TECHNIQUES AND APPLICATIONS OF VERY HIGH RESOLUTION ELECTRON-BEAM LITHOGRAPHY

A Thesis submitted to
the Faculty of Engineering of the
University of Glasgow for the degree of
Doctor of Philosophy

by

William Stuart Mackie

April 1984

ACKNOWLEDGEMENTS

Throughout the duration of the project, the advice and assistance of many people was appreciated.

Thanks are due to Professor C.D.W.Wilkinson and Dr. S.P.Beaumont for invaluable guidance and support throughout the term of work.

Appreciation is also expressed to the Science and Engineering Research Council and Plessey (Caswell) Ltd. for financial support. Also the many staff at Caswell who gave extremely valuable assistance during the portion of the work carried out there. These include D.Bramley, R.Butlin, N.Forbes, W.Jones, B.Wilson, C.Oxley and many others.

Thanks are also due to Professor J.Lamb for the use of laboratory and computing facilities in the University, and the assistance of the technical and computing staff is gratefully acknowledged.

Special thanks go to J.Crichton for his hard work at the SEM, performing many of the electron-beam exposures required for this project.

Finally, appreciation is expressed to all my friends within the Deprtment who helped to make the three years of study a most pleasant time.

Summary

This thesis describes the development of techniques of very high resolution electron beam lithography for fabrication of very small electronic devices and the depletion- mode Schottky-barrier GaAs field-effect transistor in particular.

Previous work in this Department and elsewhere had demonstrated the ability to use the techniques of electron-beam exposure and liftoff to fabricate metal lines 10nm wide, and metallic gratings with a pitch of 45nm. This work had been carried out on thin carbon substrates to eliminate the effect of electrons backscattered from a solid substrate. The aim of this project was to transfer the technology to thin semiconductor membranes and fabricate devices that might exhibit interesting characteristics because of their very smallness. It was decided to try to fabricate a FET on a thin substrate of GaAs. The dimensions of the FET were to be at the limits of electron-beam lithography.

Silicon nitride membranes were used for a series of experiments designed to find the limits of resolution for PMMA electron resist. Procedures were devised that enabled consistent fabrication of isolated 10nm wide lines, and gratings with 40nm pitch. A method of alignment was devised that enabled successive patterns to be aligned to an accuracy of 5nm over a 25 micron frame.

A method of fabricating thin membranes of GaAs was developed, and fine line lithography demonstrated on such a membrane.

On solid substrates, procedures were developed to enable exposures of $1mm^2$ to be performed with minimum feature sizes of 0.7 microns. Methods of alignment were developed to allow multilevel devices to be fabricated. These techniques were used to fabricate mesa-isolated Schottky gate GaAs FETs, with gate

lengths from 1.5 microns to 0.075 microns.

Technical problems with proton isolation prevented all the above techniques that had been developed from being put together, and a thin substrate transistor from being made.

CHAPTER 1 INTRODUCTION

The purpose of this chapter is to give a brief review of the processes involved in electron beam lithography, why it is useful to use a thin substrate if very high resolution features are required, and to explain why a FET on a thin GaAs membrane was chosen as the target device for this project.

1.1 ELECTRON BEAM LITHOGRAPHY

Electron beams have been used for delineation of sub-micron features for many years [1.1,1.2]. The key feature of the process is that certain materials, mostly polymers, have their molecular strucure changed when an energetic beam of electrons strike them. When used in lithography these materials are called resists. They are deposited onto samples as a layer 20nm to 2 microns thick depending on subsequent processing requirements. The desired pattern is scanned out by the electron beam which is under computer control. Under the action of a developer, and depending on the resist, either the exposed or the unexposed areas of resist can then be dissolved out preferentially. The resist is then used to pattern the wafer by etching or to pattern a metal layer onto the wafer by etching or liftoff.

Some aspects of lithography in general, and the reasons for some of the choices for the lithography processes used in this project in particular, are discussed in the following sections.

1.2 RESISTS - POSITIVE VERSUS NEGATIVE

Organic polymeric resists fall into two classes called posistive and negative, where the dominant exposure mechanisms are due to chain scission and cross-linking respectively. The

development of both resists is similar in concept, a solvent which will only dissolve polymer below a certain molecular weight being used to selectively dissolve the exposed areas of a positive resist, or the unexposed areas of a negative resist. After development it is usually desirable to use the resist to pattern the wafer in some way, either by etching the surface or patterning a metal layer. Metal can be patterned by etching a previously deposited layer, or in the case of positive resists, by liftoff (Fig. 1.1). The procedure for liftoff involves overcoating the developed sample with metal, then dissolving the remaining resist in a strong solvent with agitation thus 'lifting off' the excess metal to leave metal lines in the exposed areas of the sample. In order to perform liftoff successfully, it is desirable for the resist profile to have an undercut so that the metal on the substrate is not connected to the metal on top of the resist, which would prevent clean liftoff. The undercut is obtained with thick resist because of forward scattering of the primary beam (Section 1.4), and with thin resist by use of a two layer system (Sections 4.1.1.1 and 6.5.3.2). In general, positive resists give better edge definition, better resolution, and better contrast than negative resists but tend to have poorer sensitivity [1.3]. On the other hand, posistive resists tend to have poor adhesion in the presence of chemical etches and to have poor resistance to ion etching.

In this work the decision to use a positive resist was fairly clear. The two main reasons for this are that metal features with the highest possible resolution were desired and were required to be formed by liftoff (Section 1.3). The decision was made more clear-cut since one resist in particular, polymethyl methacrylate (PMMA), a positive resist, offers higher resolution than any other conventional resist, positive or negative [1.4,1.5] (but see Broers [1.6] for an unconventional negative resist with slightly better resolution than PMMA). PMMA was used for all the electron beam lithography in this work.

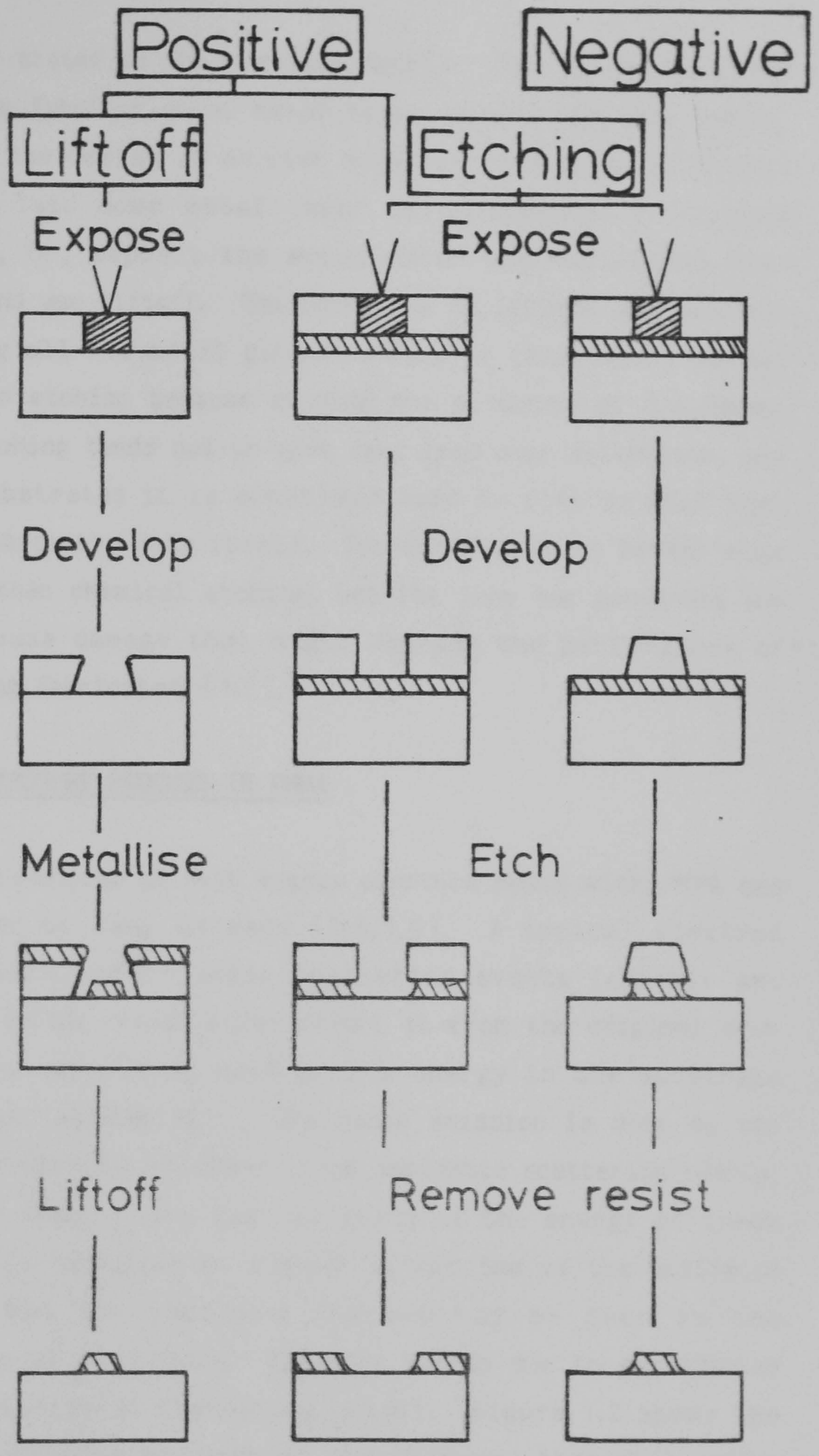


Fig. 1.1 Processing of positive and negative resists to pattern metal. Note non-vertical metal edges, due, in the case of liftoff, to lateral growth of the metal layer from the resist during evaporation, and, in the case of etching, to undercutting of the resist by the etch.

1.3 LIFTOFF VERSUS ETCHING

As was stated in the previous section there are two options open for the fabrication of metal lines using a positive resist; either use the resist as an etch mask and etch away areas on a previously laid down metal layer using chemical or plasma techniques, or, deposit the metal after the resist has been developed and use liftoff. The technique of liftoff was used for fabricating all the metal patterns used in this work. It was preferred to etching because etching has a number of drawbacks. Chemical etching tends not to give very good edge definition, and for GaAs substrates it is sometimes hard to find an etch that will not attack the GaAs itself. Ion etching gives better edge definition than chemical etching, but the ions can penetrate the GaAs and cause damage that might degrade the performance of devices being fabricated [1.7].

1.4 THE EXPOSURE PROCESS IN PMMA

The interaction of high energy electron beams with PMMA has been studied by many workers [1.8,1.9]. A typical electron suffers several small angle scattering events (elastic and inelastic) in the resist which divert it from the original beam path, before depositing most of its energy in the substrate during further scattering. The chain scission is done by the secondary electrons resulting from inelastic scattering events, as far as is known. The vast majority of the energy of these secondaries is deposited on average within 5nm of the collision point, so the the exposure process may be seen as the superposition of many 'local' exposure events due to secondaries at primary inelastic scattering points. Figure 1.2 shows the result of simulating the path of 100 electrons through 2 micron of PMMA resist on a silicon substrate, assuming the incident beam to be have an energy of 50kV, and to have a delta-function distribution. The program was written by S.P.Beaumont and calculates the path of each electron assuming Rutherford scattering for elastic events, and incorporating the method first

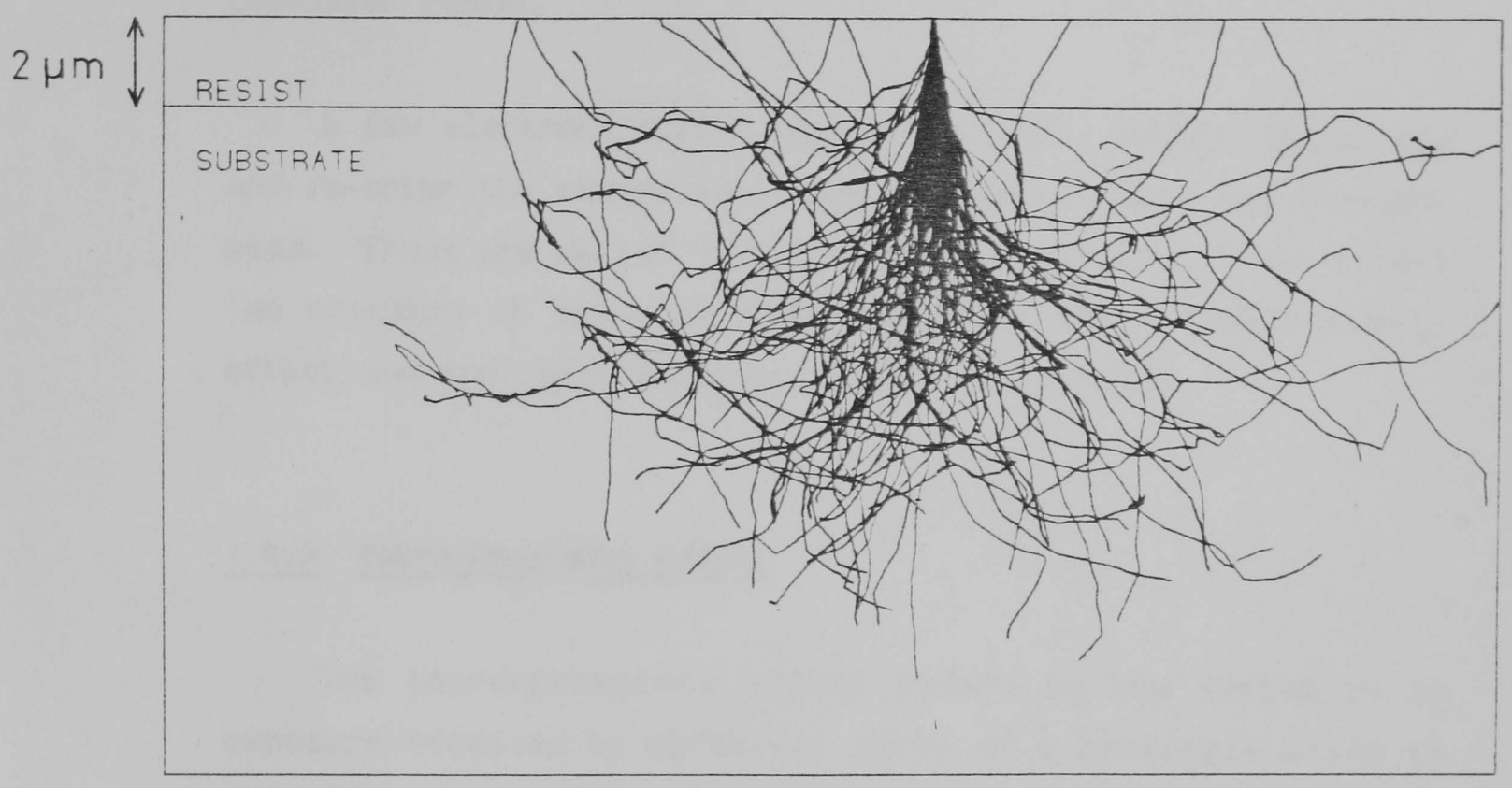


Fig. 1.2 Monte Carlo simulation of 100 50kV electrons entering PMMA on a silicon substrate. Note that some electrons are scattered back into the resist.

used by Shimizu [1.10] to take account of inelastic events. If high resolution (narrow, closely spaced) lines are required, it may be seen that thin resist should be used to minimise the effect of beam widening by forward scattering. But if the resist is too thin, (say less than 0.5 microns, for a 50kV beam), the undercut profile required for liftoff has to be provided by a two-layer resist.

A few electrons suffer high angle (>90^{o}) elastic scattering and re-enter the resist and cause exposure outside the incident beam. These are called 'backscattered electrons'. They affect the exposure in two ways referred to as the intra-proximity effect and the inter-proximity effect [1.11,1.12].

1.4.2 Intra-Proximity Effect

The intra-proximity effect refers to the variation in exposure received by different parts of a rectangle which is exposed using a uniform dwell time per picture point in the exposure field scanned out by the beam (pixel). Because the total exposure at any one point is the sum of contributions from nearby points (assuming that the pixel size is is much smaller than the range of backscattered electrons, a condition that is satisfied if the pixel size is less than 0.5 microns), it is found that the exposure received at the edge of a large exposed pad is half that received at the centre, and the exposure received at a corner is a quarter that at the centre. The intra-proximity effect causes wide rectangles to require less exposure than narrow ones. The effect is illustrated in Fig. 1.3, which shows the variation of critical dose with linewidth for lifted off lines (i.e. the minimum exposure for which lift off was successful). A wide pad requires only a third the dose of a single pass line. The anomalous peak in the curve at 4 pixels is not understood.

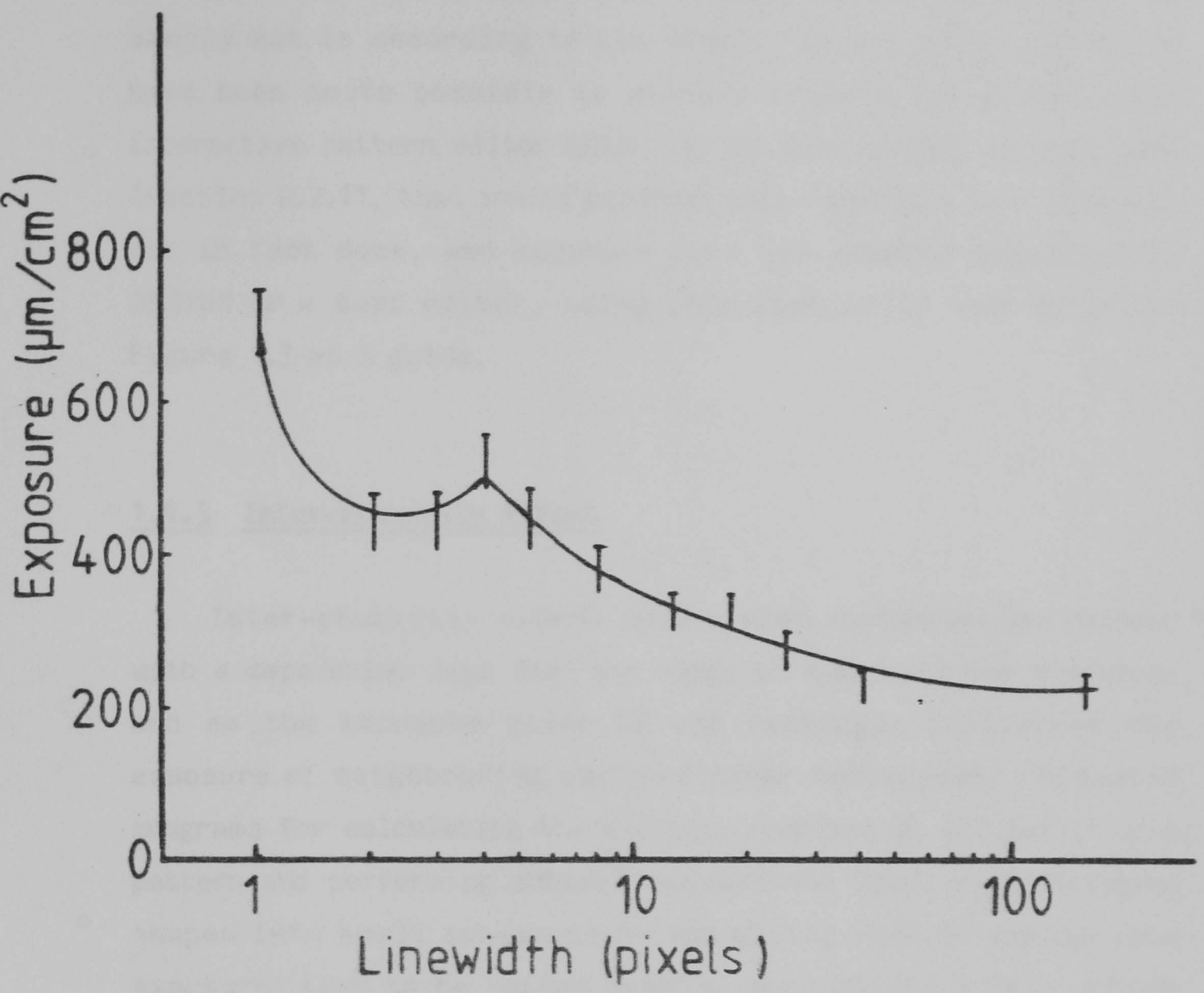


Fig. 1.3 Plot of critical exposure versus linewidth (pixels) for 0.7 microns of BDH PMMA (mol. wt. 185,000) on GaAs, pixel size 0.38 x 0.29 microns, 50kV beam, 0.5 micron spot. Development was for 30 seconds in 1:1 IPA:MiBK at 23.0°C.

In practice it is easy to make at least a first order correction for intra-proximity, once data such as that presented in Fig 1.3 has been obtained. The exposure for each rectangle is simply set to according to its width. In principle, it would have been quite possible to write a routine for DESIGN, the interactive pattern editor which was written as part of this work (Section 2.2.1), that would perform this function, but this was not in fact done, and exposure data was entered manually via DESIGN or a text editor, using data similar to that shown in Figure 1.3 as a guide.

1.4.3 Inter-Proximity Effect

Inter-proximity effect arises when rectangles are exposed with a separation less than the range of backscattered electrons and so the exposure given to one rectangle influences the exposure of neighbouring parts of other rectangles. Computer programs for calculating the exposure received at all points of a pattern and performing suitable corrections (such as fractioning shapes into small sub-sections and giving each an appropriate exposure) tend to be rather complex and require a fair amount computing power [1.13,1.14]. None of the patterns used in this work is very complex (all contain <500 rectangles, cf. a commercial chip with perhaps 100,000 elements or more), so the computing time required to perform the correction would not be excessive, but it was felt that the time spent writing such a program would not be saved during its use. All inter-proximity effect correction was done manually and iteratively (see Section 6.5.1.1).

As patterns are written more densely, inter-proximity correction becomes more difficult, and for very high resolution patterns it would clearly be useful to remove the effect of the backscattered electrons. This can be done by removing the substrate, or at least, by using a substrate thin enough that electrons in the primary beam have only a low probability of

scattering in it, thus the backscattered contribution becomes negligible. If a thin substrate is used in conjunction with thin resist, then the dominant factors affecting exposure and linewidth will be the beam diameter and the exposure mechanisms in the resist itself, such as secondary emmission from inelastic scatterng events within the resist, and molecular size effects.

1.5 THIN SUBSTRATES VERSUS SOLID SUBSTRATES

There are several reasons why it is useful to use a thin substrate for very high resolution electron beam lithography.

(1) Lack of backscattering contribution to exposure.
The limiting factors for resolution and linewidth on a thin substrate are the beam diameter (if it is greater than 10nm), processes of exposure in the resist itself (secondary electron emission), and molecular size effects during development [1.15].

(2) The specimen may be imaged using scanning transmision microscopy (STEM).
This gives a better resolution and higher contrast image than SEM mode where the same secondary electron processes that limit resolution in the resist cause the effective spot size on the specimen to be larger than the incident beam. The higher resolution and contrast of STEM make it easy to obtain good focussing and astigmatism correction, both of which are essential if consistant high resolution lithography is to be performed successfully. A further advantage is that once a pattern has been fabricated, it can be examined in high resolution STEM or TEM without further processing of the sample. The relatively poor resolution of SEM means that it is difficult to measure linewidths of a few nanometers accurately, if indeed they can be imaged at all.

(3) The superior contrast of STEM mode allows very high resolution alignment between successive lithographic steps to be performed (Chapter 5).

The obvious disadvantage in using thin substrates is the time involved in the fabrication of the substrates themselves, if indeed it possible at all for a given material. Isolated metal lines may be fabricated on solid substrates with the same linewidth as on thin substrates, but, if many closely spaced lines are required, the process latitude for lines on the solid substrate is very much reduced compared with that for a similar pattern on a thin substrate. An experiment performed by Beaumont demonstrates this [1.16]. Gratings were exposed across the edge of silicon nitride membranes so that half of each line lay on the membrane and half lay on the solid silicon. Gratings with line pitches of 1 micron and 70nm were exposed with a range of electron exposures. The silicon was etched back so that the lines that had been exposed on the solid and the thin substrates could be imaged in a TEM. The linewidth versus exposure was plotted for each of the four cases; widely spaced lines on the thin and the solid substrate, and narrowly spaced lines on the thin and the solid substrate, Fig. 1.4. The exposures have been normalised with respect to the minimum exposure for which lines were successful fabricated. For widely spaced lines there was no difference between lines on the solid substrate and the thin substrate. However for closely spaced lines, a difference between the two cases emerged. On the solid substrate the linewidth increased fairly quickly with exposure until liftoff failed at 1.4 times the critical exposure. On the thin substrate, the linewidth increased slightly up to 1.6 times the critical exposure, but above that there is a region where there is no variation of linewidth with exposure and the closely spaced lines have the same width as the widely spaced lines. The underlying reasons for the shape of the linewidth vs. exposure curve for thin substrate exposures are not well understood, but the conclusions that can be drawn from the experiment are that

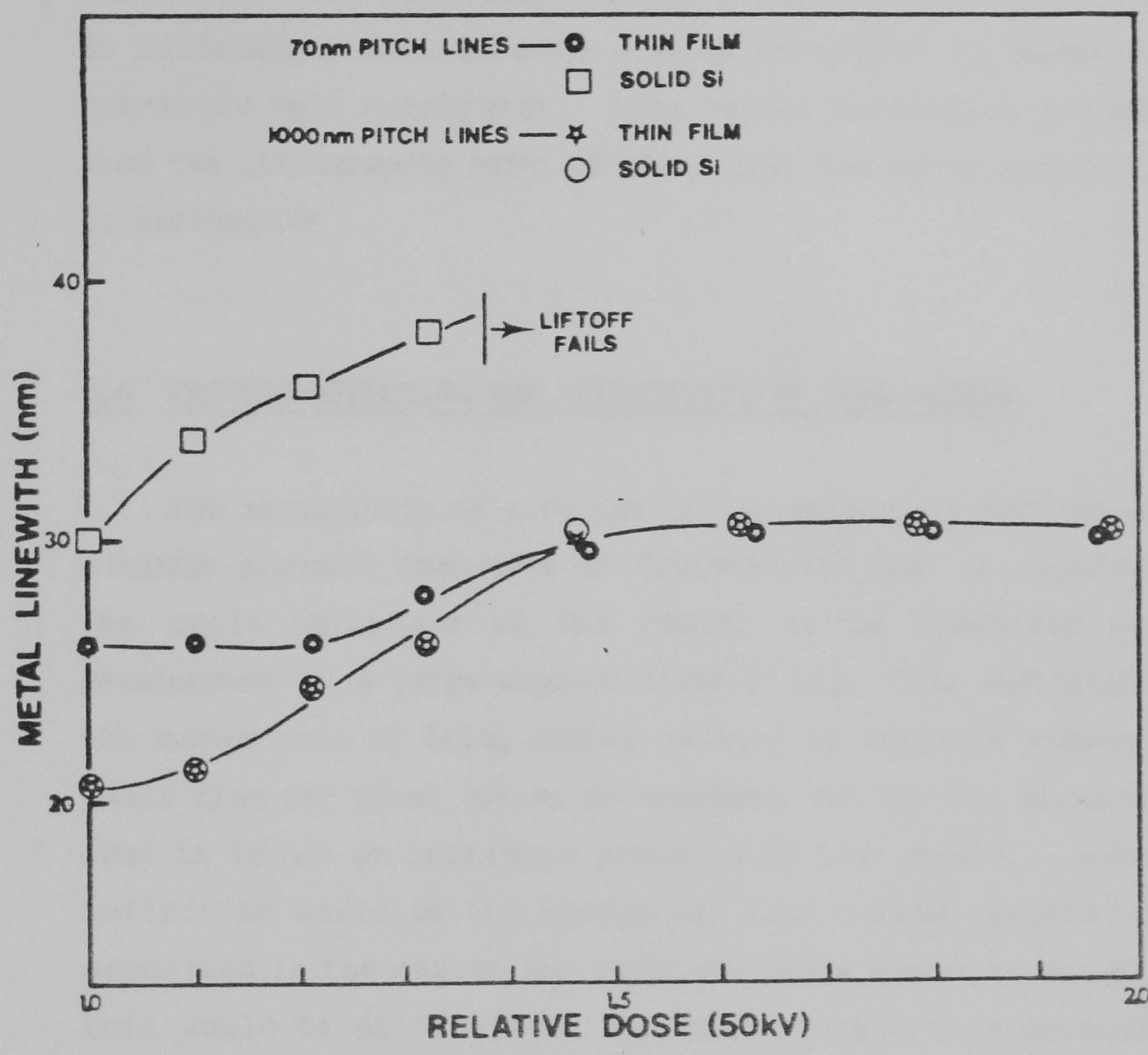


Fig. 1.4 Linewidth versus exposure for widely spaced and closely spaced lines on thin and solid substrates.

very high resolution patterns can be drawn on solid substrates but the exposure latitude for a dense pattern is very poor when compared to the same pattern when drawn on a thin substrate. Hence it is likely that most high resolution patterns which it might be desired to write could be fabricated on a solid substrate, but it is much easier to do so on a thin substrate. So although a considerable amount of effort is required to fabricate thin substrates, they confer sufficient advantage, from the lithographic point of view, that the extra work involved is worthwhile.

1.6 FACTORS AFFECTING THE SENSITIVITY OF PMMA RESIST

The sensitivity of a resist is conventionally defined as the minimum exposure (measured in Coulomb/cm^2) that is required for the whole thickness of the resist to be dissolved during development of a large exposed block [1.20]. This definition has the convenience of being simply related to the beam current and dwell time per pixel during an exposure, but has the disadvantage that it is not an intrinsic property of the resist. A better definition would be the energy per unit volume required to be deposited in the resist for complete development to occur, but this would be difficult to implement in practice because the value is not directly accessible from the beam current and dwell time. Figure 1.5 shows a typical plot to find the sensitivity (conventionally defined) of a resist. The thickness of resist remaining after development is plotted as a function of exposure. The point where the curve drops to zero corresponds to the critical exposure, also called the sensitivity of the resist. Thus the more sensitive a resist is to electron exposure, the lower the value of its sensitivity. The sensitvity of PMMA is affected by many factors, including; beam accelerating voltage, substrate material, PMMA thickness, molecular weight of PMMA, developer, developer temperature, and development time [1.21].

