to the queue, but please insert a **15 minute wait** ($ wait 00:15:00) before this job to reduce drift effects during writing.

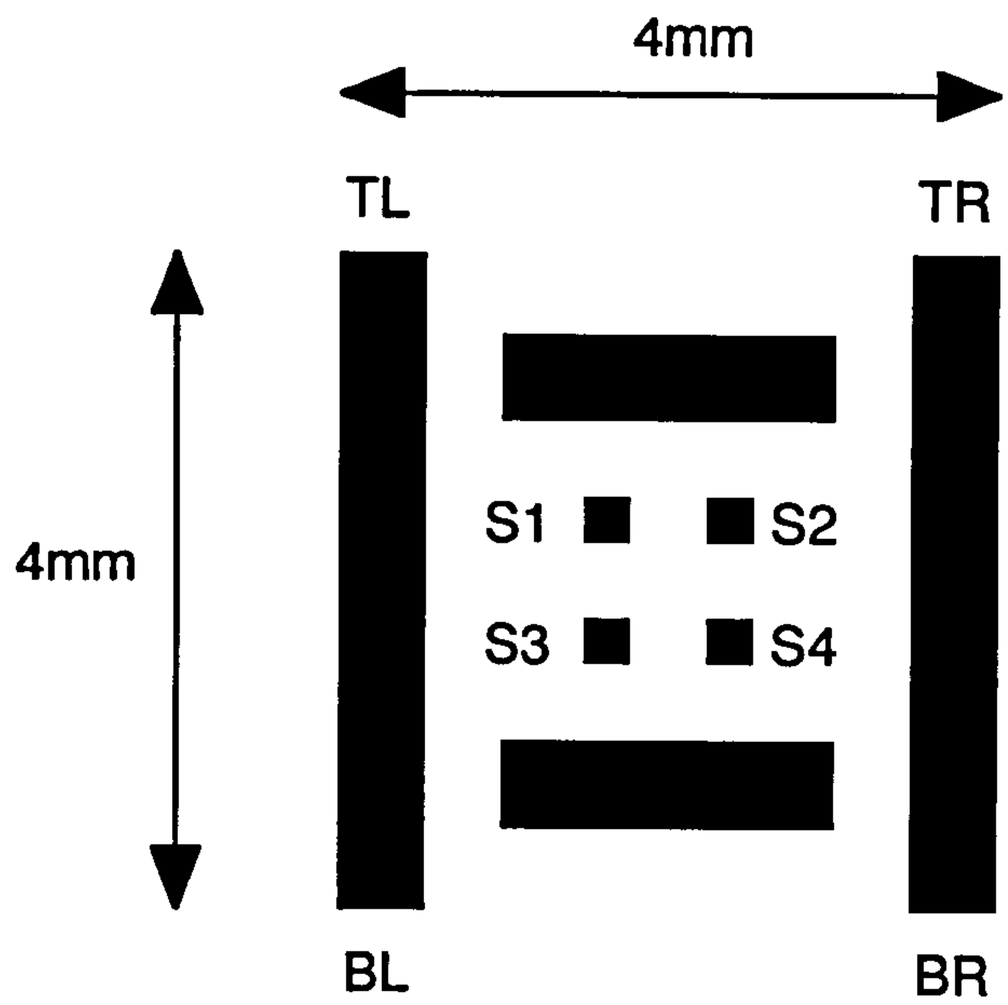


Figure 2: Finding the Cross - There are four 60μm square holes (S1 – S4). The squares are separated from each other by 650μm in x and y. The "window" edges are over 150μm wide.