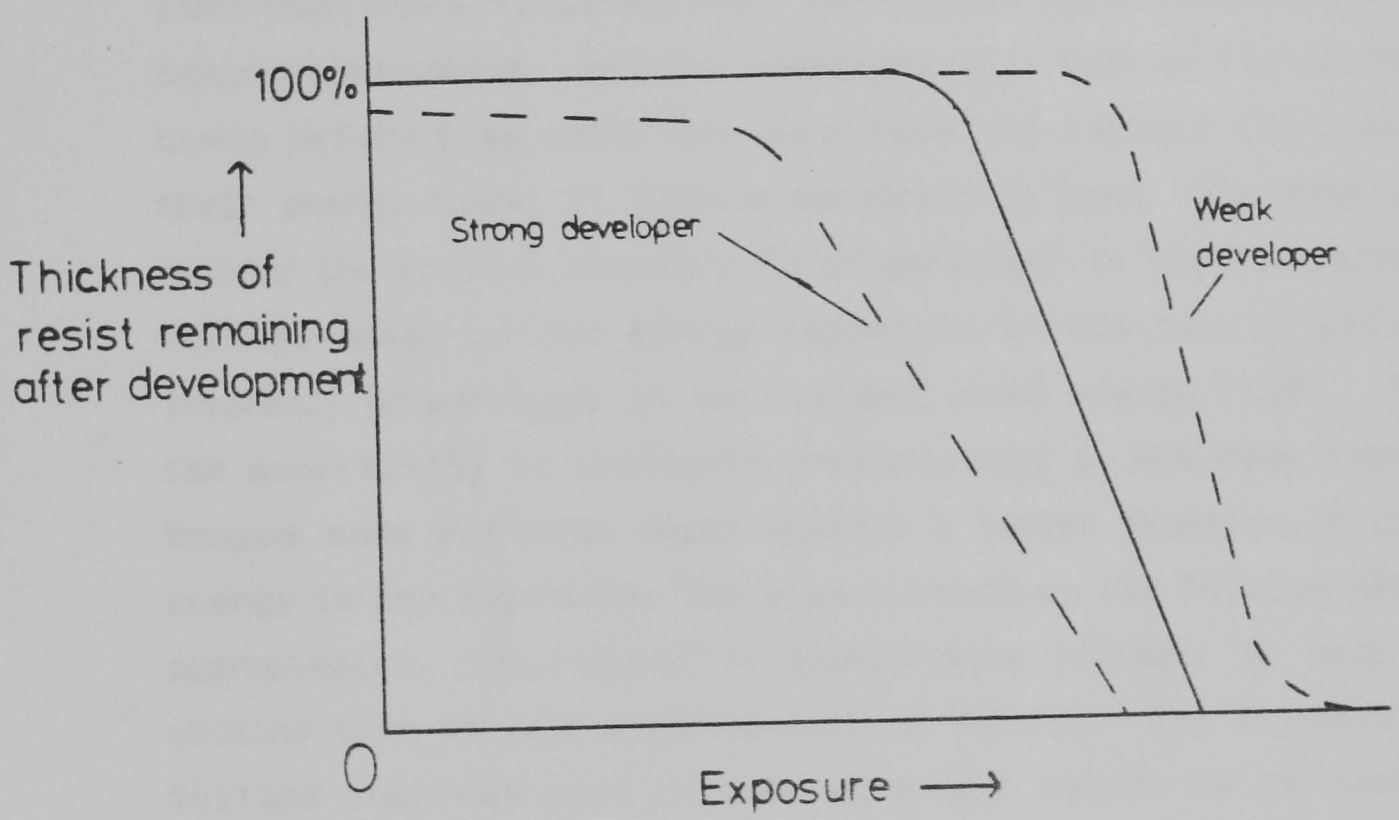


Fig.1.5 Typical sensitivity plot for PMMA. Critical exposure is where zero thickness of resist remains after development. Strong developer reduces critical exposure, but reduce contrast and dissolves unexposed resist. Weak developer has high contrast, but has 'tail' caused by inability to dissolve the few long chains left in exposed region.

The contrast of a resist is defined to be $\log_{10}$ of the slope of the straight part of the sensitivity curve [1.21]. High contrast resist/developer systems tend to give good resolution because small differences in exposure in the resist are translated into large differences in developed thickness.

1.6.1 Accelerating Voltage

The dominant exposure mechanism in PMMA is chain scission. The primary beam scatters both elastically off atomic nuclei and inelastically off electrons in the outer shells of the constituent atoms. Each inelastic collision results in a secondary electron, normally with an energy of less than 100eV, being emitted from the collision point while the primary electron carries on through the resist not usually having been deviated by more than a few milliradians. For resist thicknesses of a few hundred nanometers, primary electrons will scatter two or three times before they enter the substrate and deposit the bulk of their energy there, if a thick substrate is used. The mean free path of the primary electrons is proportional to the accelerating voltage used, so the energy deposited in the resist will be inversely proportional to the incident beam energy [1.22]. Thus the sensitivity is inversely proportional to the beam energy, because more energetic beams deposit a larger fraction of their energy in the substrate. The beam current on the Philips SEM is approximately proportional to accelerating voltage, so that the writing time is fairly independent of voltage. The accelerating voltage that was used in this work was chosen to be 50kV in preference to lower voltages, because it appeared to give slightly more reliable results for liftoff on thin substrates and reduces proximity effect on thick substrates, because collisions occur deeper in the substrate and the probability of electrons re-emerging through the resist is reduced.

1.6.2 Substrate

The substrate affects the resist sensitivity, via the intra-proximity effect because scattering events in the substrate can affect the resist. When the contribution from backscattered electrons is added up for the thousands of picture points within a rectangle, the backscattered contribution can become a sizeable proportion of the total exposure. The net effect is that high

scattering substrates give lower sensitivity values for a given resist, other things being equal.

1.6.3 Resist Thickness

The resist thickness affects sensitivity simply because exposed material has a finite dissolution rate in whatever developer is being used. For given developer conditions, more exposure will be required to develop through thicker resist in the same time as for thinner resist, because the average molecular weight in the exposed regions will have to be reduced in order that the development rate is increased. In fact, if the thickness of resist is changed, it is normally the development time that is altered, and the sensitivity stays close to its original value.

1.6.4 Molecular Weight

High molecular weight PMMA is less sensitive than low molecular weight PMMA, because more chain scissions are required to break each molecule into sections small enough to be dissolved by the developer.

1.6.5 Developer Composition

The dissolution rate of exposed regions changes with developer and so the developer affects the sensitivity. Most PMMA developers are mixtures of a weak solvent for PMMA and a non-solvent. Such mixtures will dissolve PMMA up to a certain molecular weight only, which gives contrast between exposed and unexposed regions. The stronger the developer, the faster the dissolution rate, and the lower the sensitivity, but this may be at the cost of decreased contrast due to dissolution of unexposed material. The developers used in this work were mixtures of

propan-2-ol (IPA), a nonsolvent for PMMA, and 4-methylpentan-2-one (MiBK) a weak solvent [1.20]. For thin substrate exposures a 3:1 mixture was used, which gives very good contrast, while on solid substrates a 1:1 mixture was used, in order to boost the sensitivity, and reduce the writing time of large blocks. Since only relatively low resolution was required on solid substrates, the small loss of contrast suffered through the use of a strong developer was not important.

1.6.6 Development Time

Development time is not usually a crucial factor for sensitivity, because the development times are deliberately long so that the sensitivity is "saturated". That is, the sensitivity decreases with development time up until some value called the critical time above which it levels off. To increase process latitude, normal development times are longer than the critical time. (See Section 6.4.4 for a description of how this was implemented in practice).

1.6.7 Developer Temperature

The dissolution rate is a strong function of temperature, the dissolution rate of PMMA in 3:1 developer increasing by about 35%/^{o}C, which corresponds to an activation energy of 2.43eV [1.20]. The sensitivity is not directly proportional with the development rate, since it is the ability to dissolve longer molecules of PMMA within the critical development time (which decreases with temperature) that causes the increase in sensitivity with temperature. Increasing the developer temperature from 23.0^{o}C to 38.0^{o}C increases the sensitivity by a factor of 3 but at the expense of contrast (see Section 6.4.4.3). Good temperature control is thus vital if consistant results are

to be obtained.

From the above, it is clear that figures quoted here, and elsewhere, should be regarded as pertaining only to those resist and development parameters and substrates that were used to obtain those values.

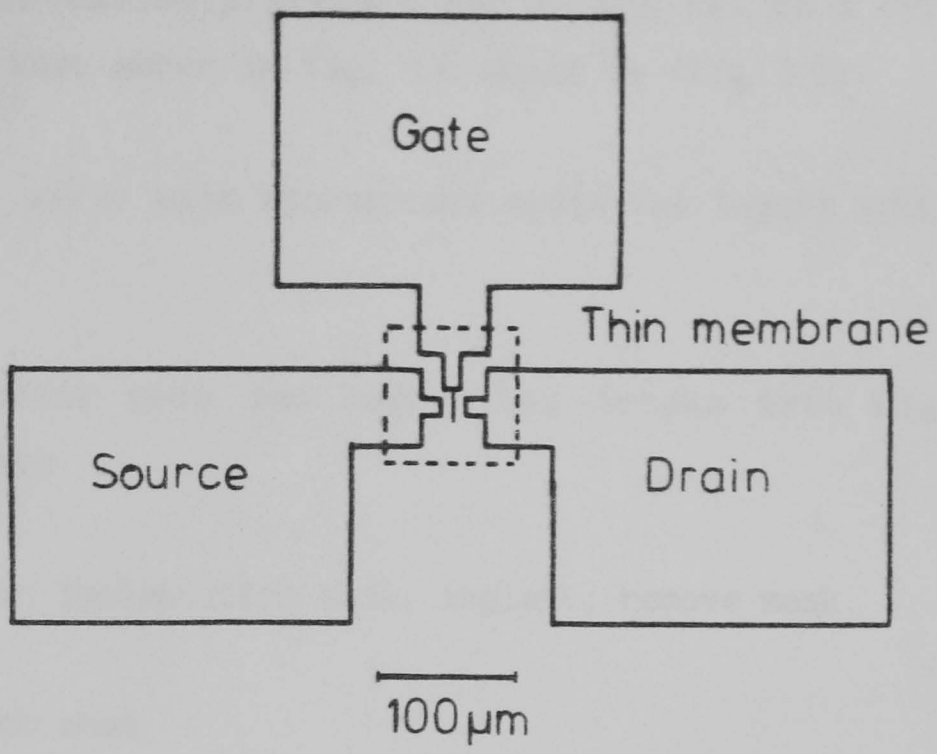
1.7 CHOICE OF GAAS FET AS TARGET DEVICE

Previous work in this Department had demonstrated the ability to fabricate 10nm isolated metal lines [1.4] and 45nm pitch gratings of 16nm lines on a 20nm thin carbon substrate, using a two layer PMMA resist system (Section 4.1.1.1), and liftoff by 'shooting' (Section 4.1.1.5). The aim of this project was to transfer this technology to silicon nitride membranes which are much easier to make and handle than carbon membranes, and then attempt to use the lithography to fabricate some device on a thin substrate of semiconductor that might show novel effects because of its small dimensions, and perhaps because of the thin substrate.

It has been predicted, and, to some extent demonstrated, that devices with dimensions of the order of a few tens of nanometers would exhibit different characteristics from larger devices of the same type [1.23,1.24,1.25]. Put simply, the reason for the expected difference is that in very small devices, electrons have a reasonable chance of passing through the device from one contact to another without scattering in the substrate; thus the electrons do not reach an equilibrium velocity in the substrate, and classical concepts such as mobility are meaningless. The initial aim of this project was to make some device with dimensions at the limits of electron beam lithography, which might exhibit novel properties.

The most obvious device to try to make first was the GaAs field effect transistor, because it is well understood for large sizes and is relatively simple to fabricate. The propagation delay time through such a device decreases as the gate length is decreased. Clearly there must be a lower limit on the delay time of FETs, either because there is a lower limit on the minimum feature size that can be made, or because as dimensions and doping levels are scaled with the gate length, some part of the device fails to keep the desirable properties of larger devices e.g. if the doping of the channel is too high, then the gate metallisation will form a leaky Schottky barrier, or if the cross-sectional area of the gate stripe becomes too small then the gate resistance might become a problem. Fig. 1.6 shows what such a device might look like when fabricated on a thin membrane. It is lithographically possible to fabricate a 100nm gap between drain and source and to fabricate a 20-30nm wide gate stripe.

If a series of transistors of various sizes could be made with decreasing gate lengths from 100-10nm (the current lithographic limit) the fastest possible device could be found, which may or may not take advantage of velocity overshoot and ballistic effects. If the speed of the fastest possible device could be measured, then given a certain computer architecture, an upper limit on the speed and 'intelligence' of future computers might tentatively be set.



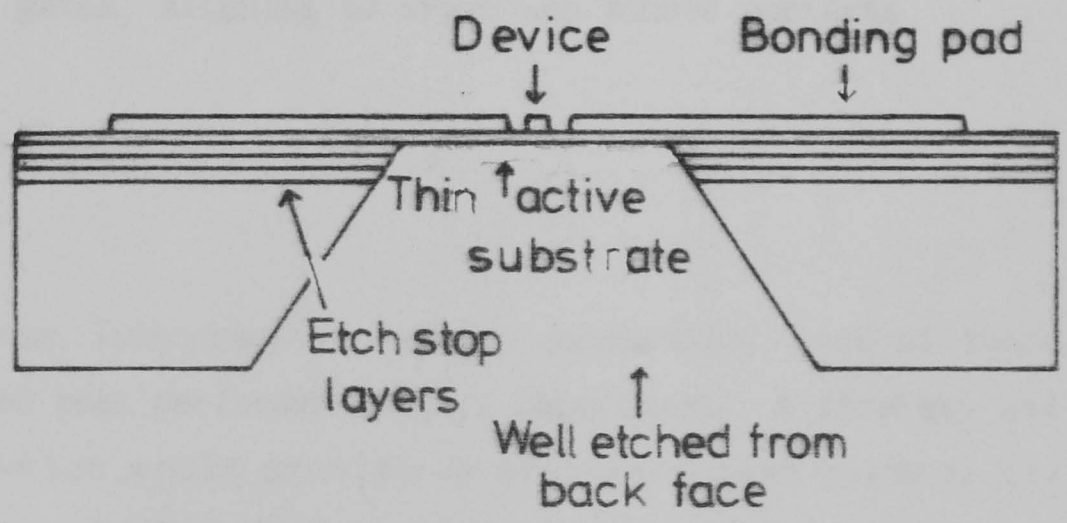


Fig. 1.6 Plan and cross-section through proposed thin membrane FET. Smallest possible device might have drain-source gap of 0.1 microns and gate length of 0.02 microns. Thin membrane formed by etching well from back of wafer and stopping on epi-grown heterostructure at front.

1.8 STRATEGY

The full fabrication procedure for a GaAs FET on a thin membrane such as that shown in Fig. 1.6 would be (Fig. 1.7):

Receive GaAs wafer with appropriate epitaxial layers grown on it

Pattern bonding pads and connecting tracks into high resolution area

Pattern proton implantation mask, implant, remove mask

Fabricate membranes

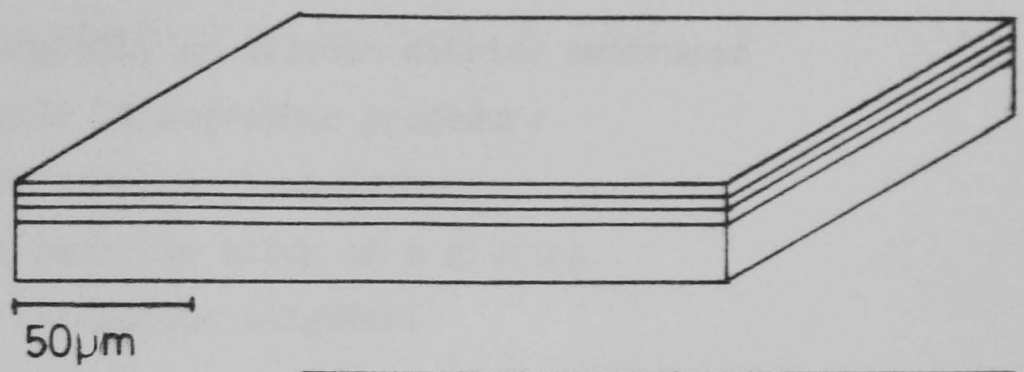
Pattern drain and source contacts, aligning to tracks

Pattern gates, aligning to drain and source contacts

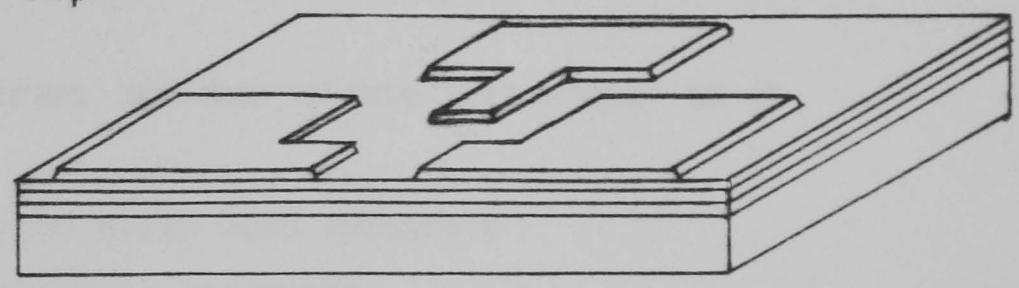
Test devices

Apart from lithography on carbon substrates, none of these procedures had been performed in this Department. A strategy was arrived at which would provide an efficient path towards the desired device. A major feature of the strategy was to test most procedures on cheap and easily fabricated silicon nitride membranes before moving onto the much more expensive and difficult to make GaAs membranes.

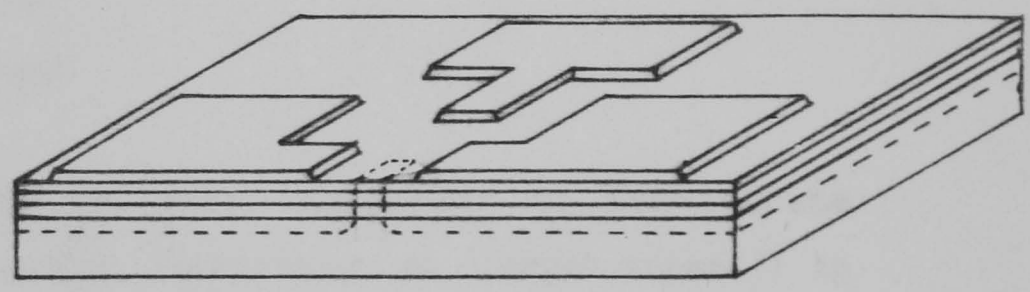
a) Wafer has active and etch stop layers grown on it



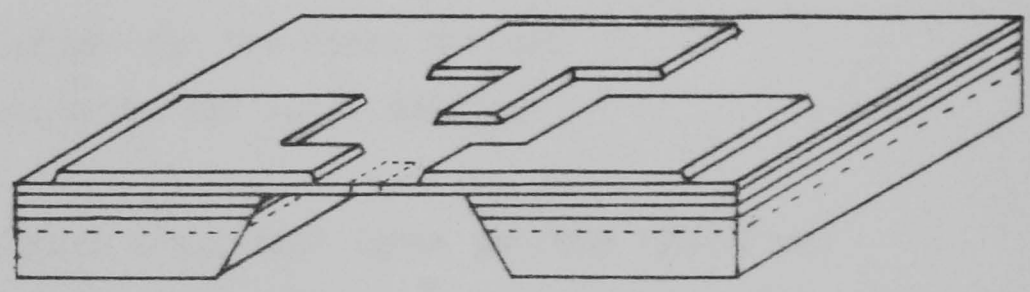
b) Pattern pads and connecting tracks



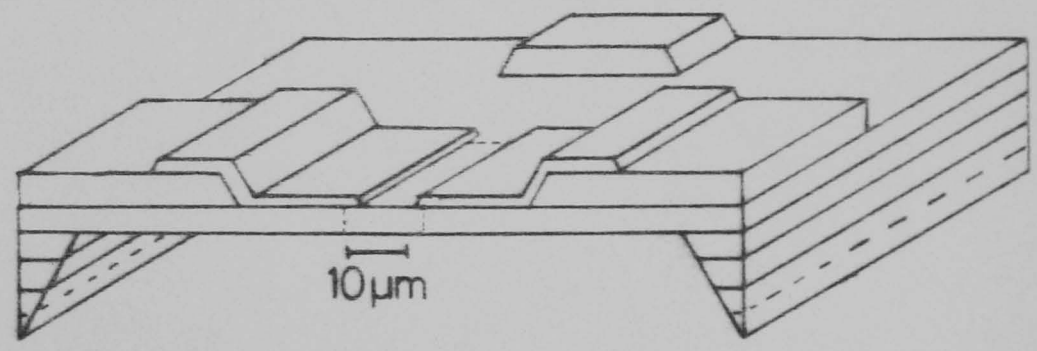
c) Isolate by proton implantation



d) Etch from back to form membrane



e) Pattern drain and source (ohmic)



f) Pattern gate (Schottky)

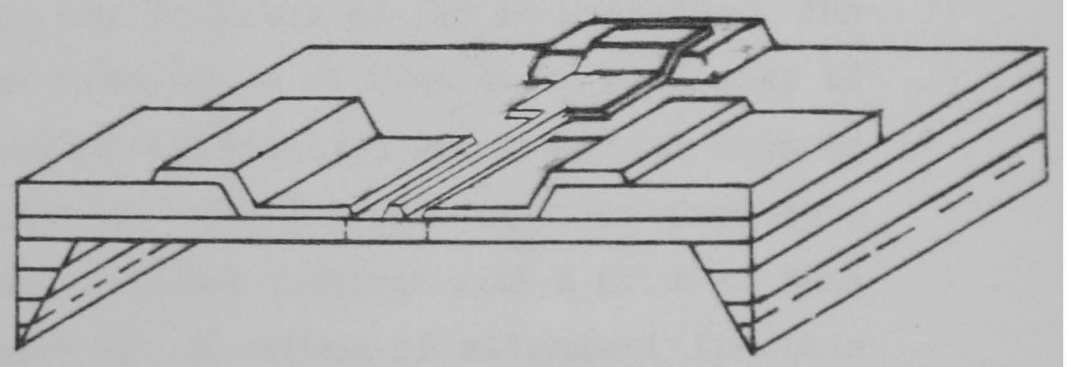


Fig. 1.7 Fabrication procedure for thin membrane FET.

Path to thin membrane GaAs FET

Find limits of lithography on silicon nitride membranes

-develop reliable lithographic procedure

-find narrowest possible linewidth

-find smallest possible pitch of a grating

-perform high resolution alignment

Fabricate GaAs membrane and demonstrate lithography on it

Perform lithography on solid GaAs substrate

-develop lithographic procedure

-find resolution

-peform alignment

Fabricate GaAs FET on solid substrate to demonstrate capability of transistor fabrication at 'large' sizes (1 to 0.1 micron gates)

-use mesa isolation for the first devices

-use proton isolation for later devices

Fabricate thin membrane transistor (gate lengths 100-10 nm)

-requires alignment from back of wafer to front for positioning of windows for etching the membranes

Test devices

1.9 RESULTS

The above strategy was followed as far as possible. The techniques developed for fabrication of 50nm thick membranes of silicon nitride and GaAs by preferential and selective etching are described in Chapter 3. The methods used to fabricate isolated metal lines 10nm wide and gratings with a pitch of 40nm are described in Chapter 4. A method of alignment for thin substrates that enables one pattern to be positioned with respect

to another with an accuracy of less than 5nm is described in Chapter 5. All solid substrate lithography was performed with a view to fabricating a solid substrate transistor. The lithographic techniques employed in making FETs with gate lengths from 1.5-0.075 microns, together with descriptions of experiments with ohmic contacts and proton implantation, are given in Chapter 6. The method and results of testing the devices are given in Chapter 7. Some overall conclusions and suggestions for further work are included at the end of Chapter 7.

References in Chapter 1

[1.1] "Electron-Beam Technology in Microelectronic Fabrication", Ch. 1, Ed. G. Brewer, Academic Press (1980)

[1.2] "The Physics of Microfabrication", I. Brodie and J.J. Muray, Plenum Press, New York (1982), p. 28.

[1.3] See ref. 1.2 p. 333.

[1.4] S.P.Beaumont, T. Tamamura, and C.D.W.Wilkinson, "Two layer resist system for efficient liftoff in very high resolution electron-beam lithography", Proc. Microcircuit 80, p. 381.

[1.5] H.G.Craighead, R.E.Howard, L.D.Jackel and P.M.Mankiewitch, "10-nm linewidth electron-beam lithography on GaAs", Appl. Phys. Lett. 42 38 (1983).

[1.6] A.N.Broers, W.W.Molzen, J.J.Cuomo, and N.D.Wittels, "Electron Beam Fabrication of 80A Metal Structures", Appl. Phys. Lett. 29 596 (1976)

[1.7] S.K.Ghandi, P.Kwan, K.N.Bhat, and J.M.Borrego, "Ion Beam Damage Effects During the Low Energy Ceaning of GaAs", IEEE Electron Device Lett. EDL-3 48 (1982).

[1.8] M.Hatzakis "Recent Developments in Electron Resist Evaluation Techniques", J. Vac. Sci. Technol. 12 1276 (1975).

[1.9] M.Parikh, "Energy Deposition Functions in Electron Resist Films on Substrates", J. App. Phys. 50 1104 (1979).

[1.10] R.Shimizu, Y.Katoaka, T.Ikuta, T.Koshikawa, H.Hashimoto", A Monte Carlo Approach to the Direct Simulation of Electron Penetration in Solids", J.Phys D 9 101 (1976).

[1.11] T.H.P.Chang, "Proximity Effect in Electron Beam Lithography", J. Vac. Sci. Technol. 12 1271 (1975).

[1.12] See ref. 1.2, p. 338.

[1.13] M.Parikh and D.E.Schreiber, "Pattern Partitioning for Enhanced Proximity-Effect Corrections in Electron-Beam Lithography", IBM J. Res. Develop. 24 530 (1980).

[1.14] W.D.Grobman, A.J.Speth and T.H.P.Chang, "Proximity Correction Enhancements for 1 micron Dense Circuits", IBM J. Res. Develop. 24 537 (1980).

[1.15] S.A.Rishton, Ph.D Thesis, Glasgow University (1984).

[1.16] S.P.Beaumont, B.Singh, and C.D.W.Wilkinson, "Very high Resolution Lithography - Thin Films, or Solid Substrates?", Proc. Tenth International Conference on Electron and Ion Beam Science and Technology, Montreal, 1982.

[1.17] P.Chaudhari, A.N.Broers, C.C.Chi, R.Laibowitz, E.Spiller, J.Viggiano, "Phase Slip and Localisation Diffusion Lengths in Amorphous W-Re Alloys", Phys. Rev. Lett. 45 930 (1980).

[1.18] R.B.Laibowitz, A.N.Broers, J.T.C.Yeh, and J.M.Viggiano, Josephson effect in Nb nanobridges, Appl. Phys. Lett. 35 891 (1979).

[1.19] S.M.Sze, "Physics of Semiconductor Devices", 2nd Ed. Wiley, 1978.

[1.20] J.S.Greeneich, "Developer Charactaristics of Poly-(Methyl Methacrylate) Electron Resist", J. Electrochem. Soc. 122 970 (1975).

[1.21] M.J.Bowden, "Electron Irradiation of Polymers and its Application to Resists for Electron-Beam Lithography", CRC Critical Reviews in Solid State Science, 223 (1979).

[1.22] Ref. 1.1, p. 94.

[1.23] M.S.Shur and L.F.Eastman, "Ballistic Transport in Semiconductors at Low-Temperatures for Low-Power, High-Speed Logic," IEEE Trans. Electron Dev. ED-26 1677 (1979).

[1.24] K.Hess, "Ballistic Transport in Semiconductors", IEEE Trans. Electron Dev. ED-28 937 (1981).

[1.25] L.F.Eastman, R.Stall, D.Woodward, N.Dandekar, C.E.C.Wood, M.S.Shur, and K.Board, " Ballistic Electron Motion in GaAs at Room Temperature", Electron. Lett. 16 524 (1980).

CHAPTER 2 THE LITHOGRAPHIC SYSTEM

2.1 GENERAL DESCRIPTION

The lithographic system is defined here to be the software and hardware that enables patterns to be scanned by an electron beam onto samples.

The system used in this work consists of several sections:-

-The interactive pattern editor DESIGN, which creates pattern files containing exposure and coordinate data.

-The program SENDPAT which converts exposure information into pixel dwell times, converts pattern files into hexadecimal code and sends them to a KIM microprocessor for storage in RAM.

-A machine code program residing in the microprocessor which delivers coordinates of rectangles to the scan generator and timing data to its programmable clock.

-The scan generator which controls the beam via D/A convertors on the scan coils and scans out rectangles in a reversing raster.

-The Philips PSEM 500 scanning electron microscope which has been modified for lithographic use.

Each of the above parts of the lithographic system is described in more detail in the following sections.

2.2 DESIGN, SENDPAT, and POSITION Programs

2.2.1 DESIGN

Patterns consist of rectangles defined within a 4096 x 4096 point coordinate array with its origin at the top left corner. The resolution of the array corresponds with the resolution of the 12-bit DACs on the scan coils which are controlled by the scan generator. Each rectangle is defined by the coordinates of its top left and bottom right corners, and these groups of four coordinates are stored sequentially in a file in a GEC 4070 computer in records of the form 'X1 Y1 X2 Y2', where (X1,Y1) defines the top left corner, and (X2,Y2) defines the bottom right corner. Lines in the file of the form '0 X 1 0', where X is an integer, indicate that the rectangles following this line should be exposed with a pixel dwell time such that the rectangles receive an exposure equivalent to X microcoulombs/cm^2. If rectangles are to be scanned with a fixed clock rate then a line of the form '0 T 0 0' may be inserted, where T is 10 times the required clock rate in microseconds (to allow for single decimal place precision). This feature is useful for scanned alignment patterns (see Section 5.2 and 6.5.2.2.1). A record containing '0 0 0 0' indicates the end of the pattern file.

DESIGN is a macro which handles pattern files and runs an interactive graphics pattern editor program. Patterns may be created or changed with it. The macro requests the terminal type and the name of the file to be edited. The editor program is then run. A square is drawn on the terminal screen and this is the "window" onto the pattern. Initially the whole frame is windowed, but selected areas of the pattern can be viewed to see fine detail. A cross-hair cursor is used as a pointer for many of the available commands. The rectangles are drawn in the same order on the screen as they will be when the pattern is exposed on a sample. The position of a rectangle in this order is called its 'scan number'. The order in which rectangles are exposed is important if hysteresis effects in the scan coils are to be

minimised (Section 2.4.3).

The commands in DESIGN are of four types.

(I) Shape Input and General Utilities

Rectangles can be input either singly, using R and S (or I), as a connected group forming a line of constant width using the L command, or as grating using the G command.

R Take current cursor coordinates and store them as the first corner of a rectangle.

S Completes a rectangle following a previous R command. The points designated by a pair of R and S commands are taken to define two diagonally opposite corners of a rectangle. The rectangle is drawn in on the screen.

I Allows coordinates of rectangles to be input directly from the keyboard (when high accuracy is needed).

L Line. Enables a series of abutting rectangles to be drawn to form a continuous track with width as input from the keyboard.

G Creates grating. A simple grating fills a defining rectangle with lines of a given width and pitch. A complex grating allows changes of pitch and/or width with any given periods.

K Deletes rectangle whose corner is indicated by the cursor.

X Assign an exposure to the indicated rectangle.

T Assign a dwell time to the indicated rectangle.

N Causes the scan number of the indicated rectangle to be returned.

Y Causes the scan number and/or the exposure or dwell time of each rectangle to be displayed below it during the next redraw.

(II) Screen Handling.

C Prints the current cursor coordinates on the screen.

D Redraws the pattern in the current frame.

Z Zoom. Redraws the pattern with the current cursor position as the centre of the window, scaling the picture by the factor input from the terminal.

W Window. Define a new window by entering its coordinates from the terminal. Pattern is redrawn as viewed through the new window.

F Redraws the pattern with the original 4096 x 4096 window.

P Plots the pattern as seen by the current window onto the Benson drum plotter for hard copy.

(III) Group Handling.

These commands act on a group of rectangles defined using the N command. The 'current group' consists of the rectangles with scan numbers lying between those of the rectangles on which the last two N commands acted, inclusive.

M Allows the current group of rectangles to be repeated in a rectangular array of any size with any X and Y repeat spacings.

A Allows the scan order to be changed. The current group of rectangles is moved in the order so that the first rectangle of

the group has a new scan number which is input from the keyboard.

V Move the current group by a translation whose X and Y shifts are entered from the keyboard.

IV End the Edit

E Exit from DESIGN editor. A file is created and the pattern data is stored in it.

Figure 2.1 shows hard copy output from DESIGN that illustrates the use of some of the commands.

In order to scan the file it has first to be converted to hexadecimal form and taken into the memory of the KIM microprocessor, which is done using SENDPAT (below).

Use of DESIGN enables patterns to be created quickly and efficiently, and to be changed easily, if required. It represents a great improvement over the previous method of pattern creation, which had to be done via a text editor, because the effects of inputting each rectangle can be seen immediately.

2.2.2 SENDPAT

SENDPAT is used in conjunction with the T command of the monitor program, EBSS (see Section 2.3), to transfer pattern data from a DESIGN file on the GEC 4070 to the RAM of the KIM microprocessor. The program converts exposure data into dwell times, and then converts the timing data and coordinate data into a hexadecimal string which is output to KIM. The dwell times are converted to three-digit hexadecimal numbers corresponding to the appropriate voltage to be applied to the input of the

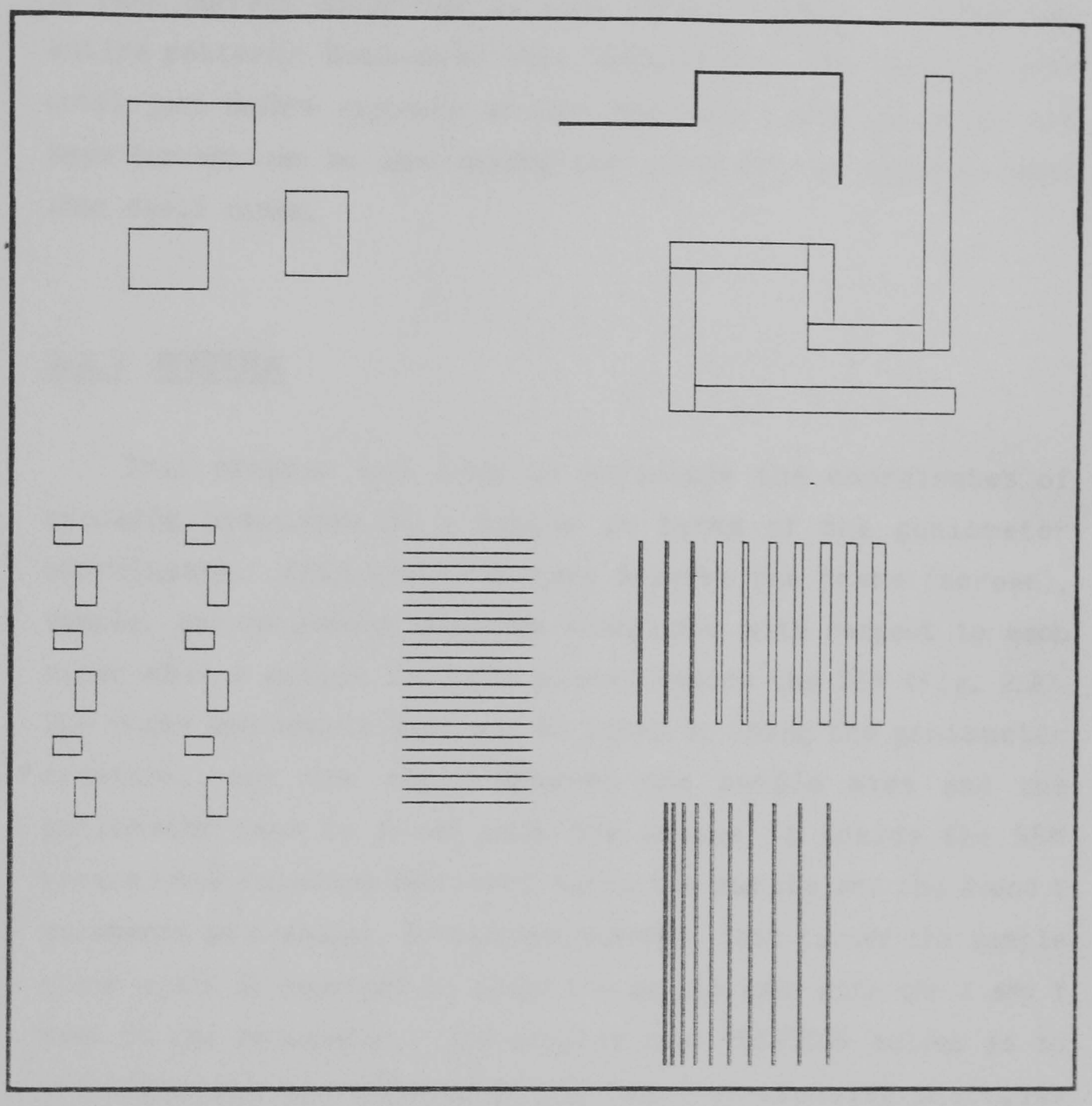


Fig. 2.1 Sample output from DESIGN.
Top left. Rectangle generated with **R** and **S** commands.
Top right. Lines generated with **L** command.
Bottom left. Pair of rectangle repeated in an array using **M** command.
Bottom right. Gratings. A simple grating plus two complex gratings, one with increasing linewidth, and one with increasing pitch.

programmable clock of the scan generator. The conversion of exposures to dwell times is done at this stage because the KIM does not have the capability to perform the calculations. It would be a more ideal situation if the calculation could be done after patterns had been stored, because corrections for changes in beam current could then be made without having to reload the entire pattern. Because of this, patterns are not usually sent until just before exposure so that the most recent value for the beam current can be used during the conversion of exposure data into dwell times.

2.2.3 POSITION

This program was used to calculate the coordinates of exposing positions on a sample in terms of the goniometer coordinates. This was necessary because the frame (screen), sample, and goniometer axes are misaligned with respect to each other when a sample is first placed inside the SEM (Fig. 2.2). The frame and sample axes can be lined up using the goniometer rotation, but the angle between the sample axes and the goniometer axes is fixed once the sample is inside the SEM because the rotation movement turns the sample and the X and Y movements as a whole. A rotation movement that turned the sample alone would be required to align the sample axes with the X and Y axes of the goniometer. The problem that POSITION solves is to transform the coordinates of the required exposing positions relative to the sample axes into coordinates relative to the goniometer axes. This is a simple coordinate transformation. The angle between the corresponding axes must be found as well as the offset between origins. This is easily achieved by finding the coordinates in the goniometer frame of two points lying on the sample frame's Y axis, normally the top left and bottom left corners of the sample for solid substrates or of the first membrane when thin substrates are used. The angle between the axes is found by calculating

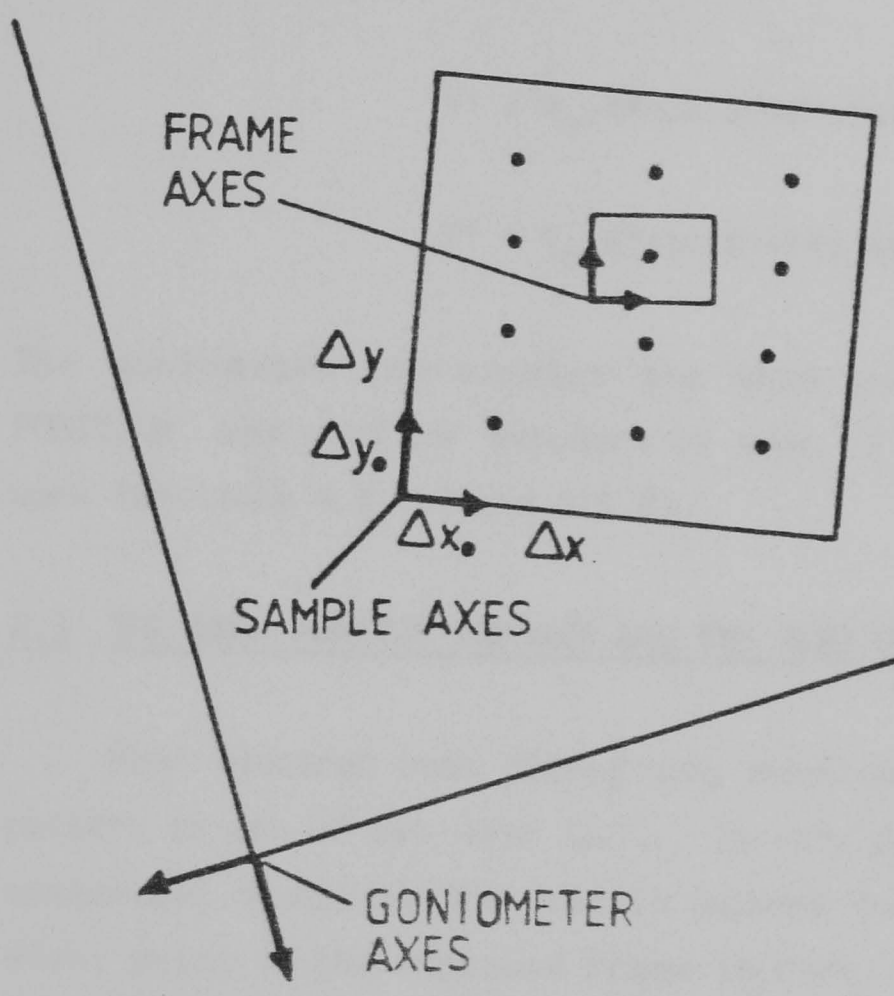


Fig. 2.2 Misalignment of axes. Sample and frame axes can be aligned by stage rotation, POSITION is used to calculate goniometer coordinates of exposing sites.

$$\theta = \arctan((Y1-Y2)/(X1-X2))$$

and the offsets are simply the coordinates of the bottom left corner, since it is the origin of the sample axes. The sample coordinates of the exposing positions are taken in by the program either from the keyboard or from a file. In either case the data consists of pairs of X and Y coordinates separated by carriage returns and terminated by a line containing two zeros. The coordinates are entered in units of millimeters. POSITION calculates the goniometer coordinates of the exposing positions using the transformations

$$X' = X_o - X*\cos(\theta) + Y*\sin(\theta)$$

$$Y' = Y_o - X*\sin(\theta) - Y*\cos(\theta)$$

The goniometer coordinates are then output to the screen. POSITION was used for exposure of most of the patterns in this work (Sections 4.1.1.2 and 6.4.3).

2.3 THE EBSS MONITOR PROGRAM AND THE SCAN GENERATOR

Most electron beam lithography machines expose the required pattern in one of two ways [2.1]. In very fast, high throughput, commercial machines, the beam is scanned in a raster that covers every point in the exposure frame in turn, but the beam is only unblanked at points that require exposure. This method of exposure has the advantage of reducing constraints on the lens design as far as hysteresis effects are concerned, but the disadvantage is that the data throughput required is very high, since data on every point in the frame must be output to the beam blanker. The other main method is called 'vector scanning'. Instead of scanning out every point in the frame sequentially, only the areas that require to be exposed are scanned. In the simpler machines all patterns are composed of rectangles within the frame and each rectangle is scanned out by a hard-wired pattern generator. The data transfer problem is greatly reduced because now only coordinate information on rectangles needs to be

transferred. When vector scanning is to be used, the lens and deflection system designs need to take account of the fact that the beam may require to step large distances going from one rectangle to the next. Compensation for proximity effect (Section 1.4) is considerably easier on a vector scan machine than a raster scan machine because patterns can easily be broken up into small parts each receiving a appropriate exposure.

A vector scan system was used in this work. Fig 2.3 shows schematically how the system works. The microprocessor feeds data into a scan generator which controls the beam via D/A converters connected to the X- and Y-scan coils. Patterns are usually exposed in an array of sites on a chip (a), the patterns are made up from rectangles which are scanned out in turn (b), each rectangle is scanned out by the beam in a reversing raster (c).

After the DESIGN file has been sent down from the GEC 4070 using SENDPAT, the pattern file is stored in RAM. The delivery of coordinates from RAM to the scan generator is controlled by a monitor program, EBSS (Electron Beam Scanning System). The KIM 6502 microprocessor has 4K of RAM. EBSS is stored in the memory up to the hexadecimal address $2945, the rest of the memory being available for pattern storage, which gives space for about 850 rectangles. The program obeys commands which instruct it to store a pattern in memory as it is transmitted from the GEC 4070 (T command), to store (R) or retrieve (G) pattern data on cassette tape, to scan a pattern once (S) or repeatedly (F), or to scan out a software generated grating of unit width lines with a given pitch (P).

The dwell time at each pixel is controlled by a programmable clock which uses a digital-to-analogue converter followed by a voltage-to-frequency converter to generate pulses periods from 2.0 to 1000 microseconds. The clock speed can be set manually, but it is usually controlled from within the pattern data. The clock period is latched until more timing data in the pattern

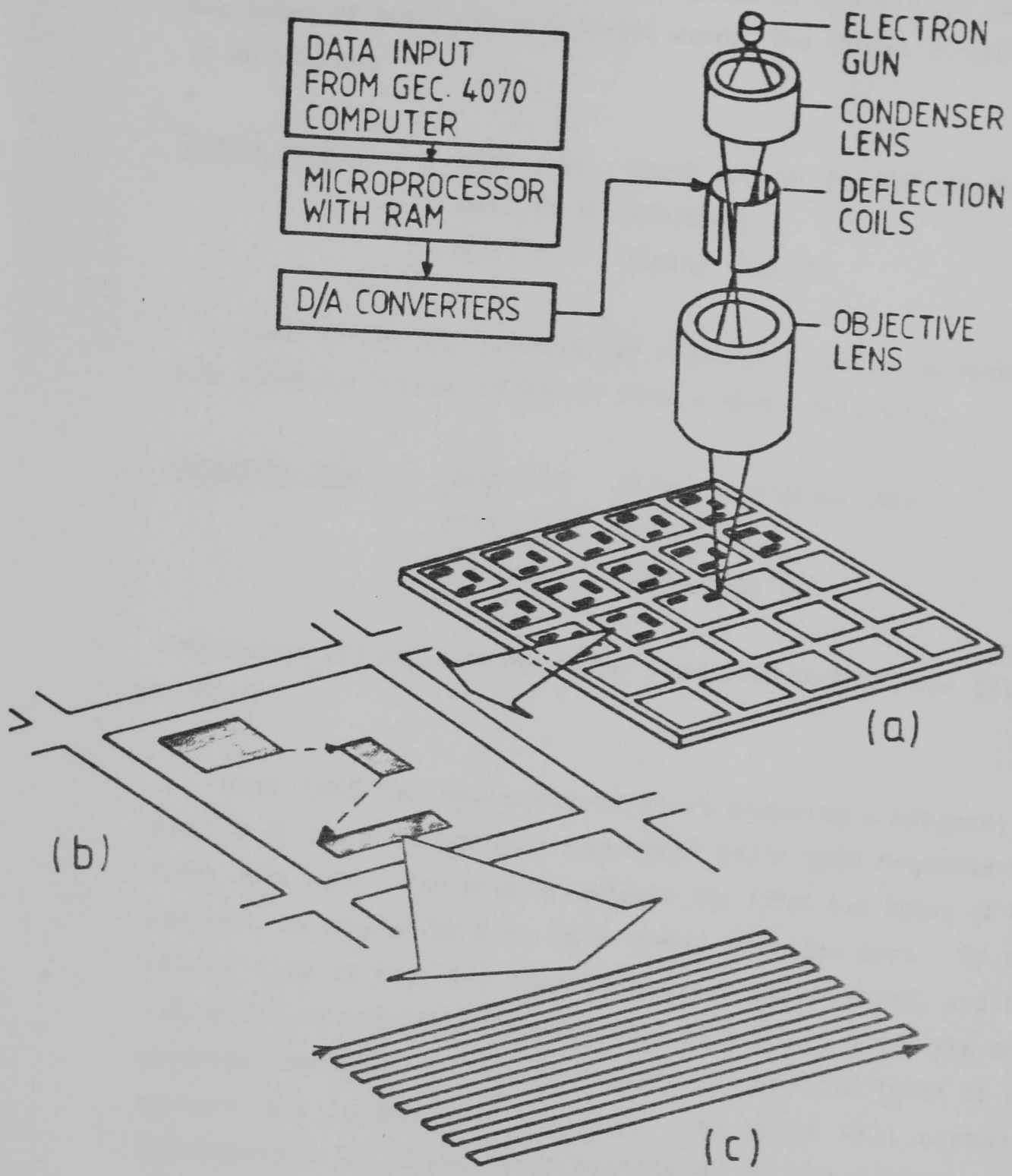


Fig. 2.3 The lithographic system.
a) Patterns are scanned out at several sites on a chip.
b) Each pattern consists of a series of rectangles.
c) Each rectangle is scanned in a reversing raster.

causes it to be changed. Thus, apart from the timing at the start of a pattern, timing data need only be inserted when a change in timing is required. Each piece of timing data occupies two bytes of RAM to form a 16-bit word. The format of each word is as follows.

Timing Data	Bits 0-11	Input voltage for V/F via DAC
	Bits 12-14	Arbitrary
	Bit 15	Timing flag (1)

Each of the four coordinates required to define a rectangle also occupies 2 bytes of RAM to form a 16-bit data word.

Cordinate Data	Bits 0-11	12-bit coordinate data
	Bits 12-13	Address bits
	Bit 15	Timing flag (0)

Coordinate information can range from $000 to $FFF, i.e. from 0 to 4095.

When EBSS is instructed to start scanning a pattern, it sends a GO pulse to the scan generator which then requests the first data word. The program outputs the first two bytes of the pattern, as stored in RAM, on a 16-bit parallel port. If the timing flag is set, the address decoder is disabled, and the timing DAC is enabled and loaded with bits 0-11. After the scan generator has returned a "ready" pulse, the next two bytes of the pattern are loaded onto the port, and these will normally correspond to X1 of the first rectangle. Figure 2.4 shows how the coordinates are loaded into registers which are then incremented or decremented to scan out a rectangle. The address bits cause the address decoder to first enable the registers labled X1 FILE and SCAN FILE, which then load the 12 data bits. The other coordinate registers are each then loaded in similar fashion. When all the registers are loaded, the beam is unblanked and is positioned at the point (X1,Y1) via the DACs which connect SCAN FILE and Y1 to the X and Y scan coils

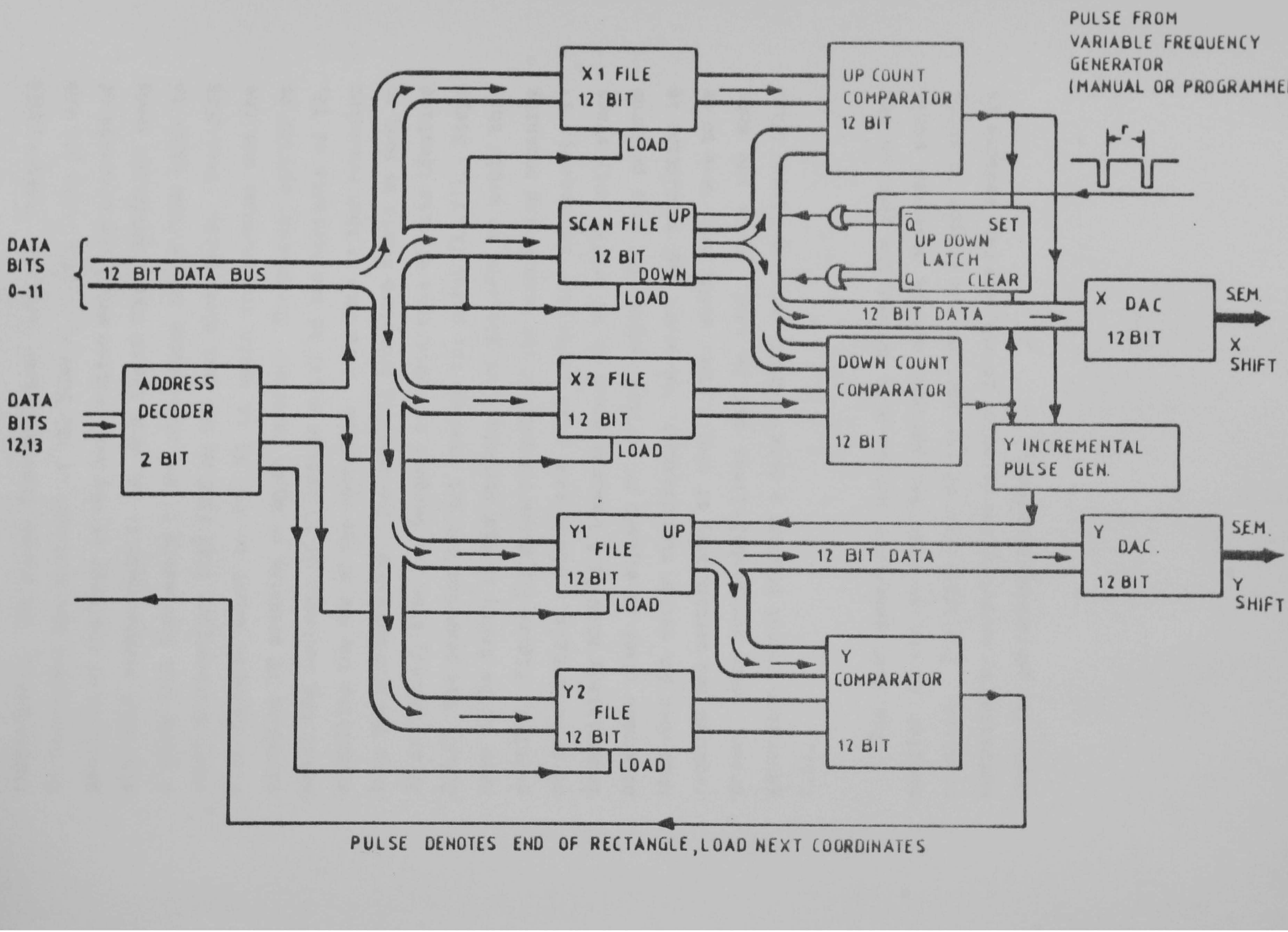


Fig. 2.4 Scan Generator.

respectively. The pulses from the clock cause the UP/DOWN LATCH to increment the contents of the SCAN FILE, and, with it, the position of the beam in the X-direction until the contents of SCAN FILE become equal to X2, when the UP COUNT COMPARATOR sends a pulse that increments Y1 and also causes the UP/DOWN LATCH to start decrementing SCAN FILE on the next clock pulse. When SCAN FILE becomes equal to X1, Y1 is again incremented and the direction of scanning is again reversed. The process carries on until the contents of Y1 become equal to the contents of Y2, signifying the end of the rectangle. More data is then requested from the microcomputer. The scanning procedure starts as soon as Y2 is loaded, even if, because of hysteresis effects (Section 2.4.3), the beam has not yet reached the point (X1,Y1). Steps have to be taken in the structure of patterns to avoid this problem. Figure 2.5 shows the path of the beam during scanning of a rectangle. Since the scan generator increments Y1 immediately SCAN FILE becomes equal to X2 or X1, only every alternate pixel is exposed on vertical edges. For many patterns this does not matter significantly, but where high resolution is required and particularly if single pixel width lines have to be drawn, only horizontal lines may be used, since the scan generator cannot produce a single width vertical lines, (Fig. 2.6).

The scan generator output also drives the spot on the monitor screen so that the pattern can be observed while scanning. The intensity of the spot on the screen is still modulated by output of the detector in use, and this feature is useful for performing alignment.

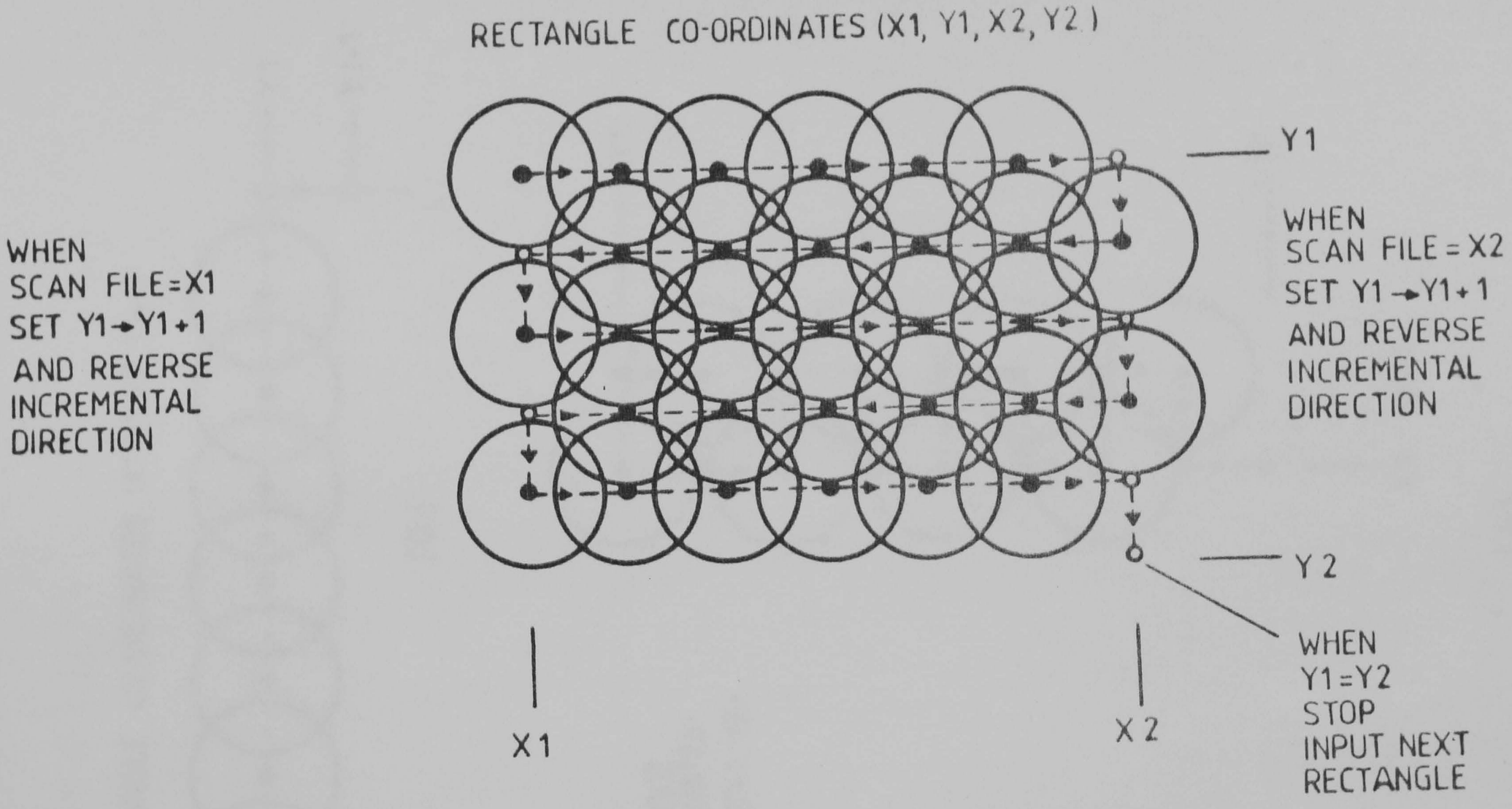
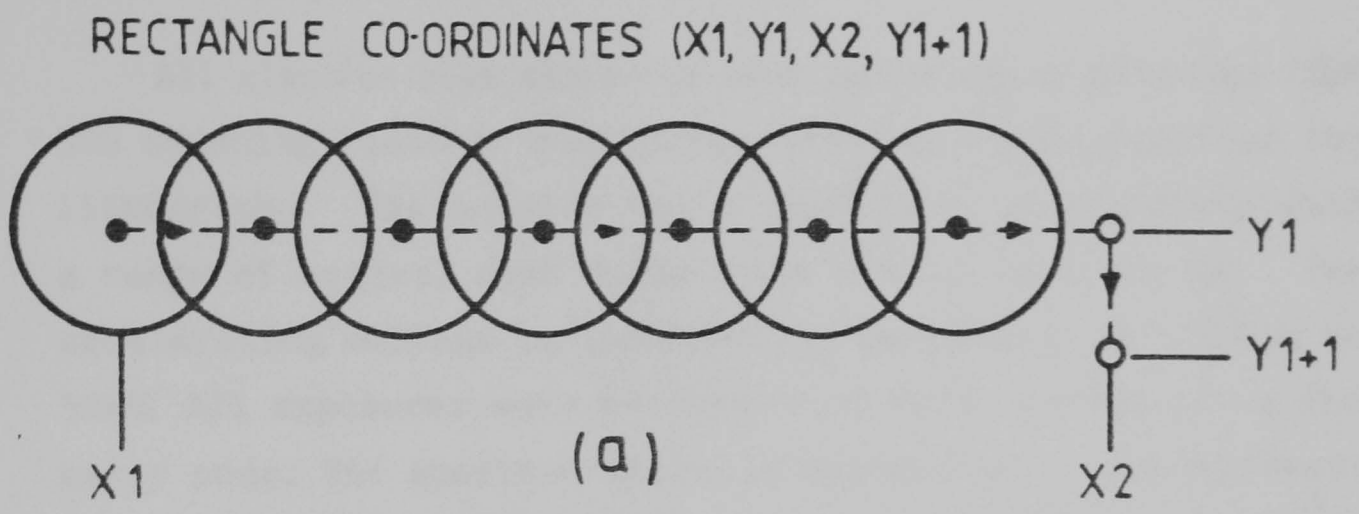


Fig. 2.5 Path of beam during exposure of a rectangle. Only alternate pixels are exposed on vertical edges because Y1 is incremented immediately SCAN FILE becomes equal to X1 or X2.