AFM Beamwriter Job Submission Form

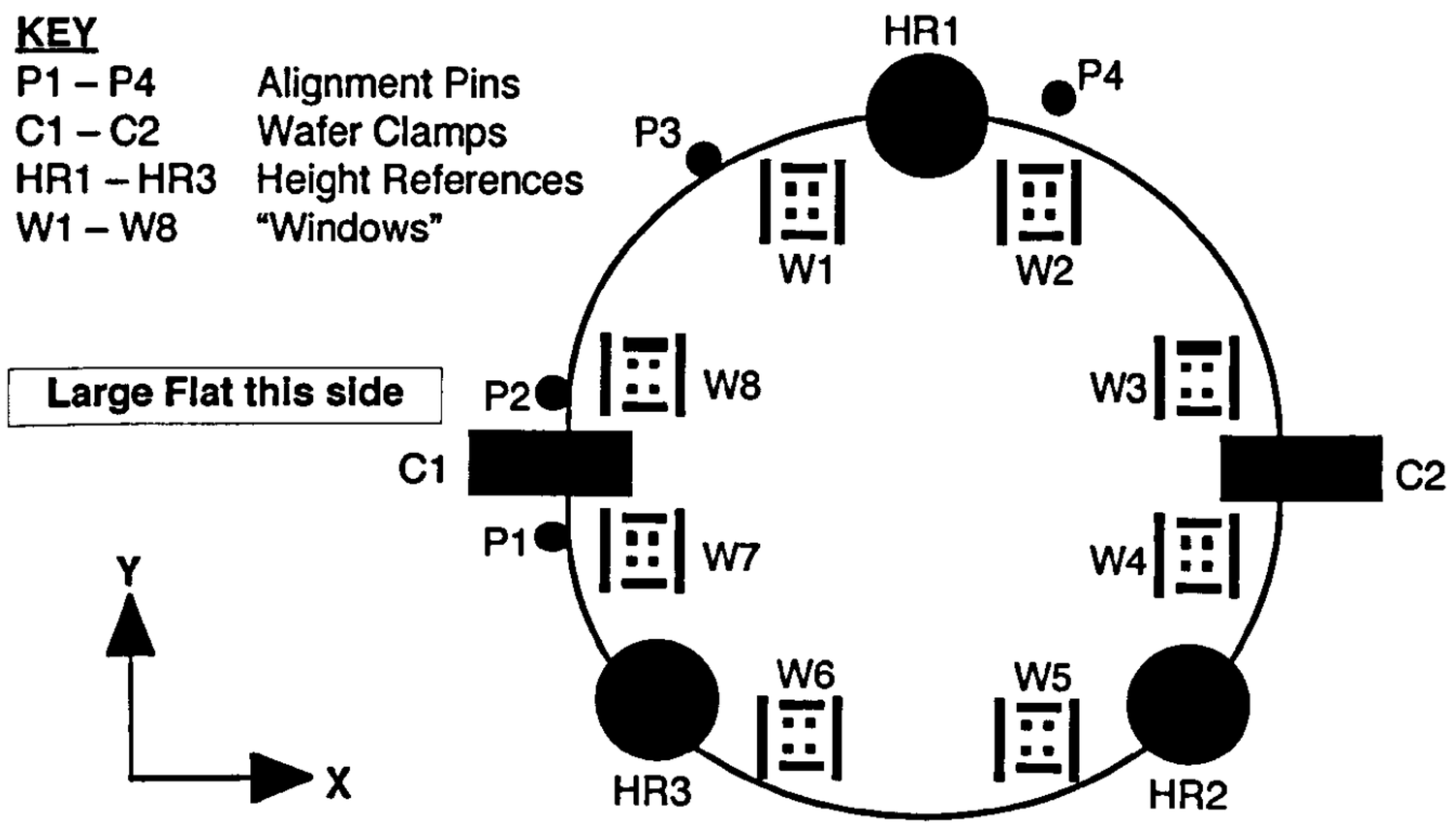
Name				Beam Energy 50kV ☐	100kV ☐
Date		Extn		Tick if job submitted before ☐	

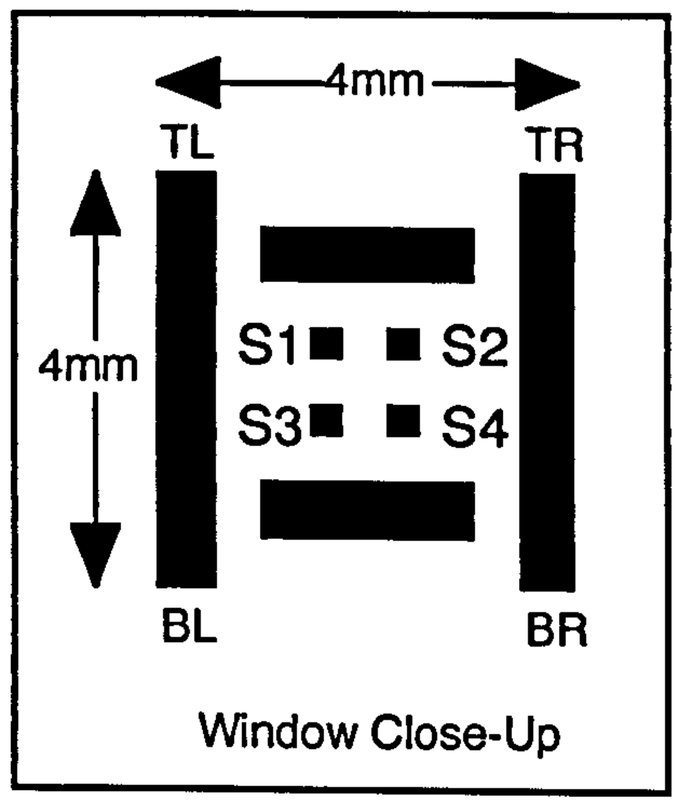
Job Name [: | | | | | | |] (Two / three initials & four digits, e.g. JK0124)

Spot Size		Job time (hours)	
Job type ☐ Registration ☐ Reg/Exp test		Priority	Low ☐ Medium ☐ Urgent ☐
Substrate	**3" Wafer**	Marker size (μm)	
Thickness (μm)	**380 μm**	Step & repeat distance (μm)	

Substrate Description

(Please circle the window being used, the bar edge and the starting position on the diagrams below)





Wafer Clamping - Note that alignment pin P4 is **NOT** in contact with the edge of the wafer. Adjust the wafer rotation by lining up the "windows" along the holder's x and y axes.	There are 4 starting points (**60μm** squares) per window, (S1-S4) separated from each other by **650μm** in x and y.