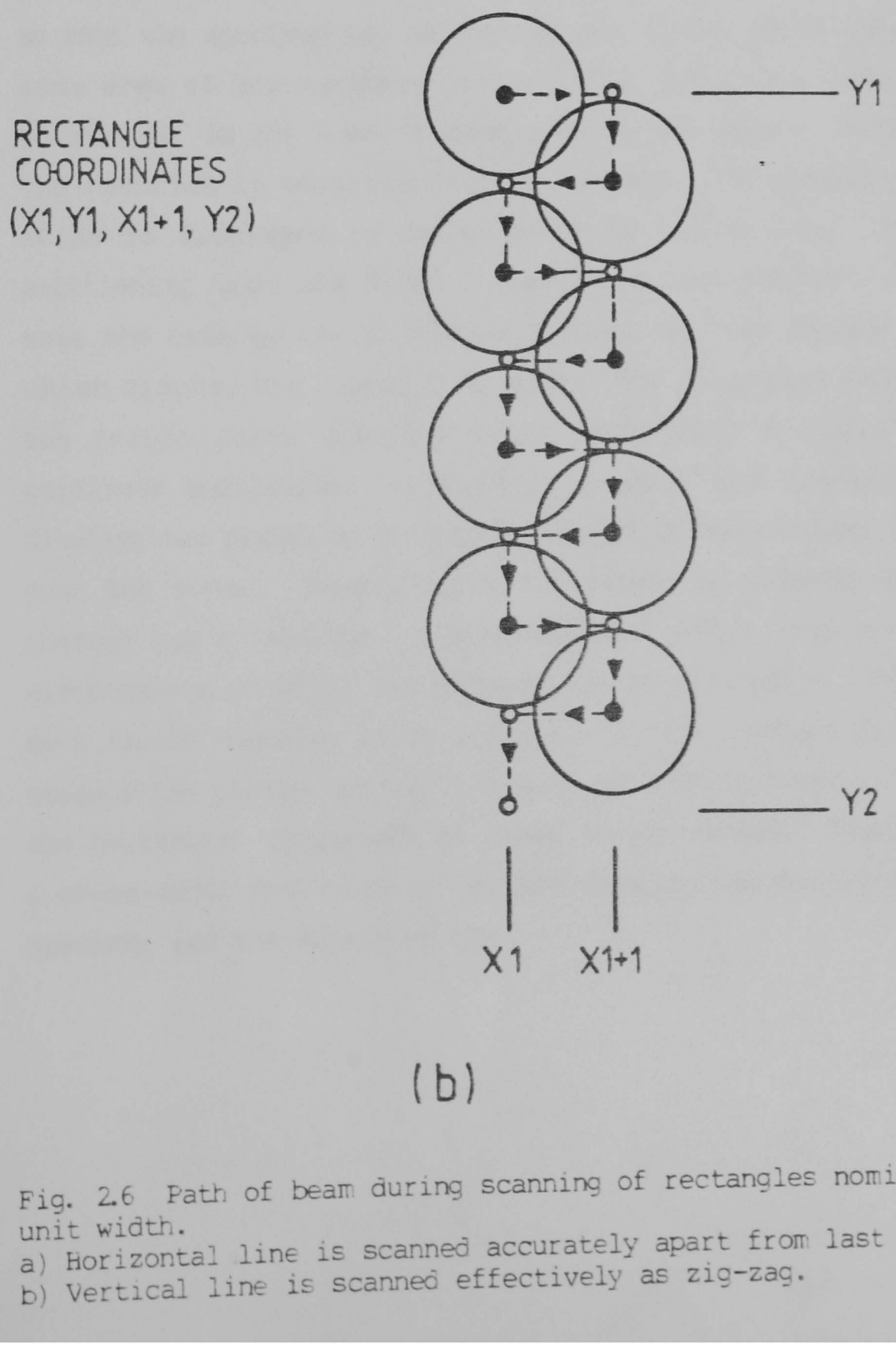


Fig. 2.6 Path of beam during scanning of rectangles nominally of unit width.
a) Horizontal line is scanned accurately apart from last pixel.
b) Vertical line is scanned effectively as zig-zag.

2.4 THE SCANNING ELECTRON MICROSCOPE

2.4.1 The Basic Machine

All electron beam exposures were performed in a Philips PSEM 500 scanning electron microscope which has been modified for lithography. The machine has a capability of operating with a range of nominal spot sizes from 8nm up to 1 micron. The accelerating voltage is continuously variable from 1.5kV up to 50kV. All exposures were performed at 50kV, except for a few early ones. The specimen stage is mounted on a five movement stepper motor controlled goniometer which has movements in X,Y, and Z, plus rotation and tilt. The goniometer is fully eucentric so that the specimen can be rotated and tilted while keeping the same area of the specimen in view (Fig. 2.7). The stage can be positioned in the X and Y directions in one micron steps, while the rotation is accurate to 1/68 degrees. The position of the stage is displayed to an accuracy of one micron. For fine positioning there are X and Y electronic beam shifters which can move the beam by +/- 10 microns. There are two monitor screens which display the signal from either the secondary detector or the transmission detector using intensity modulation and amplitude modulation. In addition a small fast phosphor screen displays the signal by amplitude modulation but without rastering down the screen. Samples up to 2.5 inches in diameter enter the chamber via a loadlock. A supplementary control box contains an electronic zoom which can increase the magnification continuously by a factor between unity and 2.5. A tilt correction control expands the picture in the direction of tilt in order to maintain the horizontal dimensions as shown on the screen. Figure 2.8 is a cross-sectional view of the SEM showing the positions of the specimen and the detectors etc.

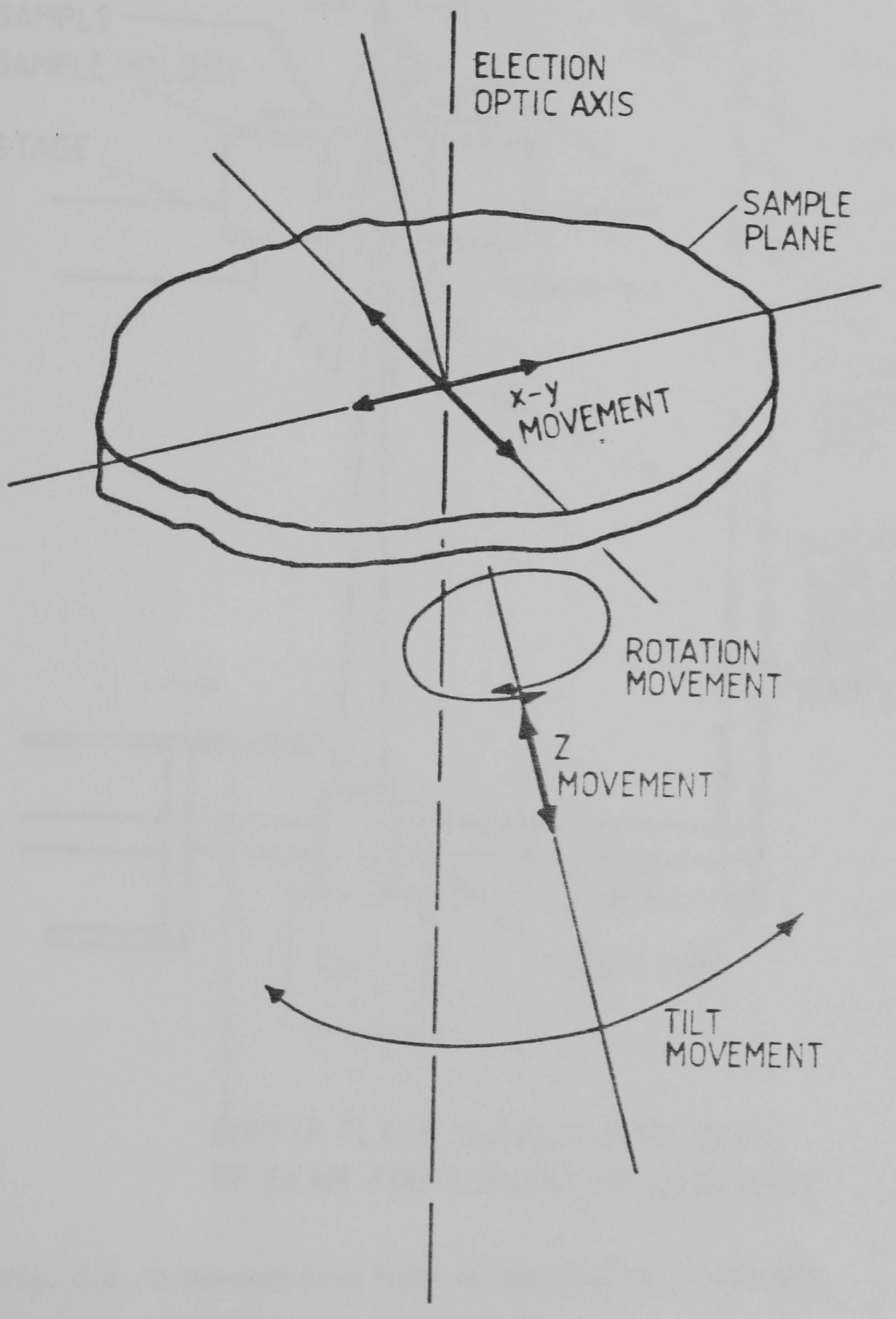


Fig. 2.7 The eucentric goniometer of the Philips SEM.

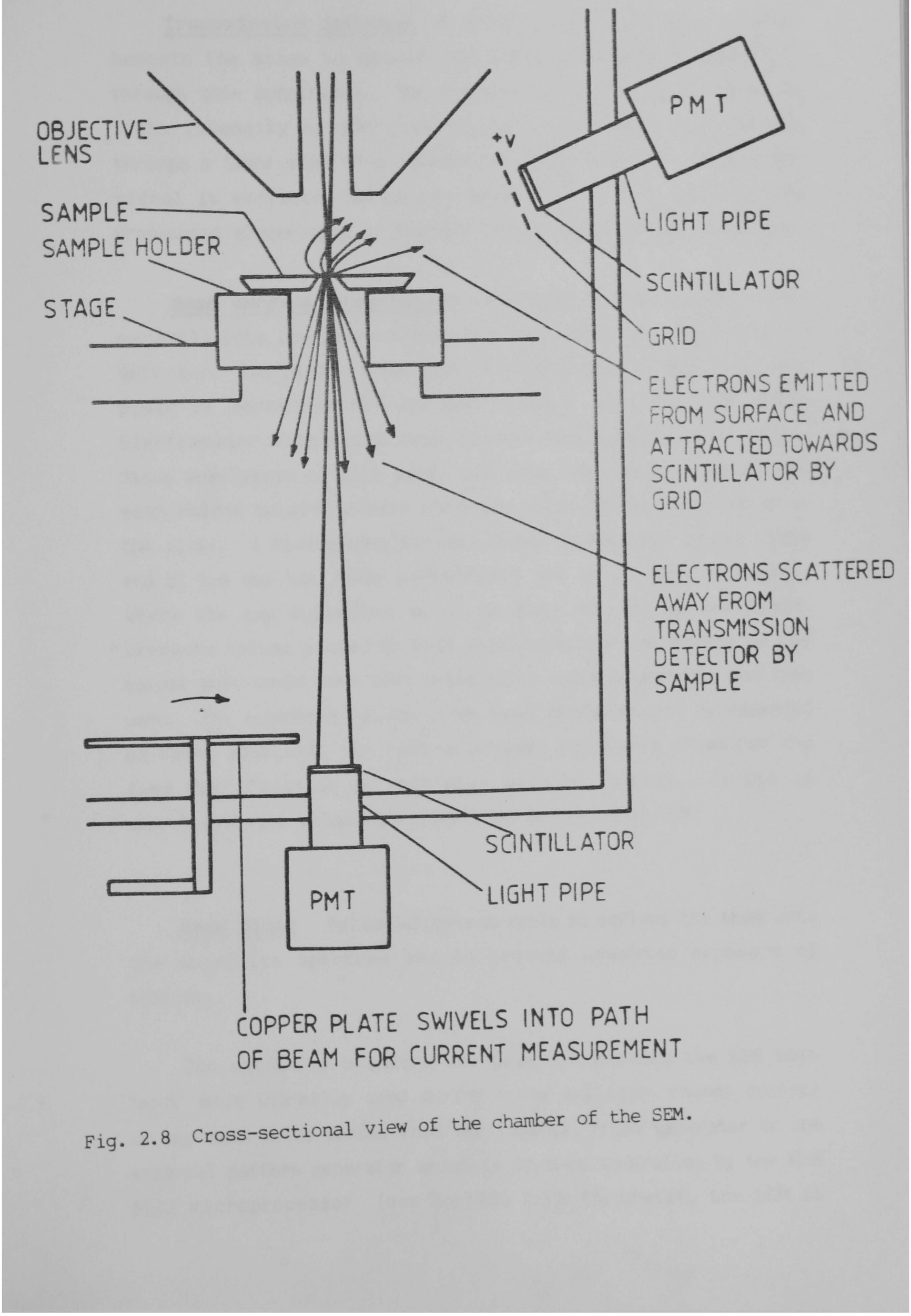


Fig. 2.8 Cross-sectional view of the chamber of the SEM.

2.4.2 Modifications

Transmission detector A scintillator has been mounted beneath the stage to detect the electron signal transmitted through thin substrates. The electron intensity is converted to light intensity by the scintillator, the light then passes through a light pipe to a Photo-Multiplier Tube (PMT) where the signal is converted to an electronic one and is sent to the processing electronics of the SEM before appearing on the screen.

Beam current measurement A copper plate can be moved manually into the path of the beam just above the transmission detector. The plate is isolated from the body of the SEM. The plate is connected through the chamber wall to a Keithley Electrometer with which beam current measurements are taken. Since completion of this work, a Faraday cup has been built into each holder to give a more accurate value for the current than the plate. A discrepancy between measurements made by the plate and by the cup was found particularly for small beam diameters, where the cup collected twice as much current as the plate. Exposure values quoted in this work have been converted to the values that would have been obtained if cup measurements had been used. The exposures values given must therefore not be regarded as being absolute, but ratios between exposures given for the same magnification or spot size will be correct. Errors in quoted exposure values are likely to be less than 20%.

Beam blank Pulses alignment coils to deflect the beam onto the objective aperture and so prevent unwanted exposure of specimen.

The action of blanking the beam and putting the SEM into 'spot' mode (normally used during x-ray analysis) causes control of the beam to be passed from the internal frame generator to the external pattern generator which is in turn controlled by the KIM 6502 microprocessor (see Section 2.3). Otherwise, the SEM is

under normal control.

Beam shifters The original single turn potentiometers of the beam shifters were replaced by ten turn potentiometers to give finer control. This modification was made in order to enable very high resolution alignment to be performed (Chapter 5)

X and Y zoom The standard Philips continuous zoom control uses a twin-ganged potentiometer to vary the X (line) and Y (frame) amplifier gains together. A switch was installed so that the ganged potentiometer could be replaced by independent X and Y gain controls each with a separate ten turn potentiometer. These controls are essential if good alignment is to be achieved.

2.4.3 Hysteresis Effects

The position of the beam on the Philips SEM moves rather slowly when travelling long distances because of the inductance of the scan coils. At 80x magnification the beam takes about 12ms to settle to within 1 pixel of the coordinates on the DACs if the jump is over half the field. Since the beam is unblanked and scanning commences immediately after loading of the coordinates, distortion of pattern elements results when the distance between rectangles is large. A single pixel width line becomes bent, with the beginning of the line closer to the previous rectangle when the jump is perpendicular to the line. If a large downward vertical jump is made to a large block, the scan generator starts the reversing scan required to expose the rectangle before the beam has reached the correct Y-position. This results in a zig-zag line with about 3 reversals above the top of a 100 micron block, which meets the main body of the block when the beam at last catches up with the DACs. The effect is important even for small jumps of 100 pixels, because the tops of rectangles are effectively scanned at a lower exposure than was specified, since the vertical pixel size is lengthened by the effect. It was found that poor liftoff resulted because of bad

resist profiles at the top edges if hysteresis effects were not accounted for.

Several techniques in pattern design were found to be useful to get rid of unwanted hysteresis effects. If a single pixel line was required to be drawn after a perpendicular jump, a small delaying rectangle was drawn close to the starting point of the critical rectangle. The positional distortion took place within the delaying rectangle, and since the jump from it to the critical rectangle was small, minimal distortion resulted. This approach was used in the patterns used to test alignment, where positional accuracy was paramount (Section 5.3). In the case of a jump to a large block, the delaying rectangle was placed inside the block so that the top edge distortion on it was hidden from view (see Figs. 6.6,6.7,6.8 for examples at the bottom of each column of contacts). Finally, if a column of blocks was being exposed, it was exposed from bottom to top so that the distortion resulted in extra exposure inside each block which did not affect liftoff drastically.

Reference

[2.1] "Electron Beam Technology in Microelectronic Fabrication", p. 186 Ed. G. Brewer, Academic Press, (1980).

CHAPTER 3 FABRICATION OF THIN SUBSTRATES

3.1 INTRODUCTION

This chapter deals with the fabrication procedures required to produce membranes of silicon nitride and GaAs suitable for use as substrates in very high resolution electron beam lithography. Details of the fabrication procedure for each is discussed including wafer structure, mask design, masking materials, etches and their charactaristics, etching procedures, and quality and reliabilty of the resulting membranes.

3.2 FABRICATION OF SILICON NITRIDE MEMBRANES

3.2.1 Introduction

The fabrication procedure for silicon nitride membranes was similar to one used by Sedgewick [3.1], except that Si_3N_4 was used to mask the backs of wafers rather than SiO_2. Silicon wafers coated on both sides with silicon nitride had windows opened into the nitride on the back face using phosphoric acid and a photoresist mask. Then the silicon was etched with sodium hydroxide solution, a preferential etch that forms truncated pyramidal wells which terminate on the nitride at the front face, thus forming the membranes (Fig. 3.1).

3.2.2 Fabrication Procedure

The method of membrane fabrication was designed with the aim of keeping the polished surface of the wafers as free from particulate contamination as possible, since this causes pinholes in the protective layer of resist on the polished face of the wafers which in turn leads to membrane failure during etching, or

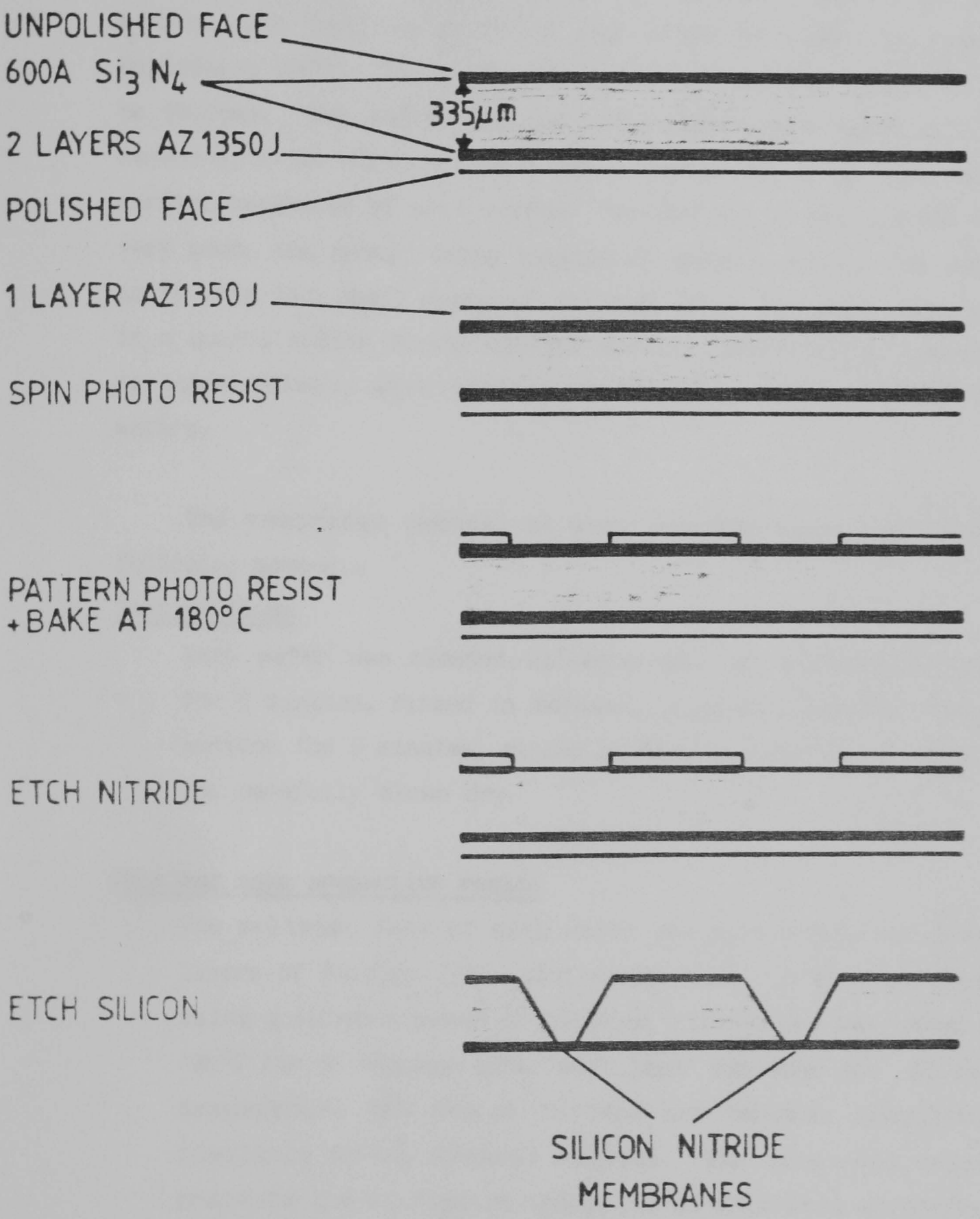


Fig. 3.1 Fabrication process for silicon nitride membranes.

to the formation of etch pits which cause undesirable striations in the PMMA electron resist when it is spun.

Three inch silicon wafers were received from the SERC Microfabrication Facility at Edinburgh University coated on both sides with silicon nitride deposited by Chemical Vapour Deposition (CVD). The thickness of the nitride was specified to be 50-70nm. The wafers had one pre-polished face which will be referred to here as the front face, while the back face had a surface roughness of +/-1 micron. The quality of the nitride was very good, the colour being completely uniform across the wafer apart from two small areas at the edge where the wafer was held in a quartz holder during nitride growth. Generally no pinholes or other defects were visible on the polished surface of the wafers.

The membranes themselves were prepared according to the following method:-

Initial Clean

Each wafer was cleaned ultrasonically in trichloroethylene for 5 minutes, rinsed in methanol, cleaned ultrasonically in acetone for 5 minutes, rinsed in flowing water for 5 minutes and carefully blown dry.

Spin and bake protective resist

The polished face of each wafer was spin-coated with two layers of Shipley 1350J photoresist each 1.5 microns thick, using a spinner speed of 2000rpm. The wafer was baked at 180°C for 30 minutes after each layer was spun on. At this temperature the resist hardens and becomes completely insoluble in any organic solvent. The toughened resist protects the nitride on the polished face from scratches, dust after scribing, and the nitride etch.

Break wafer into chips

Each wafer was scribed on a Dage Precima motorised scriber into six "chips" each 17mm x 34mm. Three considerations

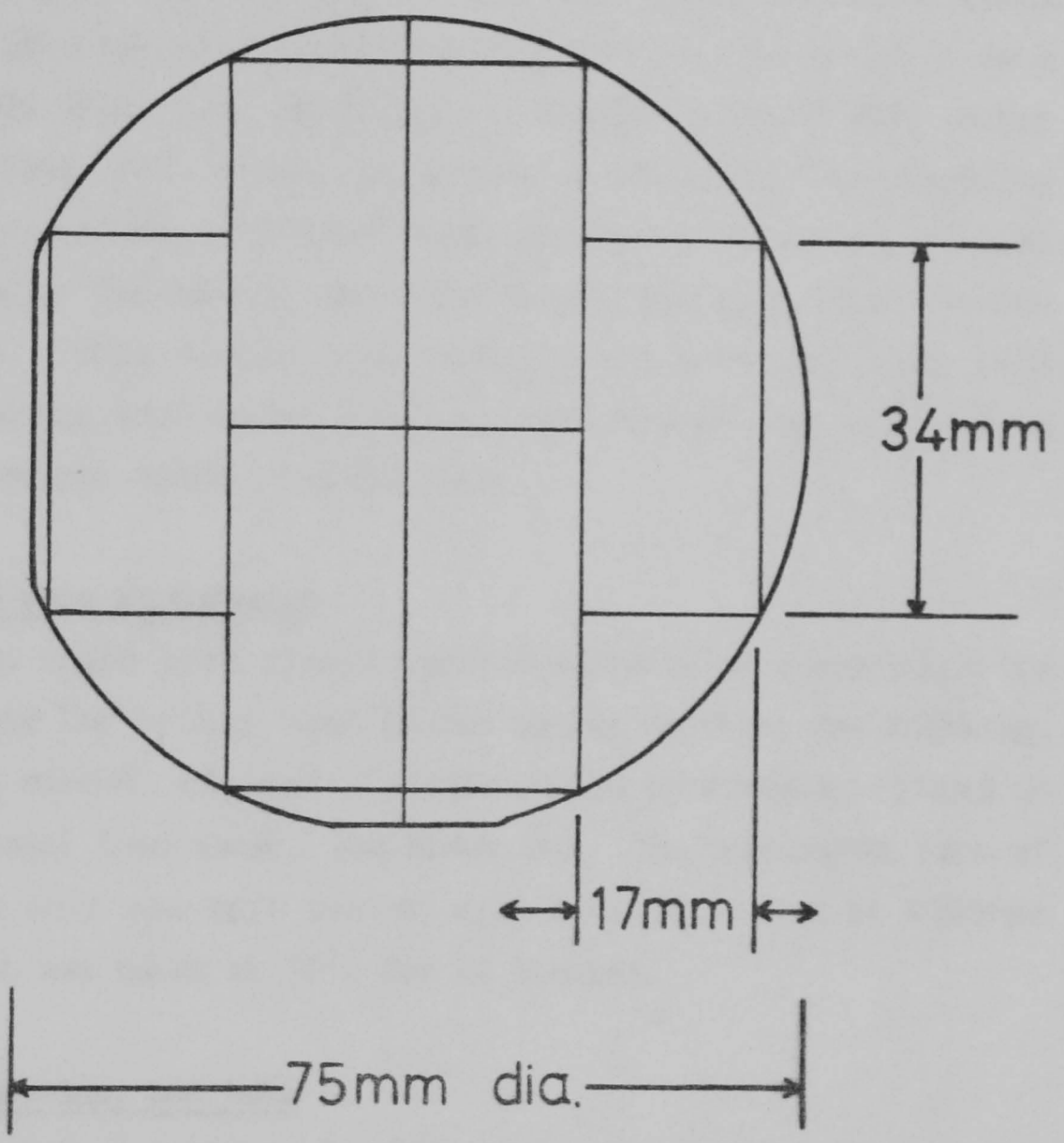


Fig. 3.2 Sectioning a 75mm wafer into six chips for membrane fabrication.

were taken into account in deciding the dimensions of the chips: 1) the chips should be large so that many membranes could be made at once, 2) the chip must fit through the opening in the etching apparatus (22mm diameter), 3) the chip shape must not cause too much waste when a wafer is scribed up. The eventual choice resulted in the wafer being cut into the six biggest rectangles that can be cut from a circle (Fig. 3.2), resulting in wastage of about 25%. After scribing, each wafer was broken up into chips by inverting it on a piece of filter paper and pressing with a scalpel blade on the edge of the wafer at the position of the scribe line. This method was found to be more reliable than breaking the wafer over a long sharp edge such as a microscope slide or steel rule.

Clean and spin photoresist

The chips were cleaned ultrasonically in soapy water to remove the silicon dust caused during scribing and breaking, then rinsed, cleaned ultrasonically in acetone, rinsed in methanol then water, and blown dry. The unpolished face of each chip was spin coated with AZ1350J resist at 4000rpm which was baked at 90°C for 20 minutes.

Expose, develop, and bake

The pattern shown in Fig. 3.3 was exposed by contact printing on a Dage Precima mask aligner. The mask contains eighteen groups of four openings each 640 microns square through which wells were etched through to the nitride at the front face so forming membranes about 120 microns square, and a grid of channels 400 microns wide which formed channels with a depth of about 90% of the thickness of the wafer during etching and were used to help separate the chip into eighteen 5 x 5mm dies. The resist image was developed in a 1:1 mixture of AZ developer and water for one minute. After rinsing in deionised water the chips were blown dry and baked at 180°C for 30 minutes in order to increase the adhesion of the photoresist in the hot silicon

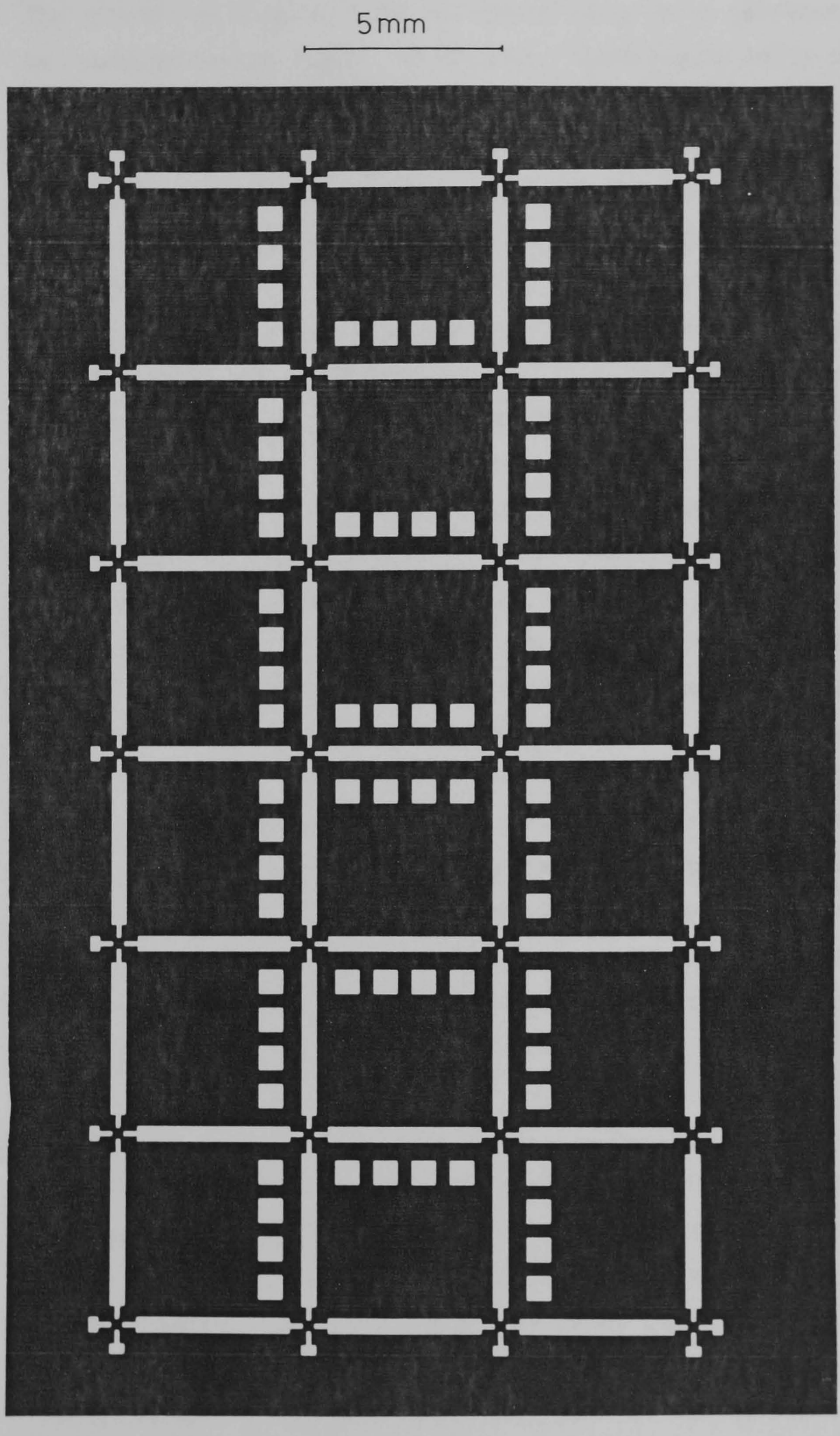


Fig. 3.3 Mask used for making membranes. Square openings are 640 microns wide, and channels are 400 microns wide.

nitride etch.

Etch windows in the silicon nitride

The nitride etching and the silicon etching were performed in round-bottommed flasks which were heated electrically so that the etchant was kept boiling and refluxing as shown in Fig. 3.4. This served to keep the temperature of the etch constant. The flasks had two openings; the larger one was used to insert and remove the samples and on the other was placed the refluxing column which consisted of two condensers, the lower with a spiral coil inside it and the upper with a straight condenser. All the joints were smeared with vaccuum grease to prevent seizing.

The phosphoric acid was supplied at a concentration of about 85%. At this concentration the boiling temperature was 185°C. Water was added until the boiling temperature fell to 146°C at which temperature the etch rate of the nitride was 3.5nm/min and the rate of undercutting was sufficiently small that less than 1 micron of undercut occurred during etching. Fig. 3.5 shows the variation of etch rate with temperature [3.2]. The chips were etched six at a time for 17-22 minutes depending on the the thickness of the nitride (about 10% over-etching was used to be sure of etching completely through the nitride). After the allotted time the chips were removed from the etch and rinsed in a beaker of deionised water.

Etch silicon

The silicon etch was sodium hydroxide solution [3.3], again refluxed while boiling. This etch has a ratio of etch rates of (100) and (111) planes of about 7:1, so when a well is etched through a square opening with sides parallel to (110) directions, a pyramidal pit results with sidewalls almost parallel to (111) planes, which are 54.7° to the horizontal. Fig. 3.6 shows the etch rate of (100) planes of silicon (i.e depth of etched wells in a (100) wafer) plotted against etch

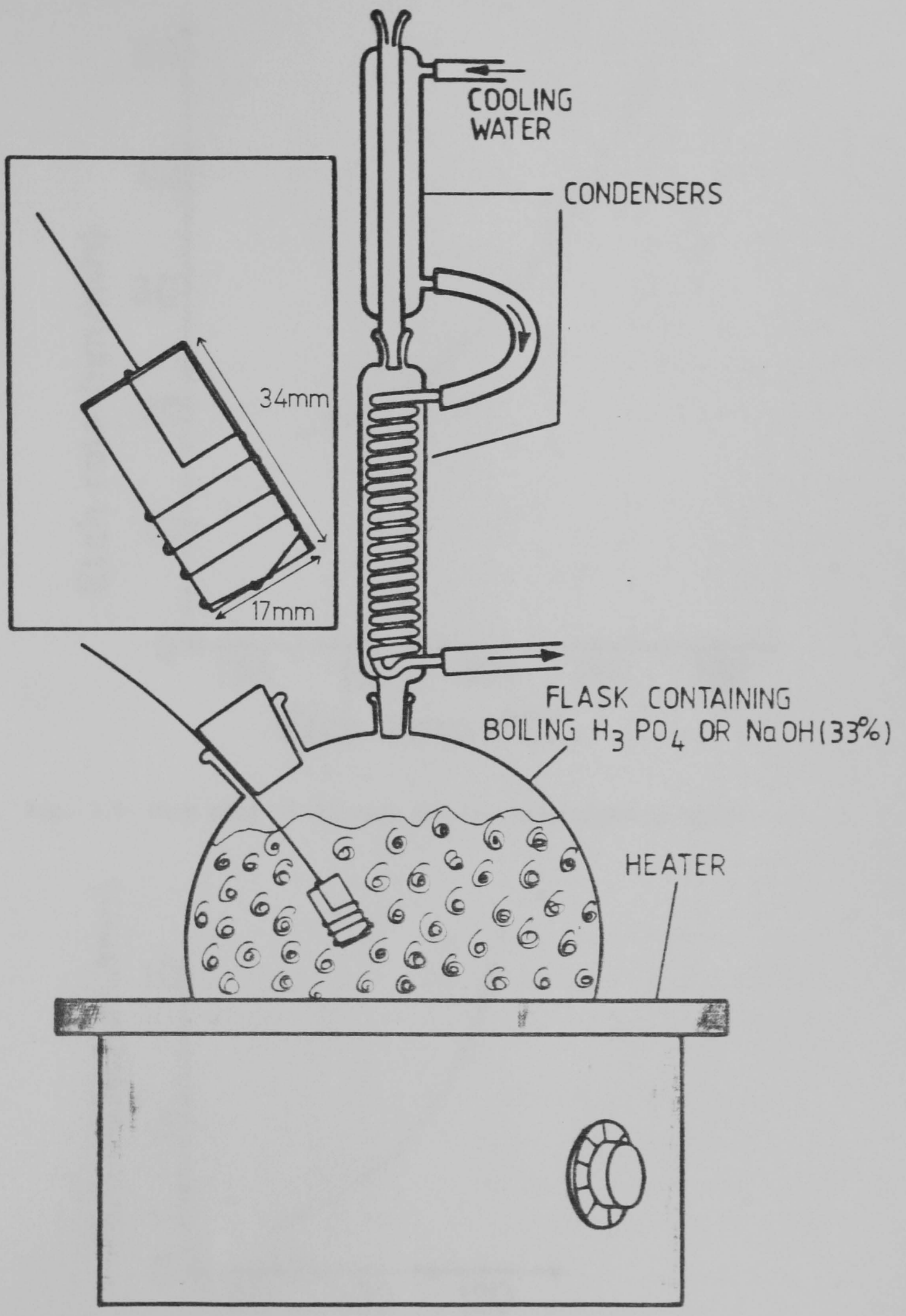


Fig. 3.4 The etching apparatus used for membrane fabrication. Etches were refluxed while boiling to keep temparature constant. Insert shows tungsten wire chip holder.

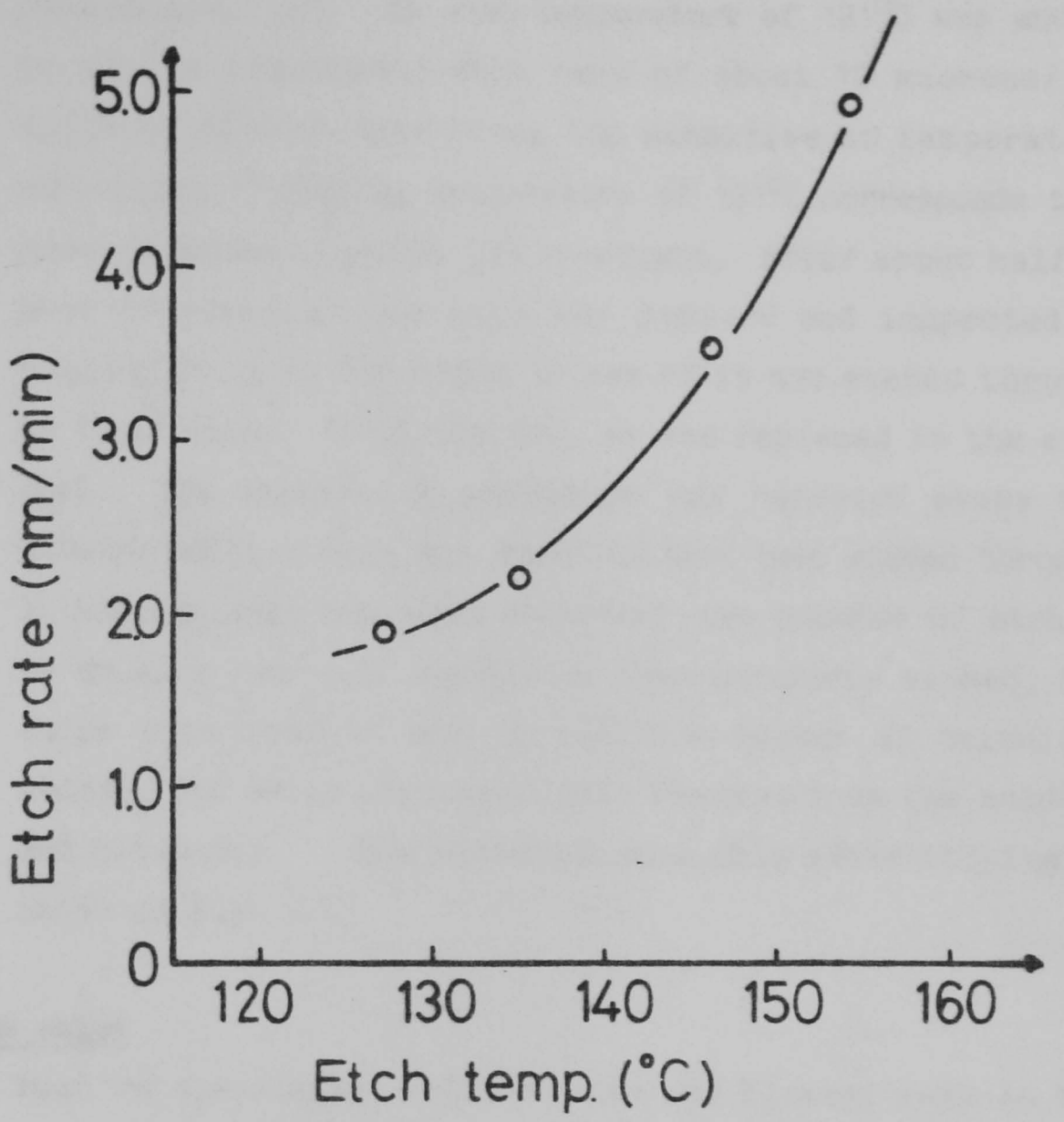


Fig. 3.5 Etch rate of silicon nitride in phosphoric acid.

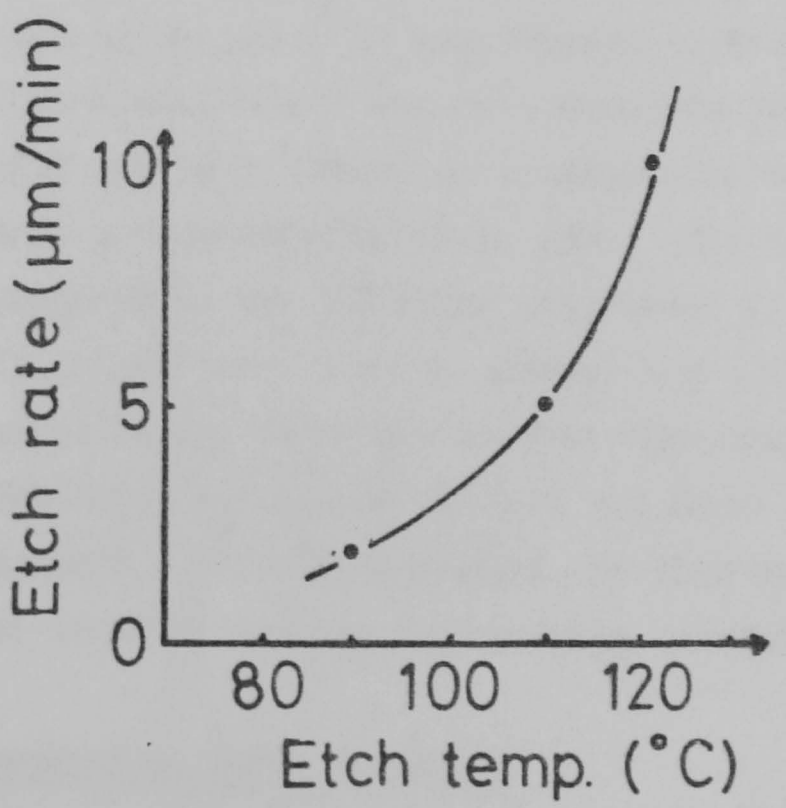


Fig. 3.6 Etch rate of (100) silicon in NaOH solution.

temperature [3.2]. An etch temperature of 121°C was chosen to give a reasonable etch rate of about 10 microns/min without the etch rate being too sensitive to temperature variations. A boiling temperature of 121°C corresponds to a concentration of about 33% by weight. After about half an hour of etching, one chip was removed and inspected by holding it up to the light to see if it had etched through to front face. If it had not, it was replaced in the etch bath. The inspection procedure was repeated every two minutes until a chip was found to have been etched through. It was replaced and after a further two minutes of etching to be sure that all membranes were properly etched, the chips were removed and rinsed in a beaker of deionised water. The chips were carefully removed from the holders and blown dry. The underside of a chip after etching is shown in Fig. 3.7.

Clean chips

Most of the resist lost adhesion and floated away in the sodium hydroxide solution within the first few seconds of immersion, but after rinsing and blowing dry there was still a residual film left on the chips. The chips were stored in this state until it was desired to break the chips up into individual "dies" and spin electron resist on them. Each chip was then cleaned in a mixture of concentrated sulphuric acid and hydrogen peroxide (30%). The cleaning solution was prepared in the following way; about 20 ml of sulphuric acid is poured into a 40 ml beaker and about 5-6 ml hydrogen peroxide was added and stirred vigourously with a glass rod. The chip was placed in the solution and left for 10-15 minutes. The chip was rinsed in flowing deionised water for at least 10 minutes before being carefully blown dry.

Spin protective layer of resist

Prior to breaking the chip up into 18 dies, the front face was protected from the dust created during breaking by spinning on a layer of AZ1350J at 4000rpm.

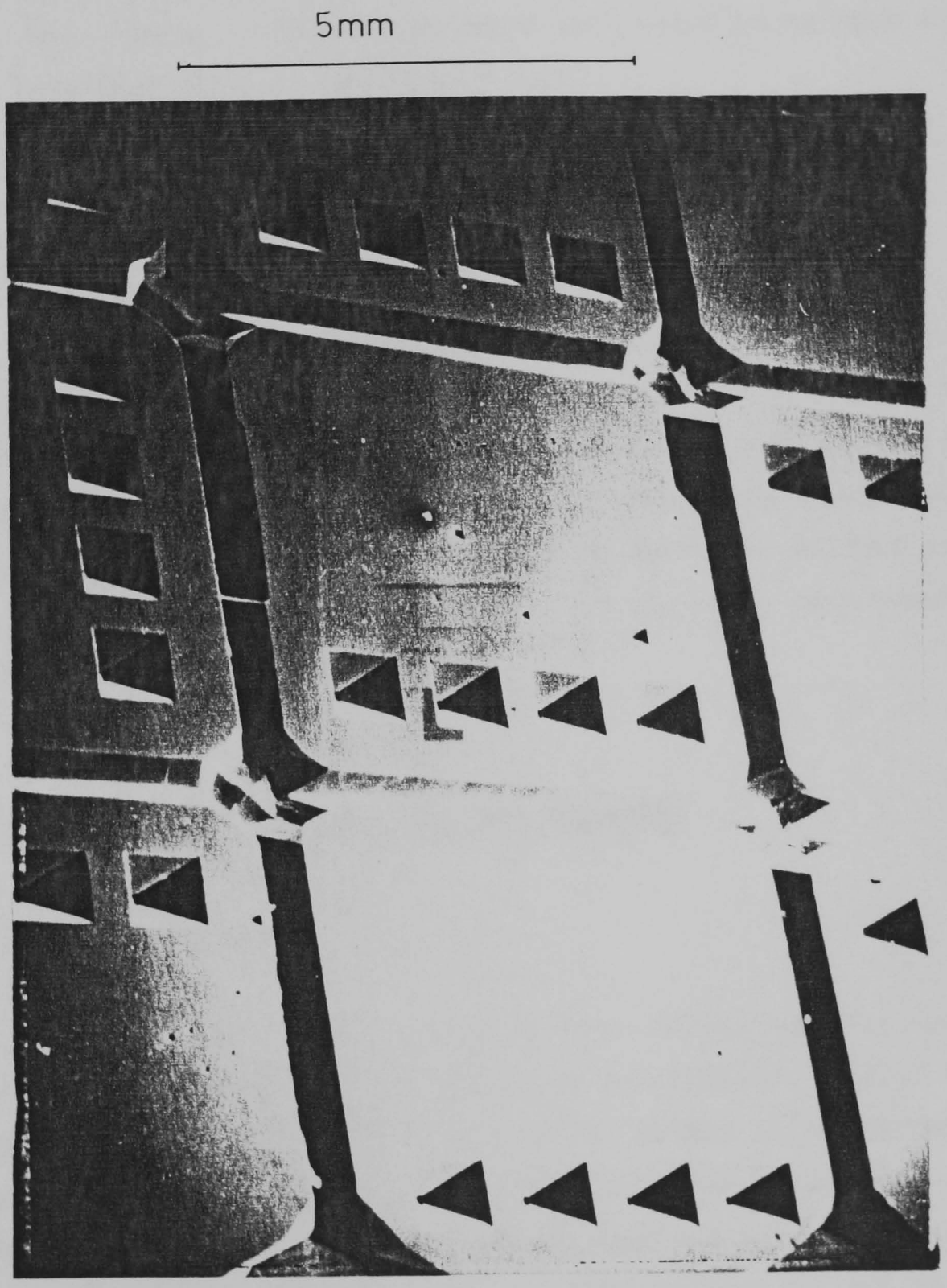


Fig. 3.7 SEM micrograph of etched chip.

Break chip into dies

The dies were separated from each other by breaking the chip over a sharp edge. This was made easier by nicking the edge of the chip with a diamond scriber on the front face above each etched V-groove. The protective resist coating was removed in acetone, the dies were rinsed in IPA and blown dry. The individual dies were then ready to be spun with electron resist (Section 4.1.1.1).

3.2.3 Reliability and Quality

When the above procedure was followed carefully, the success rate for silicon nitride membrane fabrication (ratio of unbroken membranes to number of membranes attempted) was greater than 99%. The membranes were nearly always approximately square with straight edges, but occasionally jogs of about 2-3 microns could be seen. The membranes were flat and normally contained no pinholes or other defects.

3.3 FABRICATION PROCEDURE FOR GAAS MEMBRANES

3.3.1 Introduction

The process of GaAs membrane fabrication was similar in concept to that of nitride membrane fabrication; wells were etched from the back face of a wafer to leave a thin membrane at the front. However, for the case of nitride membranes, the membranes are of a different material from the bulk of the wafer. It is easy to find an etch that will etch silicon but not silicon nitride (and vice versa). However, if a membrane of GaAs is required, there two problems. Firstly, since the membrane has to be of device quality, it has to be single crystal and lattice matched (under no compressive or tensile stress). This can only be achieved if the GaAs layer that is to form the membrane is

grown epitaxially on a GaAs substrate. Secondly, it would be quite impractical to attempt to etch through a wafer 100 microns thick and expect to be able to stop etching at a uniform distance of 0.05 microns from the front face in order to form the membranes. This problem was solved by growing intermediate etch stop layers of GaAlAs. There are several groups of etches that etch GaAs but not $Ga_{1-x}Al_xAs$ and vice versa for certain values of x. The general strategy in GaAs membrane fabrication was first to grow alternate layers of GaAs and GaAlAs on a wafer, the final layer being GaAs of the thickness and doping required for the membranes, and then to etch through a window masked on the back face towards the front face with etches that halted on successive etch stop layers until the last layer of GaAs remained to give a membrane supported only at the edges (Fig. 3.8) The main tasks in this part of the work were to find suitable etches to etch wells and stop on the epitaxial layers, and masking materials to resist all the etches.

3.3.2 Etches Tested for Suitability in Membrane Fabrication

At first it was thought that it would be possible to find an etch that could be used to etch all the way from the back face to the front and would stop on the first GaAlAs layer. In the event (see below), it was found that the selective etches that were tried were unsuitable for etching deep wells, and the approach was changed to one in which the majority of the well was etched using a fast non-selective etch, and then selective etches were used in turn to first reach and then etch the etch stop layers.

The etches were tested by etching wells through a mask formed by conventional photolithography. The mask used for testing the various etches is shown in Fig. 3.9. It contains windows of various dimensions. For each etch a suitable masking

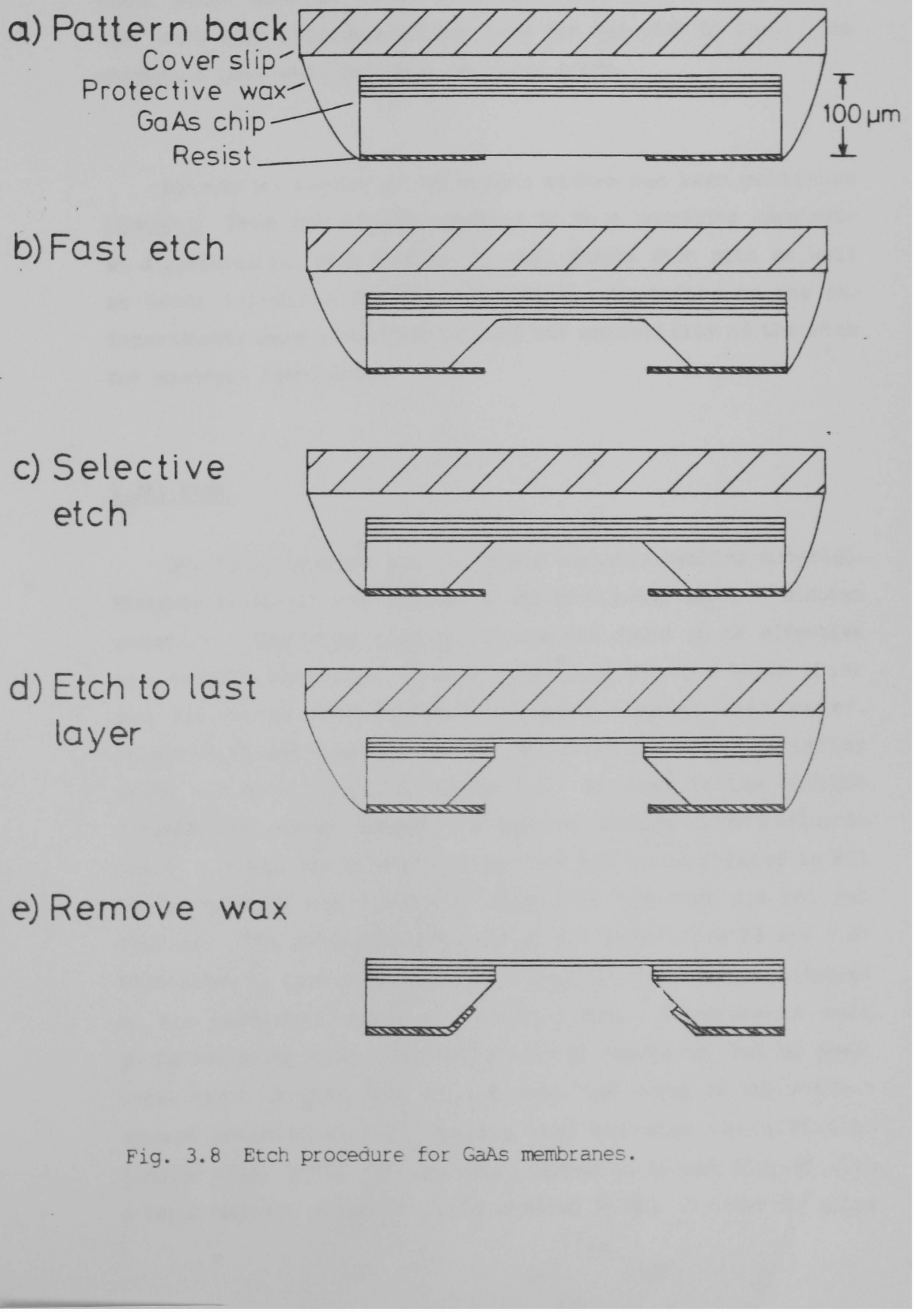


Fig. 3.8 Etch procedure for GaAs membranes.

material had to be found. The etched depths were measured using the vertical movement of an optical microscope.

Etches were used freshly made up usually in quantities of about 30ml. Several etches generated heat when freshly mixed; they were used without allowing time for the etch to cool. The chemicals used were normally 'Aristar' grade.

Data on a number of selective etches has been published [3.4,3.5]. From these I_2/KI appeared to be a promising candidate as a preferential etch leading to well shaped etch pits as well as being selective for GaAs or GaAlAs, depending on the pH. Experiments were conducted to test the suitability of the etch for membrane fabrication.

I_2/KI Etch

The first problem was to find a suitable masking material. Photoresist itself was removed by the etch after about 5 minutes immersion. Sputtered silicon dioxide was found to be effective as a masking material. A layer of silicon dioxide 100nm thick was sputtered onto the surface of a cleaned GaAs wafer. Photoresist was then spun on and patterned by contact printing using the mask shown in Figure 3.9. Windows in the silicon dioxide were opened through the mask by etching in hydrofluoric acid. It was found that the surface had to be cleaned in HCl before etching was attempted otherwise the etch did not get started. The published mixture of 0.1 mole/litre KI and 0.04 mole/litre I_2 gave an etch rate of only 0.1 microns/min compared to the published value of 1 micron/min. Experiments were performed using higher concentrations of reactants, but no great improvement in etch rate was obtained. The edges of the windows showed enhanced etching showing that the etch was diffusion limited (Fig. 3.10). Etched wells tended to become blocked with a brown deposit which had to be removed in HCl in order to allow

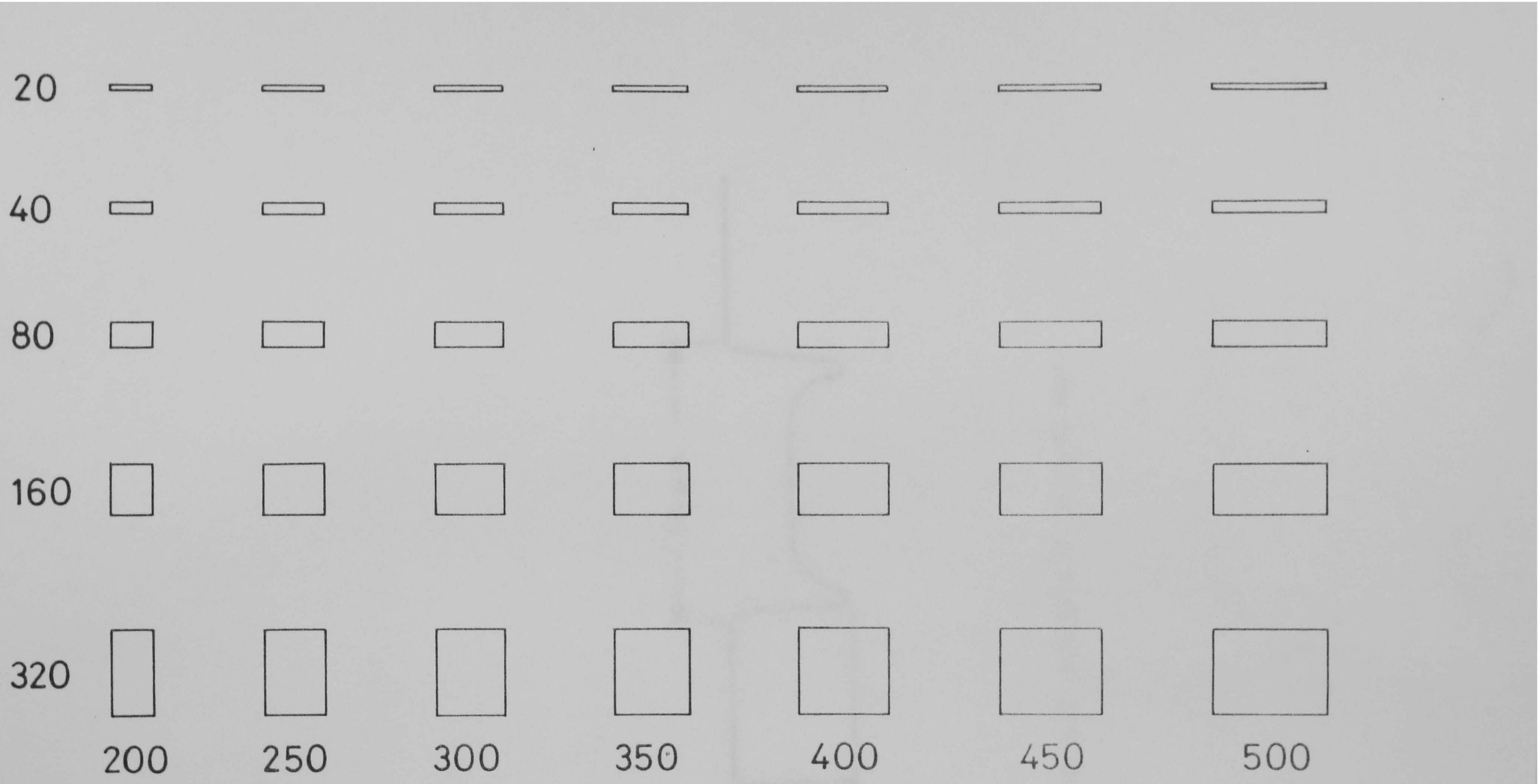


Fig. 3.9 Test mask used for finding etch rates of GaAs etches.