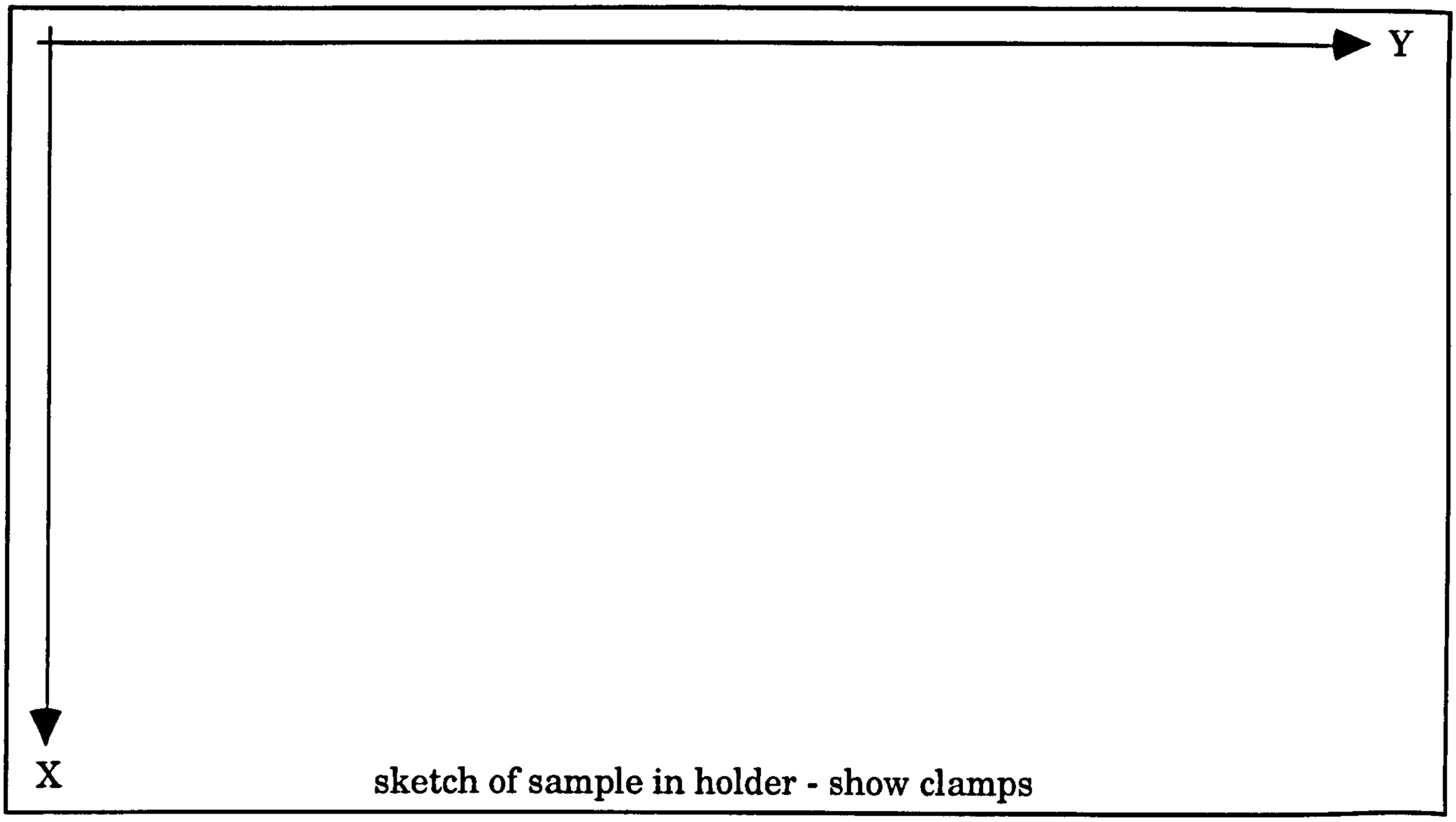
Finding the starting position:

1) getbeam 20
2) mps ff 0
3) mpos 'sub_centre'
4) mpos/rel [(,)* μm] **insert "W" distance*
5) switch monitor on
6) move [left / right*] until you reach a vertical black bar **delete as appropriate*
7) move [up / down*] until you reach the end of the bar **delete as appropriate*
8) mpos/rel [(,)* μm] **insert "S" distance*
9) **the centre of the 60μm square is the starting point: test the job as usual**

NOTE Please insert a _____ * minute wait in the batch file (after loading the holder, but prior to running the job) to minimise holder drift during writing i.e. $ wait HH:MM:SS **insert value*

Please leave this side clear for operator use

Operator		Bottom Left Corner	**x**	**y**
Date				
Holder		First Alignment mark	**x**	**y**
Position				



sketch of sample in holder - show clamps

Successful Yes ☐ No ☐	**date**	

Reasons for failure / suggestions	attempted?