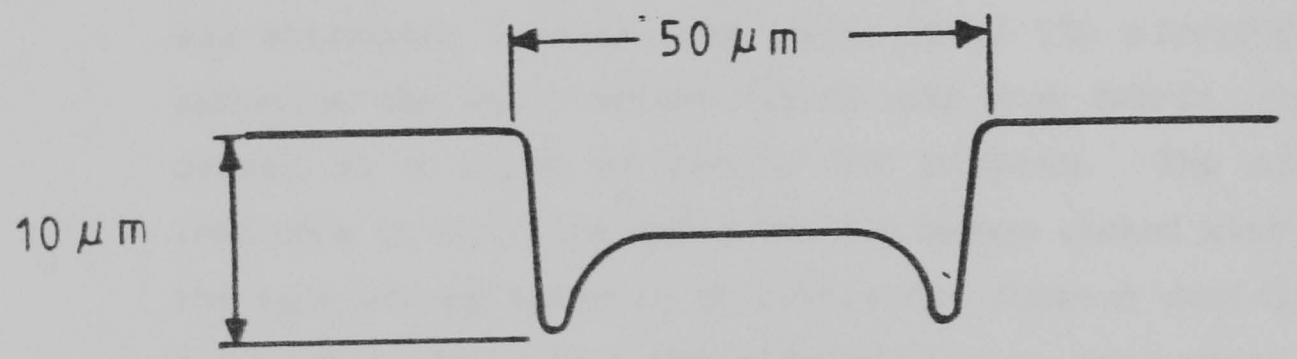


Fig. 3.10 Etch profile for I_2/KI etch.

etching to continue. Agitation by a magnetic stirrer was used in the experiments of Tijburg [3.4] and the fact that it was not used here might explain the discrepancies between etch rates that were found. The other difference in procedure was that in the published experiments a depth of only a few microns was etched whereas here a depth of tens of microns was required.

95:5 Hydrogen peroxide: Ammonia Etch

This etch is isotropic and selective [3.5,3.6], the etch rate for $Ga_{1-x}Al_xAs$ slowing exponentially with aluminium concentration, so that the etch has virtually stopped for $x>0.3$.

Etch tests were done by using photoresist as a mask. The pattern was the same as that described above (Fig. 3.9). When it was attempted to etch deep wells (over 100 microns) without agitation the wells became filled with grey debris, and etching ceased at a depth of around 100 microns. The debris was insoluble in HCl. The wells did not become choked with debris if the etch was agitated in an ultrasonic cleaner during etching, but even in this case the sidewalls were very rough and were covered in grey particulate deposits.

The two selective etches that were examined were unsuitable for etching deep wells, and so it was decided to etch most of the depth of wells with a non-selective etch, before changing to 95:5 etch to etch the last few microns.

1:8:x Sulphuric acid: Hydrogen peroxide: water mixtures

This family of etches are anisotropic and non-selective. [3.7]. Photoresist can be used as a masking material. The etch rates of some of these etches were measured. An area of a piece of GaAs was masked with wax and each piece was etched until a measurable step resulted. The steps were measured with an

optical microscope or a Talystep as appropriate. Figure 3.11 shows the variation of etch rate with water concentration.

Fig. 3.12 shows pependicular cleaved cross-sections through an well etched in 1:8:1 with edges parallel to the (011) and (01$\bar{1}$) directions. The sections show that the anisotropy of the GaAs crystal structure has a strong effect on the etching process. The (01$\bar{1}$) sections show straight sloping walls down to the well bottom which is flat. The sloping faces are at an angle of 37^{o} to the horizontal. This is similar to the angle found by Shaw [3.7] for the lower half of the walls on his etched channels. He found that the upper part of the walls made an angle of 55^{o} to the horizontal, and corresponded to A(111) planes ((111) planes with Ga atoms exposed and generally known to be the slowest etching planes in GaAs). The change of angle does not appear here. The (011) section shows a semicircular cross-section rather than the obtuse-angled sidewalls found by Shaw. The slight curve close to the top of the face in both sections shown in Fig. 3.12 indicates that undercutting of the mask has been an important factor here in determining the well shape and it has probably been responsible for the different etch profiles found here and by Shaw, who used silicon nitride as the etch mask rather than photoresist. Details of how the etch characteristics of 1:8:1 influenced the mask design for membranes are given in Section 3.4.3.2.

3.3.4 Final Selection of Etches

The final selection of etches was;

-**1:8:1** $H_2O_2:H_2O:H_2SO_4$ to etch wells to within 10 microns of the etch stop layers

-**95:5** $H_2O_2:NH_3$**(30%)** to etch to the first etch stop layer, and subsequently to etch the second such layer and stop on the

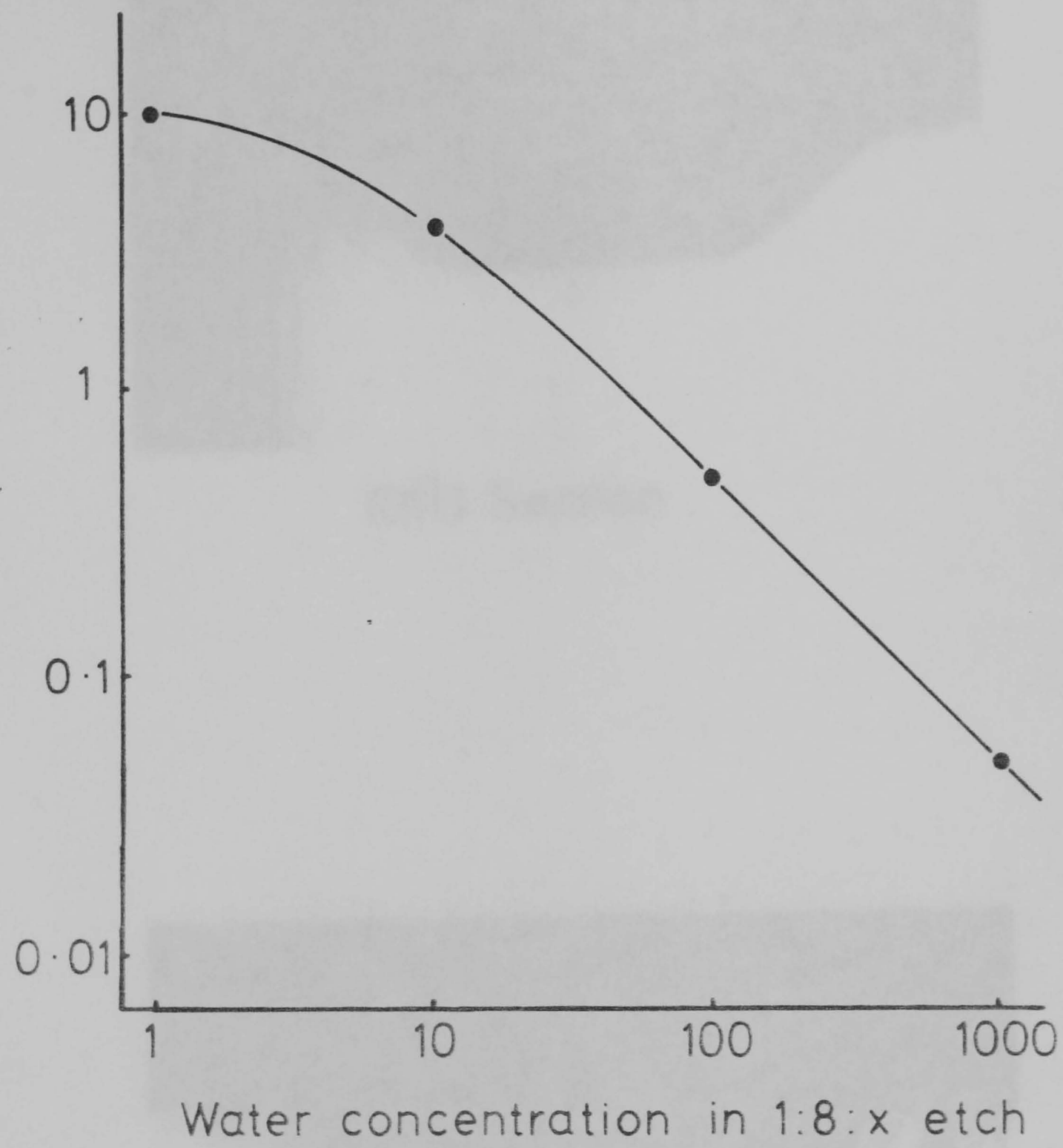
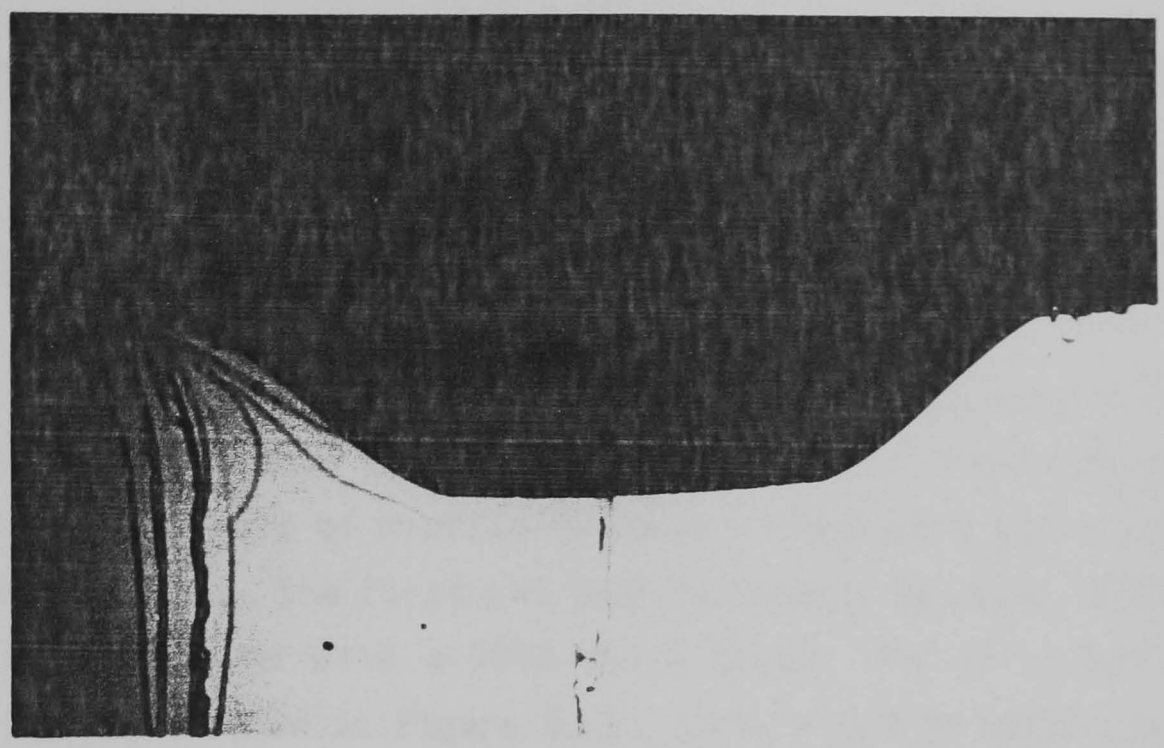
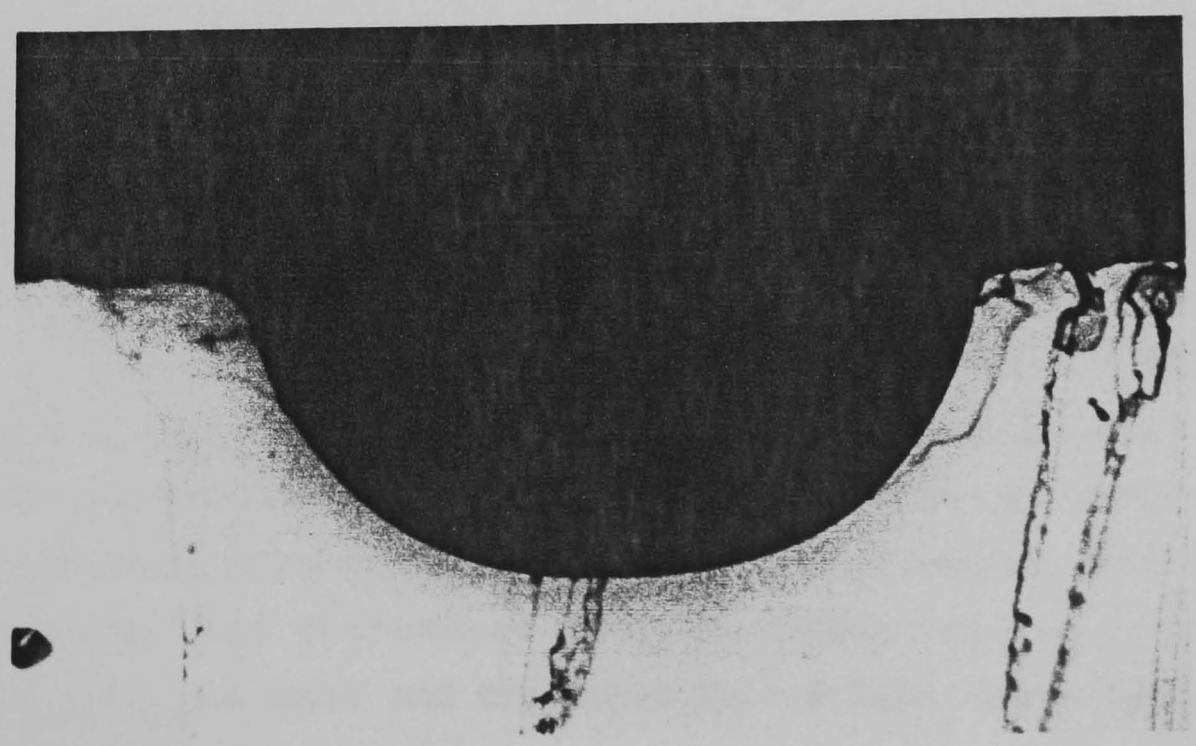


Fig. 3.11 Etch rate of 1:8:x H_2SO_4:H_2O_2:H_2O system for GaAs.



(01$\bar{1}$) Section



(011) Section

Fig. 3.12 Cross-sections through wells etched in GaAs with 1:8:1 etch. The depth of the well was 120 microns.

third.

-**20% HF (hydrofluoric acid)** to etch the $Ga_{0.4}Al_{0.6}As$ layers [3.6].

3.3.5 MOCVD Wafer Growth

The heterostructures were grown by metal-organic chemical vapour deposition at the SERC III-V Compound Growth Facility at the University of Sheffield. Three wafers were grown for these experiments, the first two with the top GaAs layer 1000A thick and the third with a 500A thick layer. The structure of the wafers is shown in Figure 3.13. First a 0.05um buffer layer was grown to provide a growing surface largely free from defects for the first layer of $Ga_{0.4}Al_{0.6}As$. On top of this four layers were grown, viz. 1um $Ga_{0.4}Al_{0.6}As$, 0.3um GaAs, 0.15um $Ga_{0.4}Al_{0.6}As$ all nominally undoped, and 0.1um or 0.05um GaAs doped n-type with 10^{17} carriers cm^{-3}. The lower three layers act purely as etch stops while the top layer will become the thin substrate.

A full account of the growth procedure appears in [3.8]. Particular attention was paid to attaining a very low concentration of O_2 in the reactor, since the growth of $Ga_{0.4}Al_{0.6}As$ is very sensistive to oxygen contamination; the system was purged with H_2 for at least 18 hours before growth was started.

The etch stop layers were grown at a temperature of 820°C to ensure that the Al-containing layers were of good quality, while the active layer was grown at 660°C in order to achieve the relatively high doping levels required. A cooling period of 9 minutes was introduced between depositing the uppermost $Ga_{0.4}Al_{0.6}As$ layer and the doped active GaAs layer. Subsquent SIMS analysis showed that during this time lapse residual trimethyl aluminium was purged from the reactor resulting in an abrupt transition between the two layers. Although the structure required five epitaxial deposits the resulting morphology was

GaAs n-type	0.05μm
$Ga_{0.4}Al_{0.6}As$	0.15μm
GaAs	0.3 μm
$Ga_{0.4}Al_{0.6}As$	1 μm
GaAs buffer	0.05μm
SI substrate	400μm

Fig. 3.13 Structure of wafer used for fabrication of GaAs membranes. Layers were grown by Metal-Organic Chemical Vapour Deposition (MOCVD) on a semi-insulating substrate.

excellent with fewer mesa defects per square centimetre than observed on single GaAs structures of comparable thickness grown in the same system.

The room temperature mobility and free carrier concentration were determined at Sheffield by the Van der Pauw technique using tin dots alloyed under H_2. In the case of the wafer with the 500A thick top layer, the mobility was found to be 1388 $cm^2V^{-1}sec^{-1}$, and the free carrier concentration was $1x10^{17}$ cm^{-3}. A correction for surface depletion [3.8] was applied to the measured carrier concentrations; however no correction was made for the conductivity of the underlying $Ga_{0.4}Al_{0.6}As$/GaAs structure which was nominally undoped. The Al mole fractions of the GaAlAs components of the structure were measured at Sheffield by x-ray double crystal diffractometry using either the [400] or assymetric [115] reflections.

3.3.6 Etching Proccedure

3.3.6.1 Masking Materials

The back face of the wafer had to be patterned with a material that would withstand attack from all three of the etches used during membrane fabrication. A good masking system was found to be 40nm of gold on a thin layer of nichrome pattened by liftoff, with 1.5 microns of AZ1350J spun and patterned on top of the gold. The photoresist was a good mask against 1:8:1, while the gold gave protection against the 95:5.

3.3.6 Etching Procedure

3.3.6.1 Masking Materials

The back face of the wafer had to be patterned with a material that would survive exposure to all three etches. A good

masking system was found to be 42nm of gold on a thin layer of NiCr patterned by liftoff, with 1.5 microns of AZ1350J spun and patterned on top of the gold.

3.3.6.2 Mask Design and Etching of Wells in 1:8:1

It was desired to make membranes with a side length of between 100 and 200um. Once the size of the membranes had been decided upon, the properties of the 1:8:1 etch determined both the thickness of the wafer and the shape of the opening on the back of the wafer. To see why this was the case, consider what would happen during 1:8:1 etching of two windows with dimensions as shown in Fig. 3.14 and defined in resist on a (100) GaAs wafer with the crystal directions also as shown. Across the $(01\bar{1})$ direction the profile will develop in the same way for both windows, with a flat bottom and sidewalls at an angle of about 37^{o} to the horizontal (Fig. 3.15 (a)). The extent of undercut at each side is about 0.95 times the depth etched. The width of the bottom decreases by approximately the same amount as the depth etched. The profiles across the (011) direction, however, will differ depending on the width of the mask opening. For the wide (320um) opening the profile begins with a flat bottom and curved walls, but, as etching proceeds, the profile tends to semicircular (Fig. 3.15 (b)). For the narrow (20um) opening the flat bottomed phase is not seen for large etched depths and the profile is always approximately semicircular (Fig. 3.15 (c)).

It will be seen that the width of a membrane will be crucially influenced by the curvature at the bottom of the well before the selective etches are used. In order to keep the membrane width small, the wafer thickness should be under 120um. However, if the wafer is much thinner than 100um it becomes too fragile to handle. Once the thickness of the wafer has been chosen, the shape of the mask opening required to give the desired membrane size was determined by the etching characteristics of the 1:8:1 etch. A mask opening of 20x200um

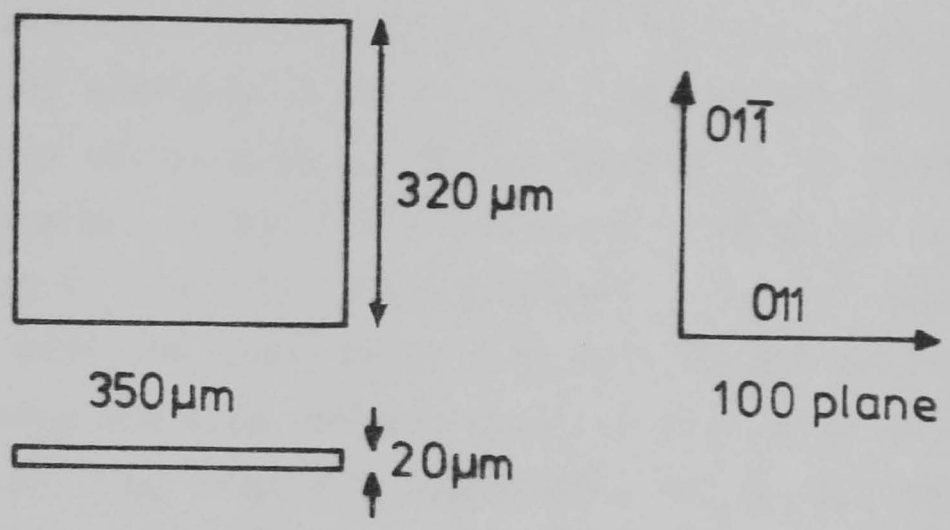


Fig. 3.14 Mask to demostrate profile through wide well and narrow well.

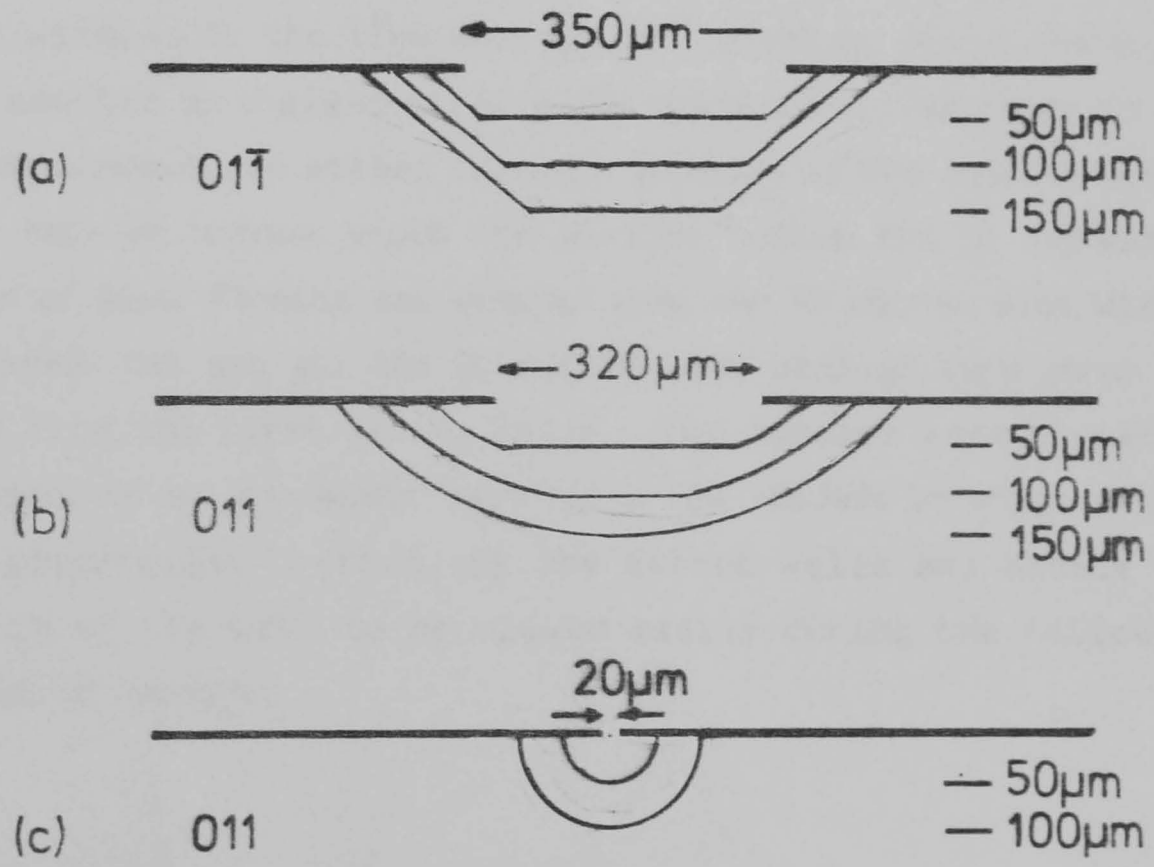


Fig. 3.15 Profiles through wells etched through mask in Fig. 3.14. Narrow well does not show flat bottommed phase and so membrane dimensions depend on wafer thickness.

was used for membrane fabrication in these experiments.

Samples from MOCVD wafers were waxed down onto glass slides to protect the front surface from the etches. The samples were then thinned by hand-polishing with 5um carborundum followed by etching with 1:8:1. The orientation of the sample was found at this point by examining a small cross that had been scribed onto the back face of the sample. The anisotropy of the etch causes the arm parallel to the $(01\bar{1})$ direction to virtually disappear, while width of the other arm increases. The back faces were patterned with the gold/resist composite described in Section 3.3.6.1, using the mask pattern shown in Fig. 3.16, and making sure that the long axis of the membrane windows was parallel to the $(01\bar{1})$ direction. The small windows are where the membranes will be, the line of windows of decreasing width were used to determine when to stop the 1:8:1 etch. The fact that the etch rate through a narrow opening is less than for a wide one was used to find the stop point of the 1:8:1 etch independently of its etch rate. Fig. 3.17 shows a section through the smaller four windows in the line during 1:8:1 etching. Since the wafer was mounted on a glass slide with white wax it was easy to see which windows had etched through. Because of the dependence of etch rate on window width the windows bottom out in decreasing order of size. Etching was stopped when the 40 micron wide window bottomed out and all the 20 micron wide windows were about 10-15um from the first GaAlAs layer. The samples were placed in methanol in an ultrasonic bath for a few seconds in order to tear the gold/resist overhanging the etched wells and enable the bottom of the well to be viewed easily during the following stages of etching.

3.3.6.3 Selective Etch Procedure

After etching in 1:8:1, the wafer was removed from the glass slide by melting the wax and sliding the wafer off. The wafer was then cut up into 2 x 3mm chips each containing two etched wells.

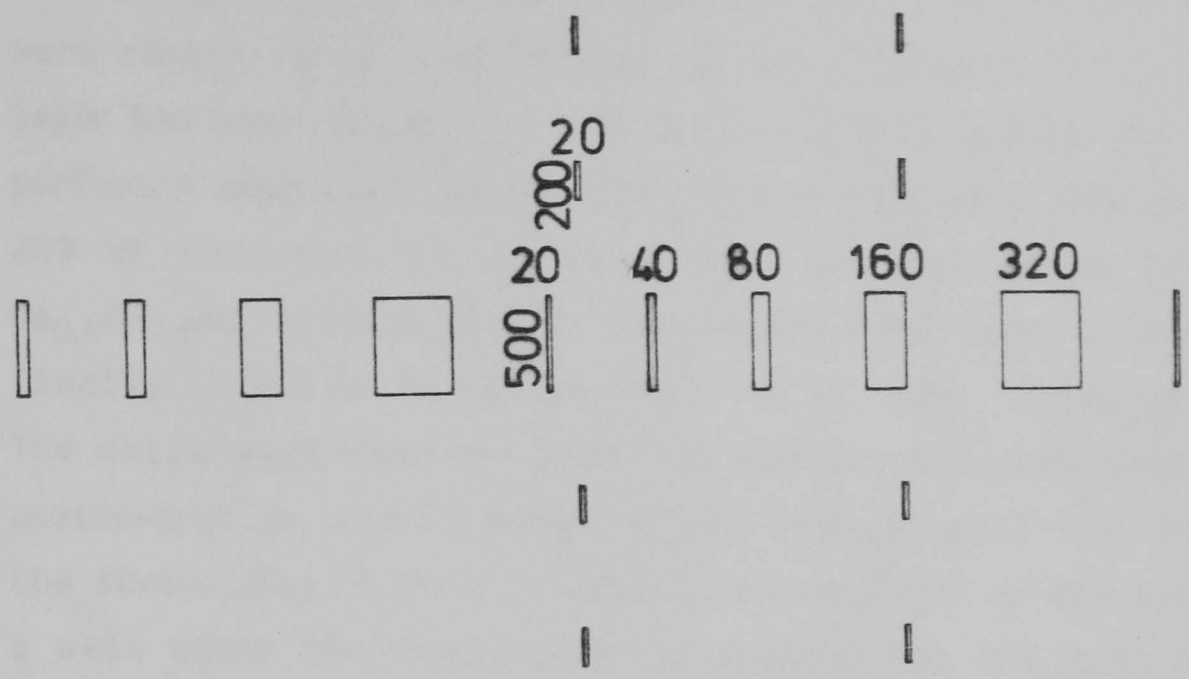


Fig. 3.16 Mask used for membrane fabrication, dimensions of windows marked in microns. The small windows form the membranes, the line of larger wells were used to find when to stop 1:8:1 etch.

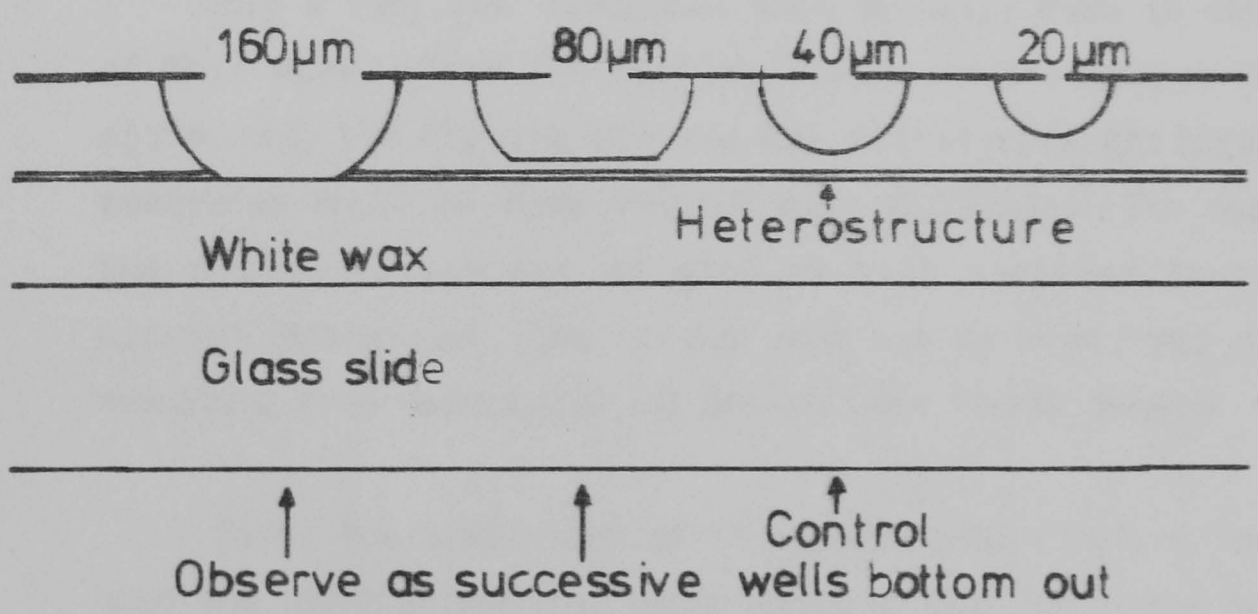


Fig. 3.17 Cross-section through wafer during 1:8:1 etching. Narrow wells etch more slowly than wide ones, etch is stopped when 40 micron well bottoms out.

For ease of handling, the chips were then waxed down separately onto str of plastic cut from a bottle. They were then etched in the 95:5 etch for periods of 5 minutes after each of which the samples were rinsed, dried, and checked to see if the first $Ga_{0.4}Al_{0.6}As$ layer had been reached. At this point the well bottoms had become perfectly smooth and mirror-like. The samples were then etched in 20% HF for 1min 15 sec to remove the 1um thick layer of $Ga_{0.4}Al_{0.6}As$, then in 95:5 to remove the 0.3um layer of GaAs, and finally in 20% HF to remove the 0.15 um layer of $Ga_{0.4}Al_{0.6}As$. The chips were removed from the plastic strip by placing it upside-down in trichloroethylene and waiting until they slid off the strip. Fig. 3.18 is an optical micrograph of the inside of a well after the completion of etching. At the bottom is a 150x150um membrane of GaAs 500A thick. Fig. 3.19 shows a 100x100um membrane viewed from the top. Using a Michelson objective it was found that while the membranes themselves were flat, there was a gentle slope of 1um over 20um out from the edges of the membranes on the solid GaAs. Thus the membranes were raised slightly above the surrounding substrate.

3.3.7 Quality and Reliability of GaAs Membranes

Only a very few membranes were actually made in the course of this work. Once the etching times and procedures had been optimised, the etching process was fairly straightforward and membranes could be made without much difficulty. The quality of the membranes was not as good as that achieved for silicon nitride membranes, the colour was not as even, and pinholes resulting from defects during growth were fairly common.

Since the completion of this work, others have successfully used the above procedure, demonstrating that the above procedure is not too operator dependant. It should be noted however, that the optimum etch times change from wafer to wafer since no two wafers have identical structures (even if it was intended that they should have).

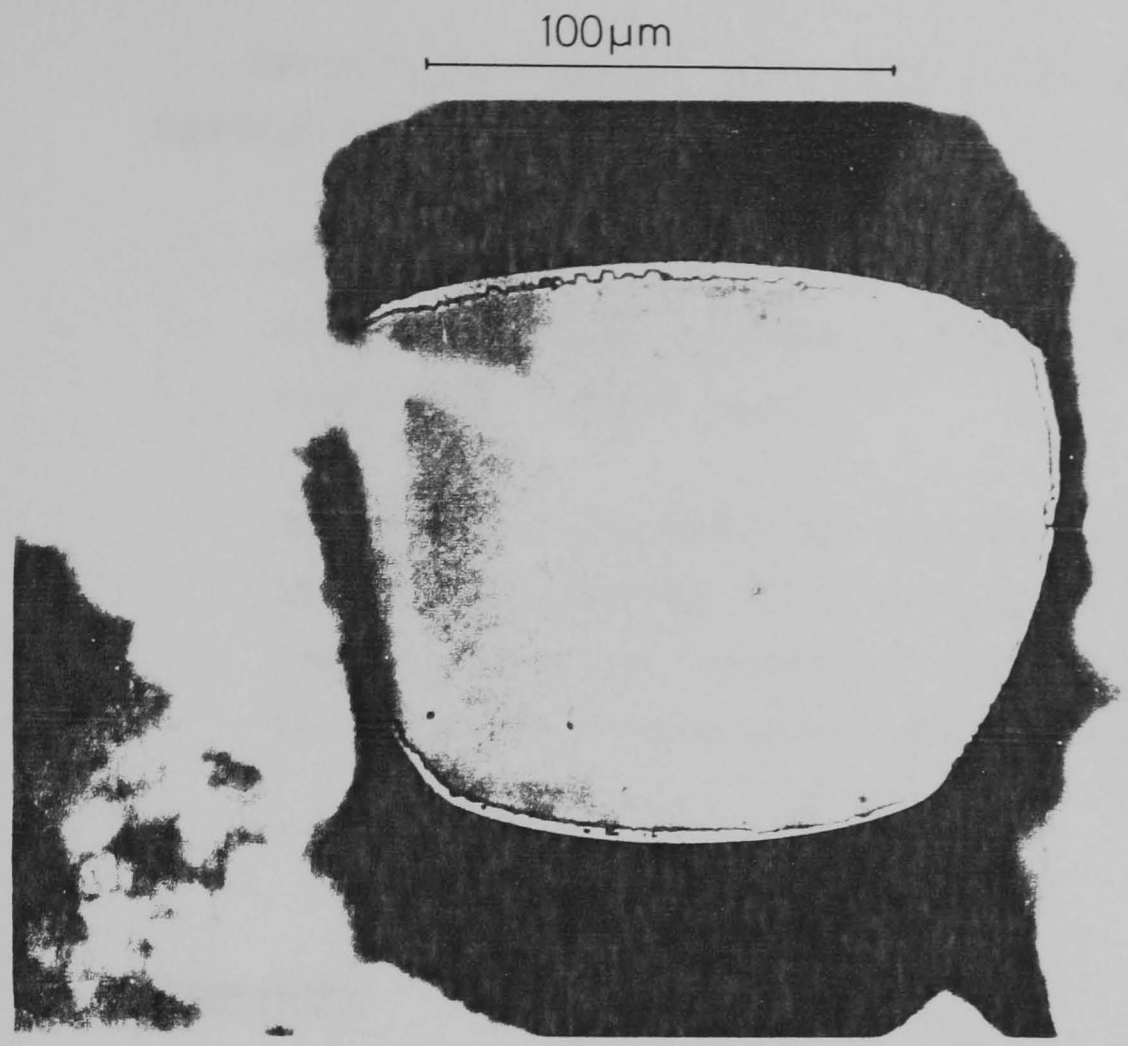


Fig. 3.18 View of a well after completion of etching. Remnants of GaAlAs layer can be seen round edge.

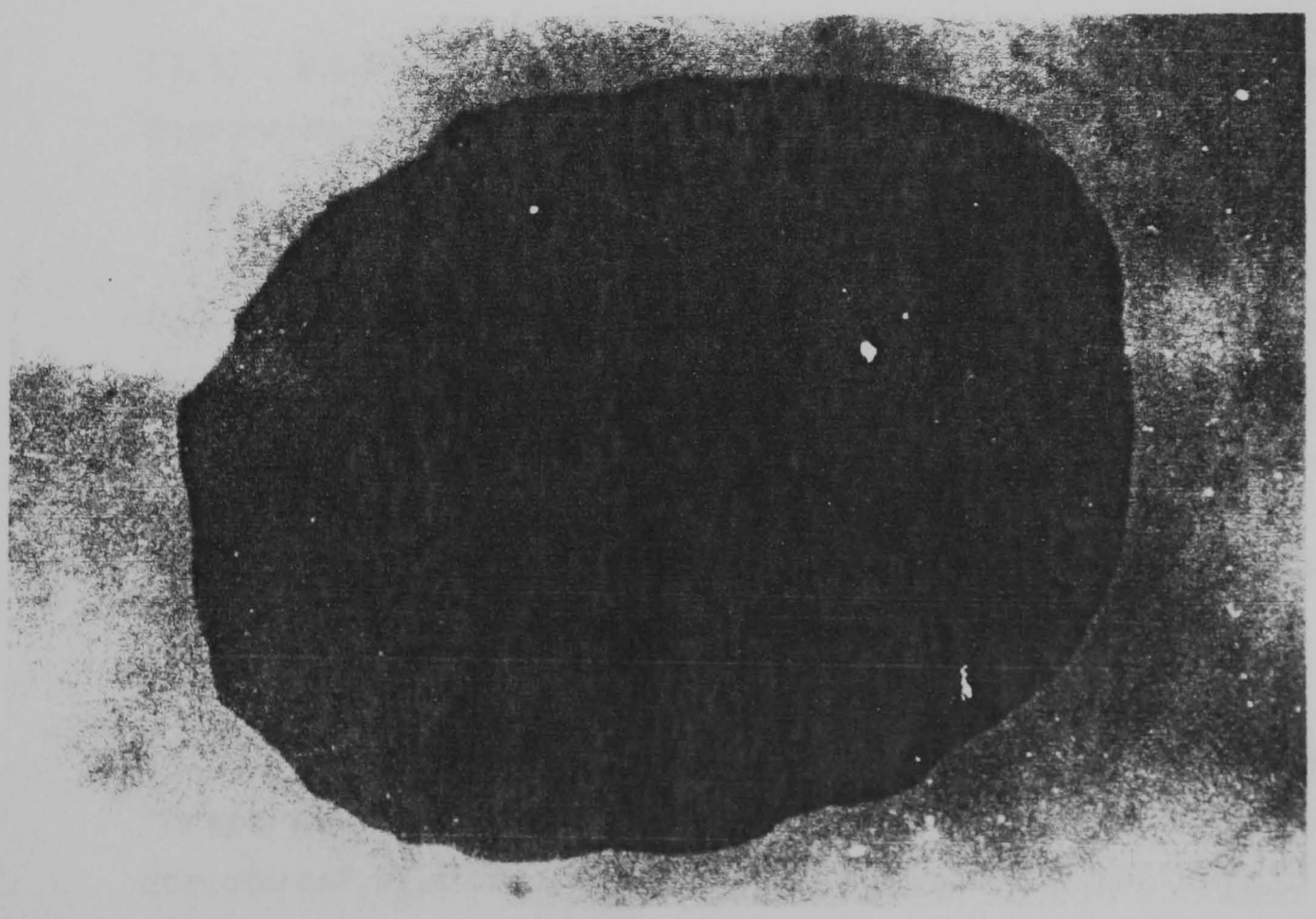


Fig. 3.19 A completed GaAs membrane, 50nm thick, and with dimensions of about 100 x 120 microns.

Certain improvements have also since been made to the fabrication procedure:-

-The masking material that is now used is Ni/Cr-Au-PMMA, with the openings patterned by electron beam exposure and plasma etching of the gold.

-The procedure used above for determining the etched depth of the wells during 1:8:1 etching is no longer used, instead,samples are removed from the etch periodically, and the etch depth checked with an optical microscope.

References

[3.1] T.O.Sedgewick, A.N.Broers, and B.J.Agule, "A Novel Method of Fabrication of Ultrafine Metal Lines by Electron Beams", J. Electrochem. Soc. 119 1769 (1972).

[3.2] M.Losco, unpublished (1980).

[3.3] W.C.Rosvold, W.H.Legat, R.L.Holden, "Air-Gap Isolated Microcircuits - Beam Lead Devices", IEEE Electron Dev. ED-15 640 (1968).

[3.4] R.P.Tijburg and T.van Dongen, "Selective Etching of III-V Compounds with Redox Systems", J. Electrochem. Soc. 123 637 (1976).

[3.5] R.A.Logan and F.K.Reinhart, "Optical Waveguides in GaAs-AlGaAs Epitaxial Layers", J. App. Phys. 44 4172 (1973).

[3.6] R.J.M.Griffiths, I.D.Blenkinsop, and D.W.Wight, "Preparation and Novel Properties of GaAs Layers for Novel FET Structures", Electron. Lett. 15 629 (1979).

[3.7] D.W.Shaw "Localised GaAs Etching with Acidic Hydrogen Peroxide Solutions", J. Electrochem. Soc. 128 874 (1981).

[3.8] S.Mackie, S.P.Beaumont, and C.D.W.Wilkinson, "High Resolution Lithography on Active Semiconductor Membranes", Proc. Tenth International Conference on Electron and Ion Beam Science and Technology", Montreal (1980).

[3.9] A.Chandra, Solid State Electronics, 22 645 (1979).

CHAPTER 4 LITHOGRAPHY ON THIN SUBSTRATES

4.1 LITHOGRAPHY ON SILICON NITRIDE SUBSTRATES

The purpose of performing lithography on thin silicon nitride substrates was to assess the potential of electron beam lithography for making very small devices, which would eventually be made on GaAs membranes. This was necessary because of the time taken to develop a GaAs membrane fabrication procedure. Moreover, there exists a more or less permanent shortage of GaAs substrate material. Firstly, a lithographic procedure was devised (4.1.1), which gave lifted off metal patterns reliably and consistantly. Once this had been achieved, experiments were performed to find (a) the lower limits of linewidth and pitch resolution for metal lines fabricated by liftoff (4.1.2.1), (b) whether complex arbitrary patterns could be written (4.1.2.2), and (c) whether a method for very high resolution alignment could be devised (Chapter 5).

4.1.1 The Lithographic Procedure on Silicon Nitride Membranes

The lithographic procedure for silicon nitride membranes was developed from previous procedures used for exposures on carbon films [4.1,4.2]. The features which are most similar are the two layer resist system, development, and liftoff by 'shooting'.

4.1.1.1 Resist Deposition

Dies containing silicon nitride membranes were made ready for resist deposition as described in Section 3.2.2. The PMMA resist was deposited by the well-known technique of spinning. PMMA is normally received as a white powder which is prepared for spinning by dissolving it in an organic solvent (xylene,

chlorobenzene, or methyl iso-butyl ketone) to a concentration of 1-15% of the total weight of solution depending on the final thickness required. Solutions were stored in glass bottles with ground glass stoppers, which were thouroughly cleaned before use.

The coventional method to hold samples onto the spinner chuck is by vacuum, but this becomes impractical for small samples and particularly when the sample contains silicon nitride membranes, which have a tendancy to break under vacuum. The chucks supplied with the spinners for very small samples do not have an o-ring seal but have a single 1mm diameter hole at the centre of the chuck. The seal between the sample and the chuck is rather less than perfect, and during spinning, as the excess resist flows off the edge of the sample, some of it tends to get sucked under the sample into the vaccuum line which eventually becomes clogged. A more serious problem is that the wells underneath the membranes can become filled with resist, thus rendering them useless as far as STEM and thin substrate lithography is concerned. A further problem that arises even for solid substrate samples, is that the resist forms a thick lip round the edge of the sample where the surface tension is sufficiently high that not all the resist can leave the sample. This region of thicker resist can extend as far as 1mm in from each edge if a thick layer (more than 100nm) of resist is being spun. This can severely restrict the area on a sample that will be useful for lithography. A special chuck was designed and built that overcame these problems to a large degree. It was designed for the standard 5mm square silicon nitride membrane samples in particular, but it can can be used for any sample with a maximum edge length less than 6mm. The idea behind the design of the holder was that the resist should have an easy, continuous path to flow off the sample, and that there should be no capillary path under the sample to the back of membranes. The specimen is held in two clamps which are reverse bevelled where they touch the sample, so that the sample is held down firmly. The sample sits on a 2mm square pedestal which is in the centre of a 6mm square recess (Fig. 4.1). This is so that the underside

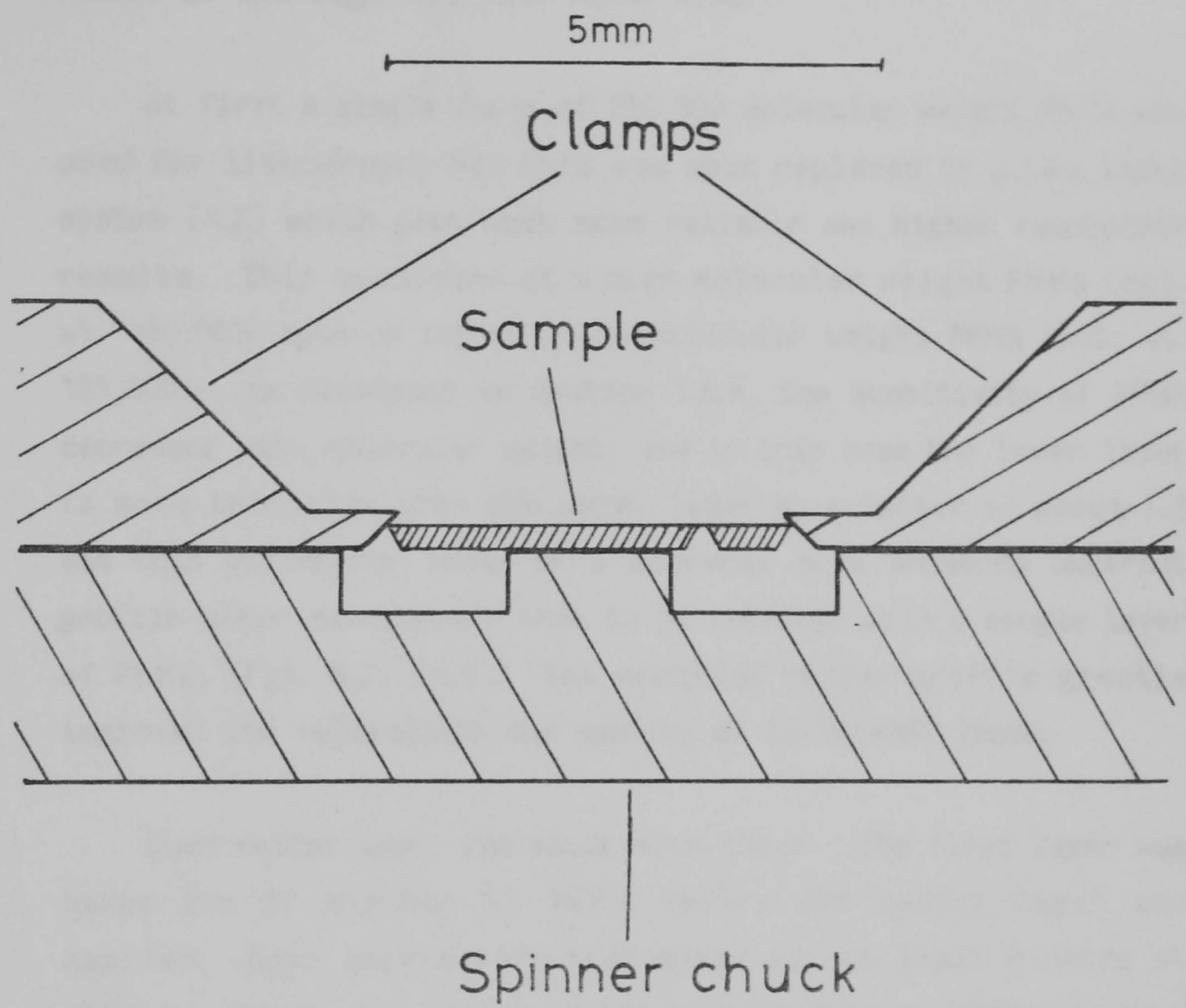


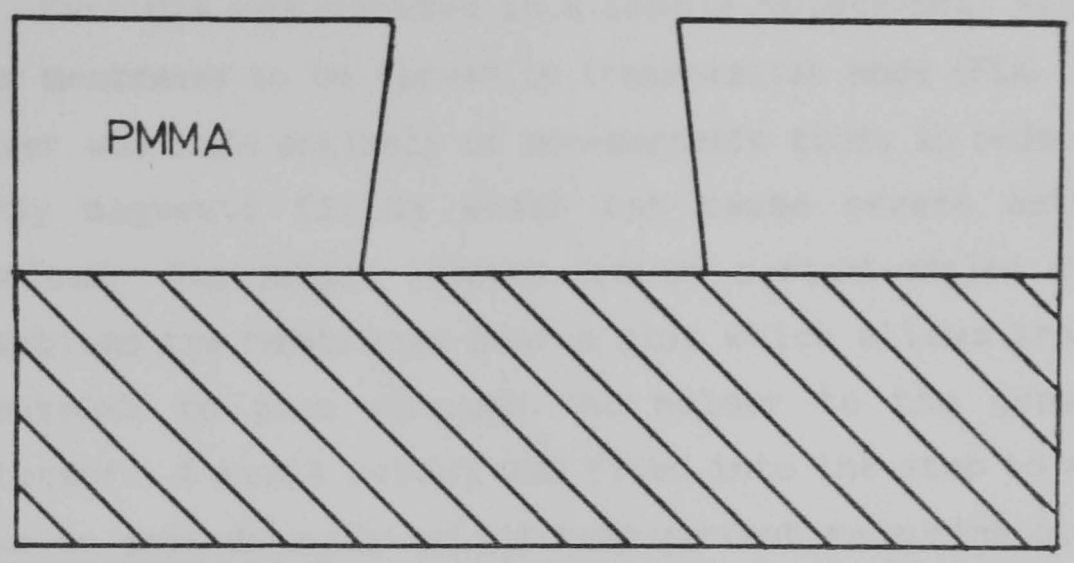
Fig. 4.1 Cross-section through special spinner chuck used for membrane samples. Resist has smooth path off sample, and there is no capillary path to underside of sample.

of the specimen does not touch the holder at its edges, and so resist cannot reach the back of the membranes by capillary action. When the special holder is used, the regions of thicker resist at the edges are only 0.2mm wide.

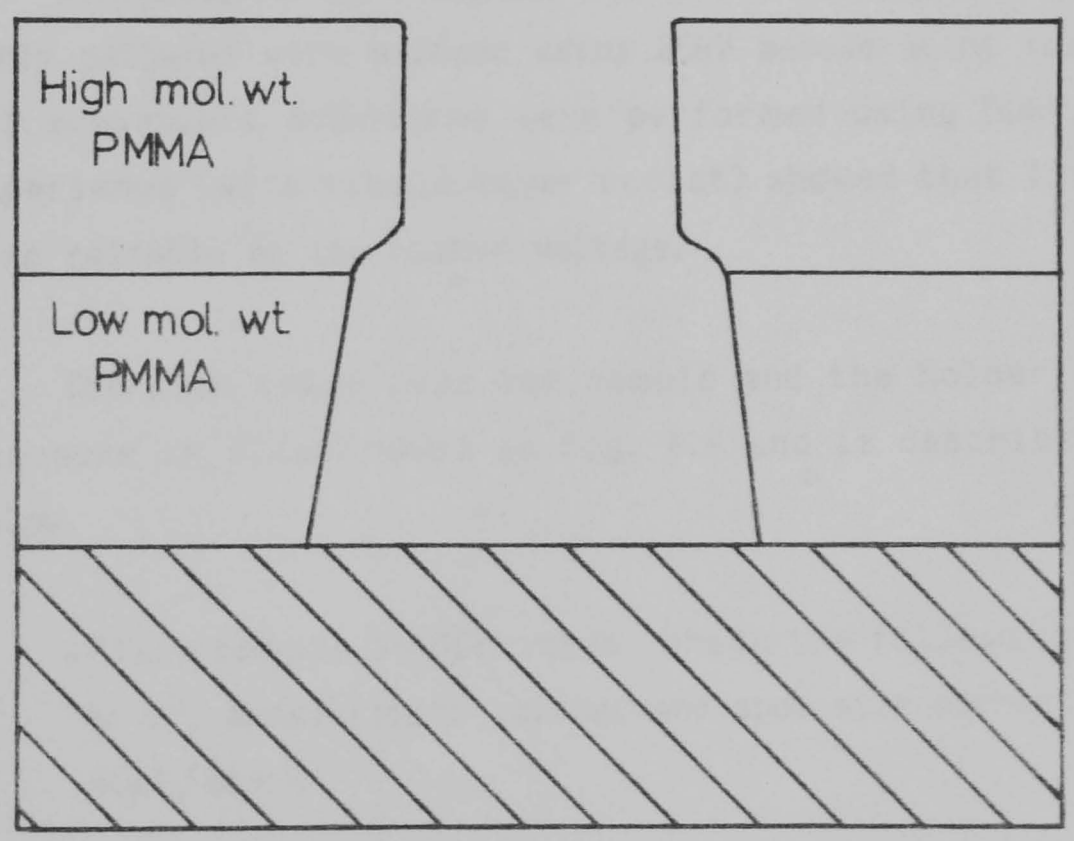
At first a single layer of 280,000 molecular weight PMMA was used for lithography but this was soon replaced by a two layer system [4.2] which gave much more reliable and higher resolution results. This consisted of a high molecular weight PMMA (mol. wt. 450,000) spun on top of a low molecular weight PMMA (mol. wt. 185,000). As discussed in Section 1.6.4, the sensitivity of PMMA decreases with molecular weight, and in this case the lower layer is more sensitive than the upper layer by a factor of about 1.3 and this difference leads to a somewhat more enhanced undercut profile after development than is be obtained with a single layer of PMMA, (Fig. 4.2) [4.3]. The modified resist profile greatly improves the reliability and quality of lifted-off lines.

Each resist layer was about 40nm thick. The first layer was baked for 30 minutes at 180°C before the second layer was applied. Both layers were then baked for at least 8 hours at 180°C to remove all traces of the casting solvent (MiBk was used at first, xylene was used later).

In order to facilitate accurate focussing and astigmatism correction during the exposure procedure, a thin layer of gold, nominally 5nm thick, was evaporated on the back face of the membranes. The gold films were then annealed at 180°C for 30 minutes. The annealing step caused the gold to coalesce into well defined grains with diameters from 5-20nm, an ideal size for accurate focussing and astigmatism correction, essential if the highest resolution results are to be obtained.



a) Single layer resist



b) Double layer resist

Fig. 4.2 Developed profiles of;
a) single layer resist, and
b) double layer resist.

The more sensitive lower level of the double layer resist develops to a greater distance from the centre of the exposed line, and thus a good undercut profile for liftoff is formed.

4.1.1.2 Exposure Procedure

Each die was mounted in a sample holder that allowed all four membranes to be viewed in transmission mode (Fig. 4.3). The holder was made entirely of non-magnetic brass in order to avoid stray magnetic fields which can cause severe astigmatism problems. The sample locates against a right-angled step which positions the membranes over a slot which allows transmitted electrons to pass through the holder to the transmission detector. A small cutout was filed into the step to allow the beam to pass directly to the beam current measuring plate which could be swung into position just above the transmission detector (Section 2.4.2).

All patterns were exposed using a nominally 8nm spot; a few early patterns were exposed using 25kV accelerating voltage, but all subsequent exposures were performed using 50kV, because experience (with single layer resist) showed that liftoff was more reliable at the higher voltage.

The path taken over the sample and the holder during an exposure is illustrated in Fig. 4.4 and is described briefly below.

-Place sample in SEM, then check the following; tilt set to 0^{o}, accelerating voltage and spot size correct (normally 50kV, 8nm).

-Move to beam current measuring position on holder (left 2.7mm, down 0.7mm from starting position of goniometer when specimen is inserted). Measure beam current. Use SENDPAT to transfer patterns from GEC 4070 computer to RAM of KIM microprocessor.

-Move to middle of left edge of Square 1. Focus and align edge with vertical of screen.

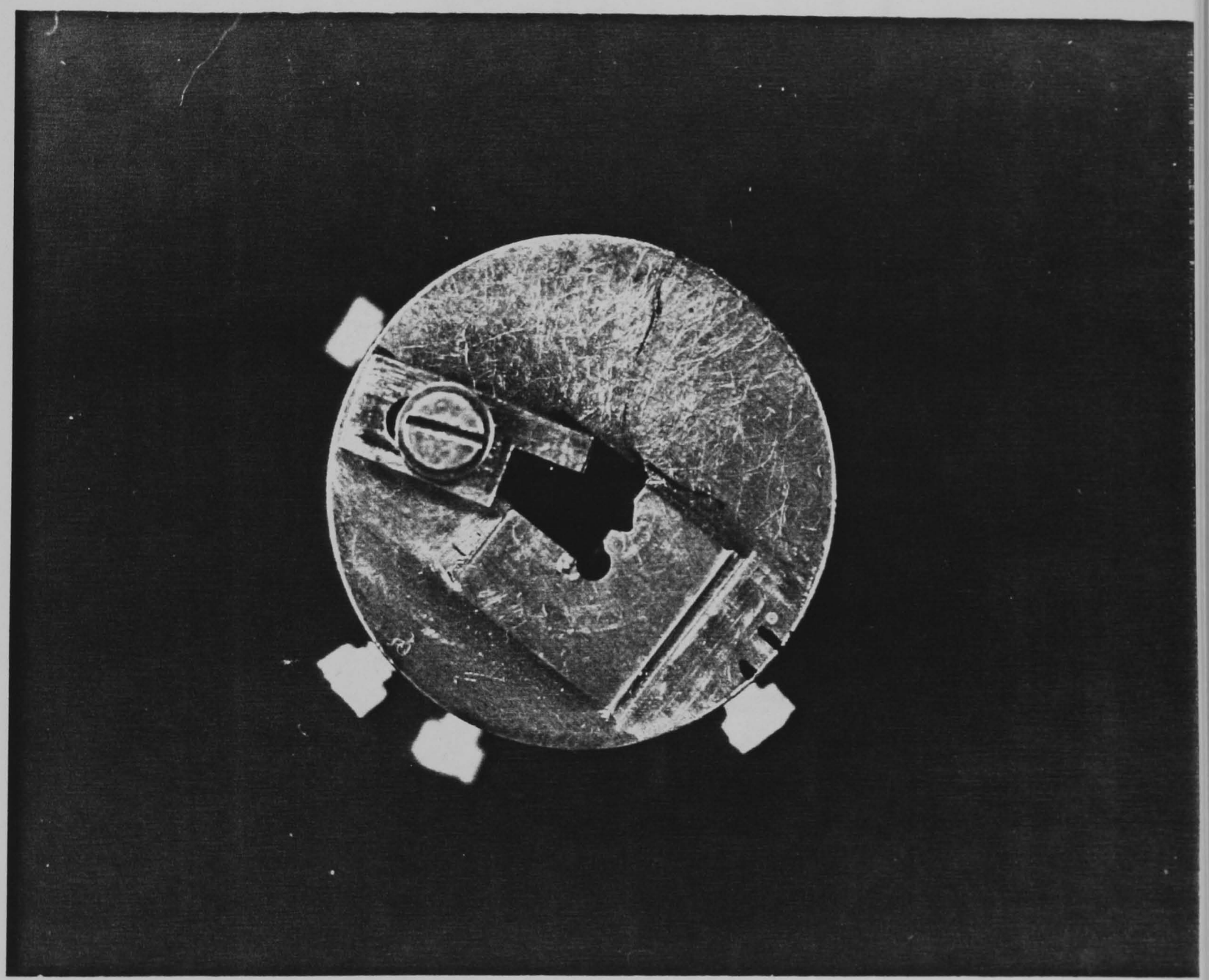


Fig. 4.3 Holder used during exposure of membrane samples. Sample is clamped against step so that membranes lie over slot and transmitted electrons can reach transmission detector. Filed cutout in step allows beam to reach current measurement plate without passing through sample.

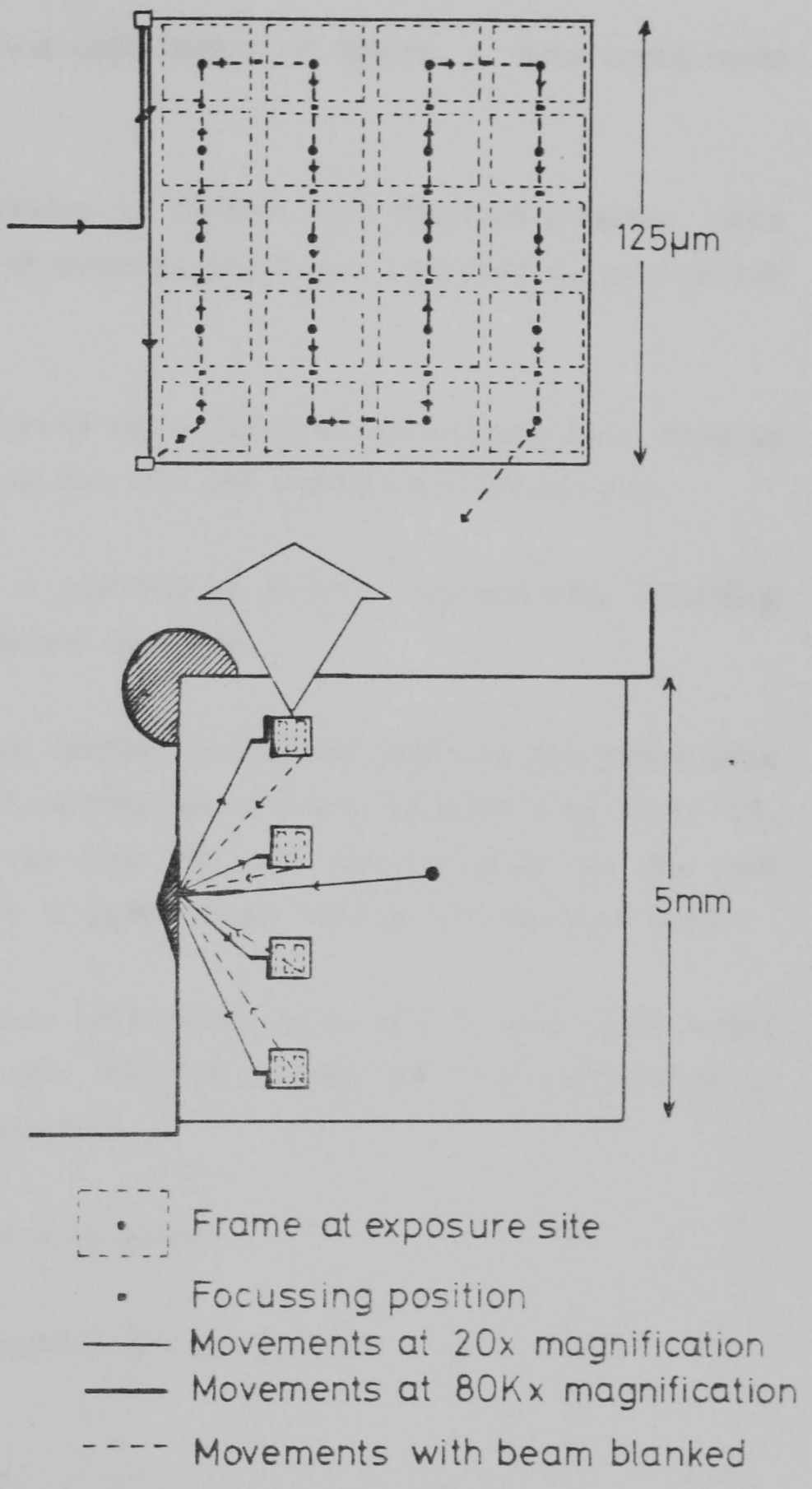


Fig. 4.4 Path of centre of frame over sample during exposure. Beam current is measured at cut-out before exposuring each membrane, and is checked at the end.

-Move to Top Left corner of Square 1. Focus and correct astigmatism. Note coordinates of corner on goniometer.

-Move to Bottom Left corner of Square 1. Note coordinates of corner.

-Enter coordinates of corners into POSITION program. Note coordinates of exposing positions relative to goniometer axes.

-Focus 10 microns below first exposing position. Move to first exposing position and expose required pattern.

-Expose rest of patterns on Square 1 as required, focussing below (or beside) each one.

-Move to beam current measuring position and check beam current. If current has changed by more than (say) 10%, reload patterns from GEC 4070 computer using the new beam current value to give correct timings for exposure data.

-Move to bottom left corner of Square 2. Enter goniometer coordinates into POSITION program and note coordinates of exposing positions.

-Expose Square 2 as Square 1

-Repeat for Squares 3,4 if required.

4.1.1.3 Development

The exposed patterns were developed by dipping in 32ml of a 3:1 mixture of iso-propyl alchohol (IPA) and methyl iso-butyl ketone (MIBK) for 30 seconds at a temperature of 23.0°C. The samples were then either immediately blown dry, or rinsed in IPA

and then blown dry. It has been shown that development continues for about 7 seconds after the start of rinsing [4.4], but no difference in line quality was detected between the two methods.

The developer used here (3:1) was used largely for historical reasons. It gives good contrast, but at the same time is not so weak that it leaves straggling resist in the bottom of developed regions. Other work in this Department [4.3] has shown that 2:1 could be an even better choice, especially if short development times were to be used. Since the main purpose of this project was to fabricate a device, however, it was decided not to spend time attempting to optimise development absolutely. The 3:1 mixture, which, even if it was not optimal, was at least capable of developing features of the required resolution.

Each sample was examined in an optical microscope after development to check for proper positioning and exposure of patterns.

4.1.1.4 Evaporation

Metallisation for liftoff was peformed in a vacuum chamber pumped to 10^{-5} Torr with a liquid nitrogen trapped diffusion pump. Gold/palladium (60:40) alloy was used for all the work on thin substrates. It has a grain size of about 6nm when thermally evaporated from a tungsten filament onto a substrate at room temperature.

The evaporation arrangement is shown in Figure 4.5. Samples were placed in a special holder 150mm directly above the source and perpendicular to the direction of the beam. It was found that deviations from the perpendicular as small as 5^{o} could cause problems with sidewall coverage and lead to poor liftoff.

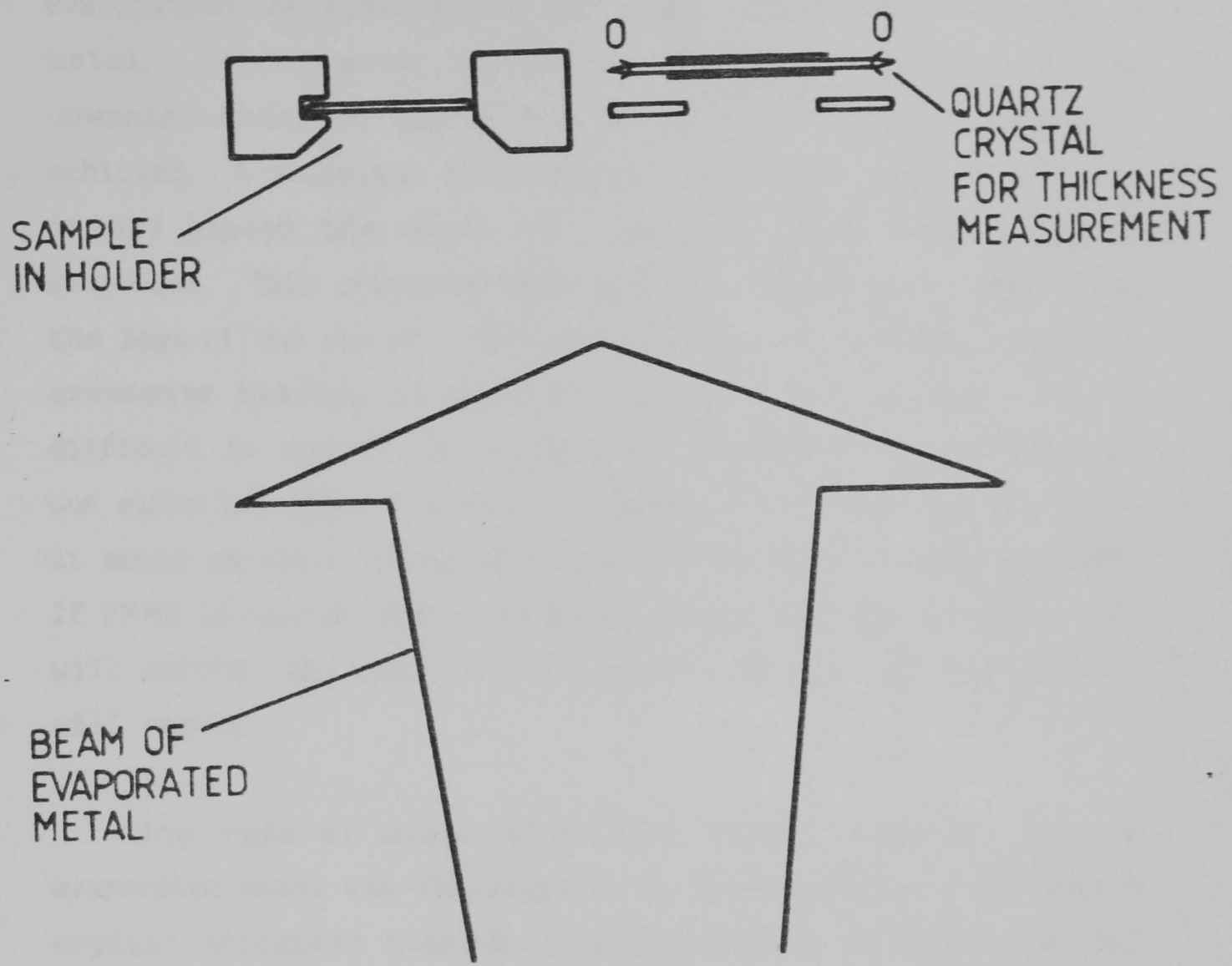


Fig. 4.5 The evaporation set-up.

The source size is important if clean liftoff is to be achieved. The source must be kept small so that during the evaporation the sidewalls of the resist do not become coated in metal. If the metal on the substrate is connected to the unwanted metal on top of the resist, good liftoff cannot be achieved. A stainless steel shield with a 5mm diameter hole in it was placed 3mm above the tungsten basket containing the evaporant. This prevented evaporant arriving at the sample from the legs of the basket. The shield protects the sample against excessive heating as well as collimating the beam. It is difficult to quantify how important protection against heat is, but since the glass transition temperature of PMMA is only 120^{o}C it seems sensible to minimise any heating effects where possible. If PMMA is heated above its glass transition temperature, it will deform, the undercut profile will be lost and poor liftoff will result.

The rate of evaporation and total thickness of the evaporated metal was displayed on an Intellemetrics IL002 quartz crystal thickness monitor. Evaporated metal is allowed to fall onto the surface of a quartz crystal. The monitor measures and displays the change in resonant frequency of the crystal due to the extra mass of the deposited metal on it. The metal thickness can then be found from a calibration graph plotted from previous experimental data. The Au/Pd alloy was deposited at a rate of about 0.3nm/sec to a total thickness of 8-12nm. If the metal is evaporated too quickly, the resulting stress can cause it to peel off.

4.1.1.5 Liftoff

The excess metal on top of the resist was removed by dissolving the resist in a strong solvent and providing sufficient agitation that the waste metal on top of the resist was washed away and did not stick to the surface of the sample. An ultrasonic bath is often used to provide the agitation when

solid substrate exposures are being lifted off, but silicon nitride membranes cannot survive this treatment. Instead the procedure known as 'shooting' was employed (Fig. 4.6) [4.1]. The sample is immersed in a beaker of methanol (a non-solvent for PMMA) and chlorobenzene or methyl ethyl ketone (MEK) is fired through a hypodermic syringe from below the surface of the methanol onto the front face of the sample. The diameter of the jet of solvent is only about 0.75mm so the sample has to be moved around in the jet so that each membrane receives an equal dose of solvent. The method provides enough agitation to remove the excess metal cleanly and prevent it sticking, while not damaging the membranes. Membranes rarely broke during this process, and liftoff was nearly always successful. After shooting, samples were rinsed in clean methanol and blown dry.

4.1.1.6 Examination

Every sample was examined in an optical microscope to check that liftoff had been successful. Most samples were then examined in STEM mode in the Philips PSEM 500. If it was desired to photograph the patterns or to take measurements of linewidths, the gold focussing dots had first to be removed by sputtering in argon.

If high quality pictures or very accurate measurements were required then the specimen was examined in a JEOL TEM located in the Department of Natural Philosophy in this University. All measurements of linewidths and centre-to-centre spacings quoted in Section 4.1.2.2 were taken from TEM micrographs.

4.1.2 Experimental Results on Silicon Nitride Membranes

The first aim was to practice and improve the whole lithographic process as it then stood in order to get reliable results. The first test pattern consisted of 12 groups of 4

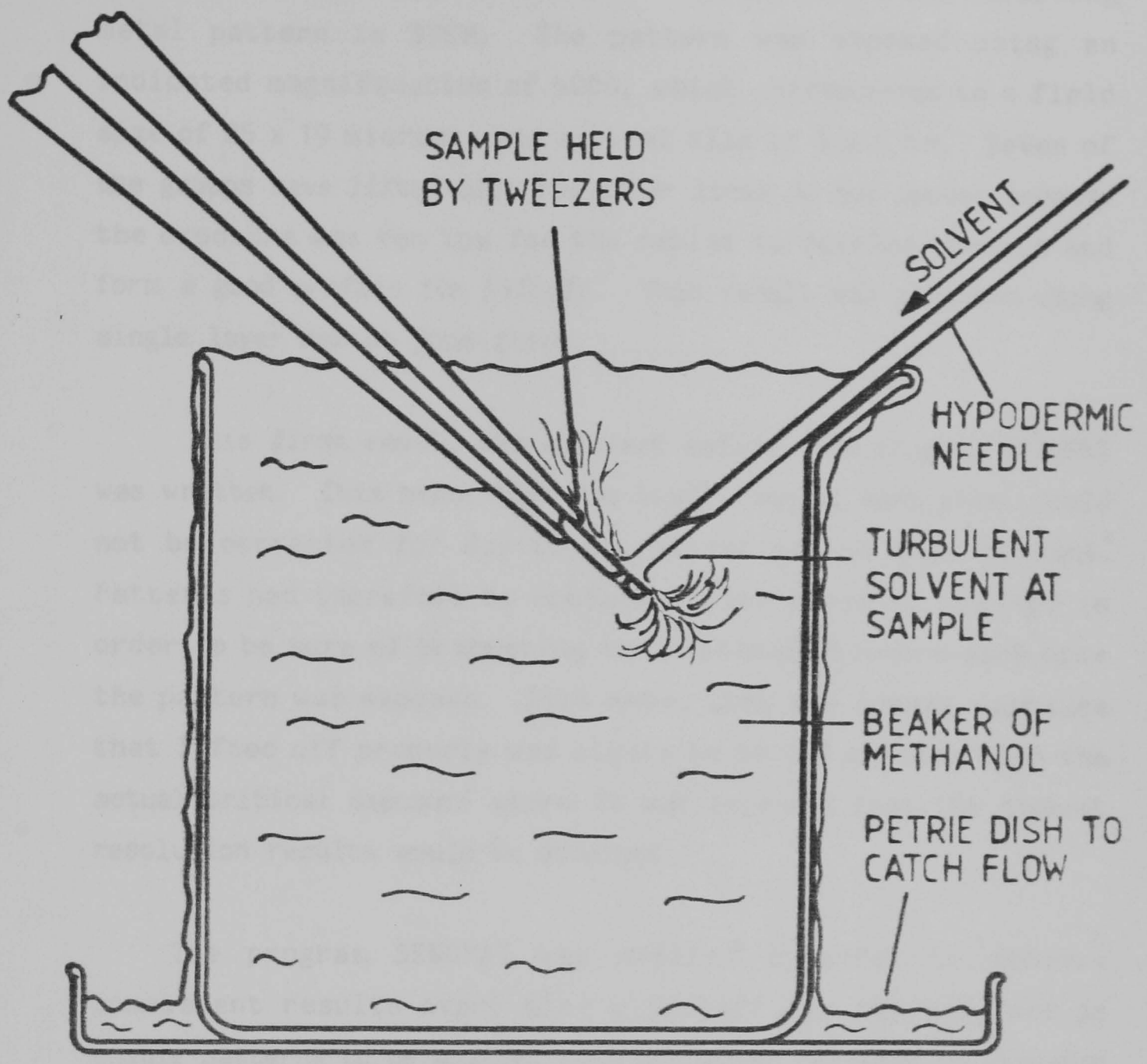


Fig. 4.6 Liftoff by 'shooting'. The methanol in the beaker provides support to the membranes when the solvent is fired through the hypodermic needle to dissolve the remaining resist and wash away the unwanted metal on top of the resist.

single pass lines, and was designed to find the critical exposure for liftoff. The exposure given to each group of lines increased by about 20% from group to group. This was the first pattern that was successfully lifted off. Fig. 4.7 shows the resulting metal pattern in STEM. The pattern was exposed using an indicated magnification of 5000, which corresponds to a field size of 25 x 19 microns, and a pixel size of 6 x 5 nm. Seven of the groups have lifted off; the other lines do not appear because the exposure was too low for the resist to develop through and form a good profile for liftoff. This result was obtained using single layer resist 50nm thick.

This first result was obtained before the program SENDPAT was written. This meant that the dwell time at each pixel could not be corrected for day to day variations of beam current. Patterns had therefore to contain a wide range of timings in order to be sure of bracketing the critical exposure each time the pattern was exposed. This meant that the lowest exposure that lifted off properly was likely to be 10% greater than the actual critical exposure where it was expected that the highest resolution results would be obtained.

The program SENDPAT was written in order to achieve consistent results every time a pattern was exposed, and so enable patterns to be written with a smaller increase in exposure from group to group. This way, the critical dose could be found more accurately, and hence the smallest possible linewidth and smallest possible centre-to centre spacing could be found. A pattern designed to find the critical exposure for single pass lines is shown in Figure 4.8. It consists of 64 groups of three lines, the exposure given to each group rising in a geometric series from about 3000 microcoulombs cm^{-2} to 10000 microcoulombs cm^{-2}. The pattern was designed to be exposed at a magnification of 10000x, in order to smooth out 'lumpiness' in the exposure along the lines due to the finite size of the pixels. The pattern was created using DESIGN (Section 2.2.1), and it required one G command followed by an M command to create the whole

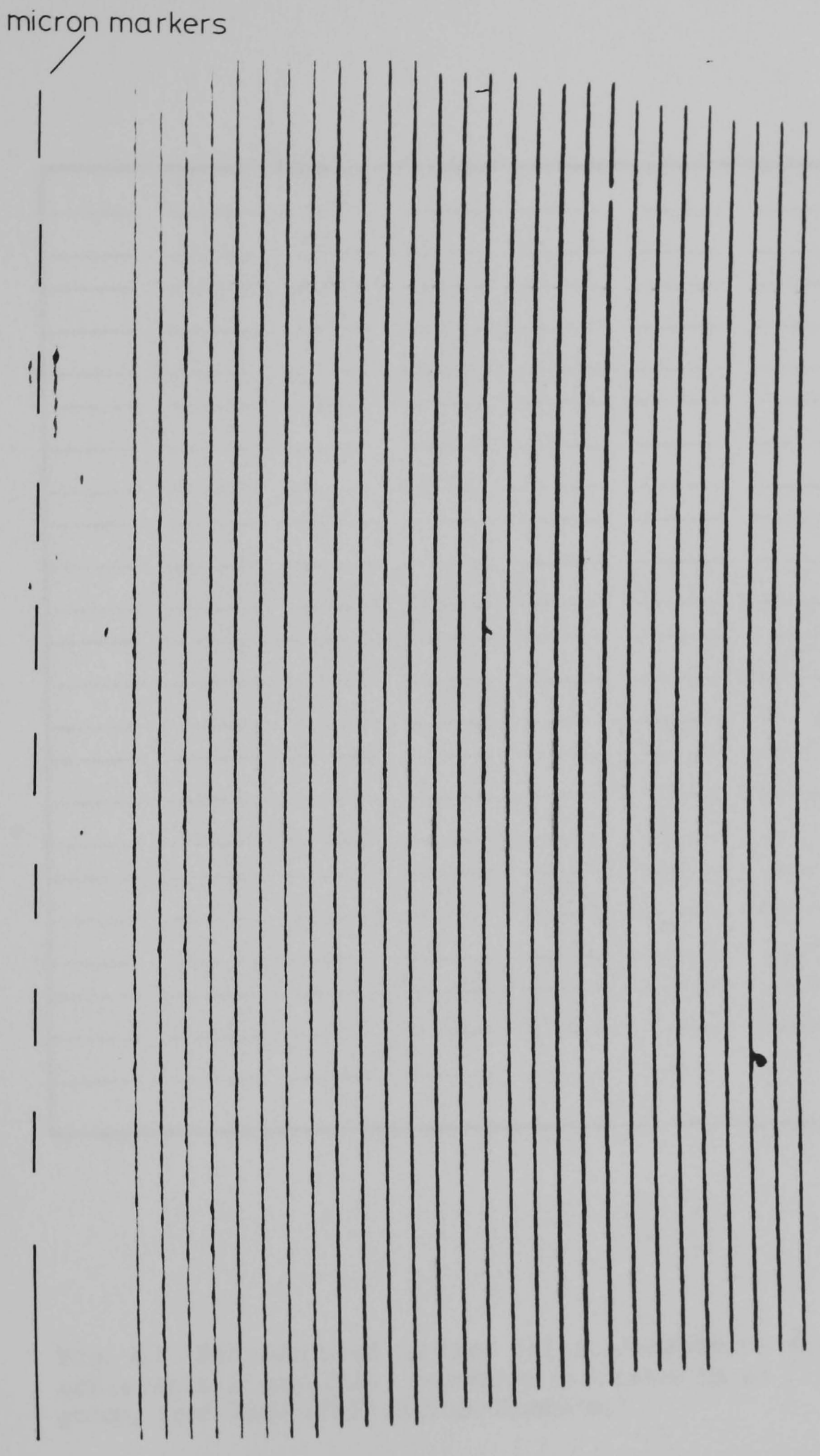


Fig. 4.7 The first successfully lifted off metal pattern. The exposure decreases by 20% from each group of four lines to the next.

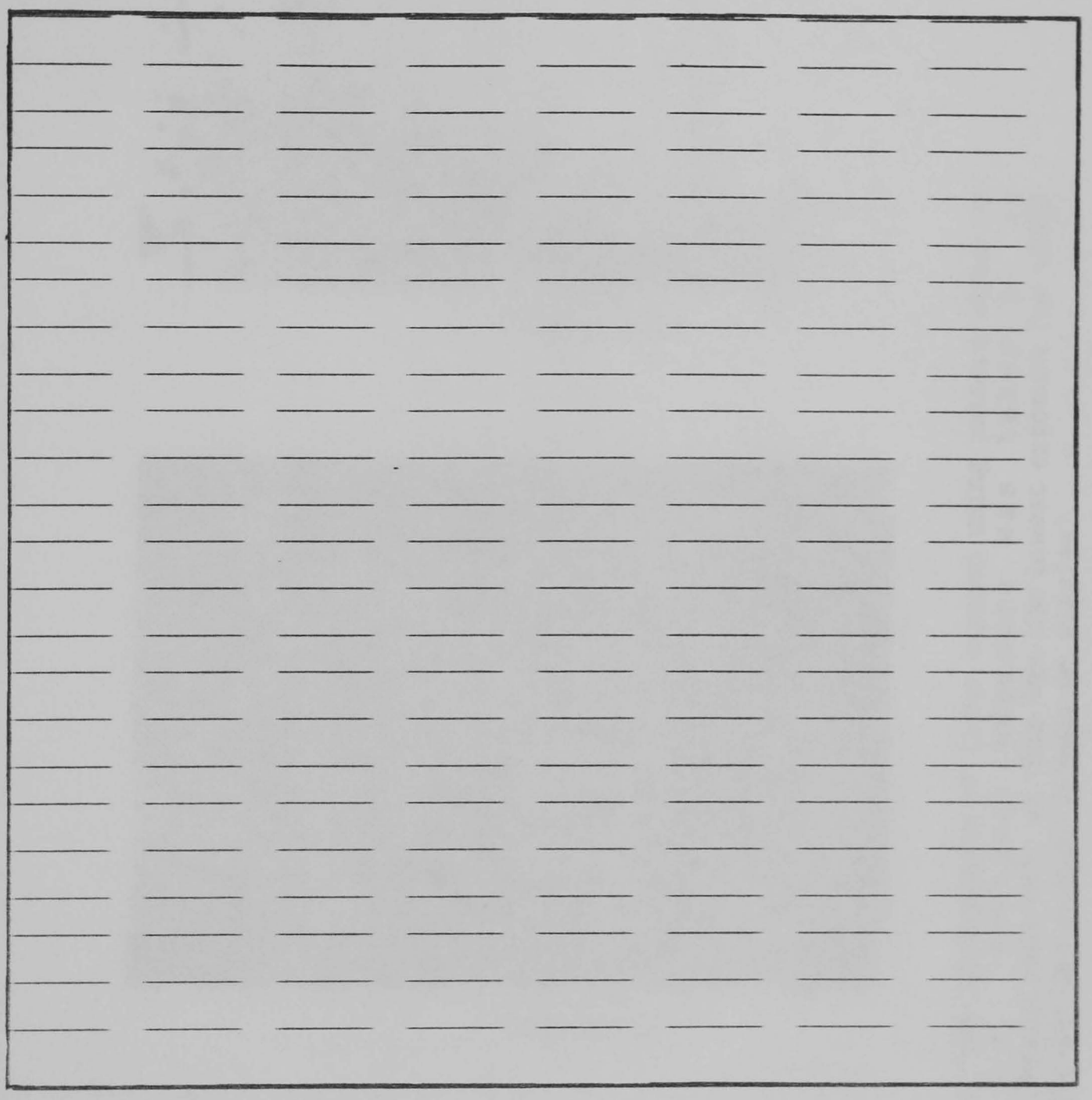
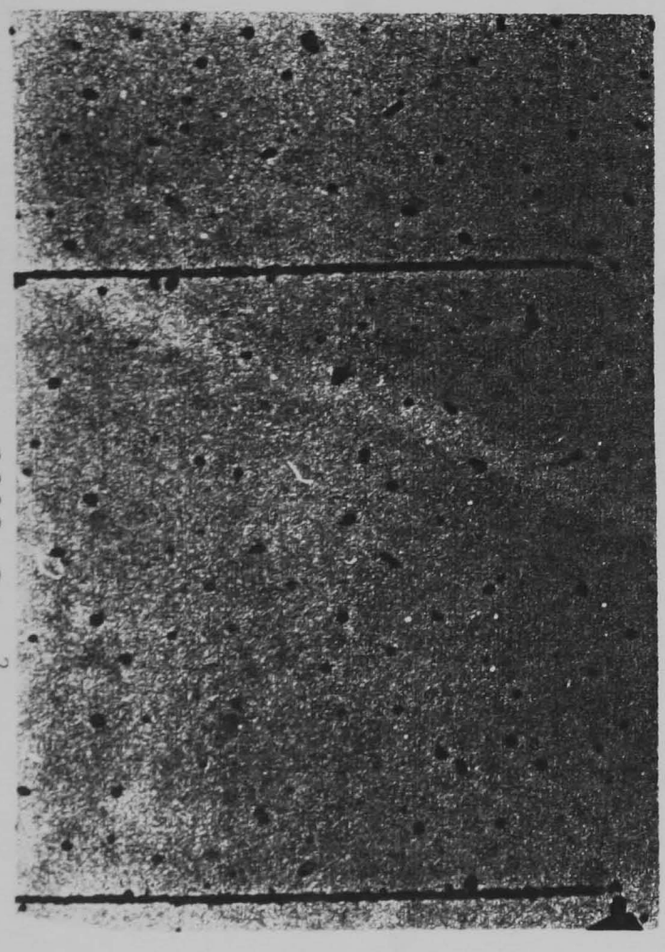
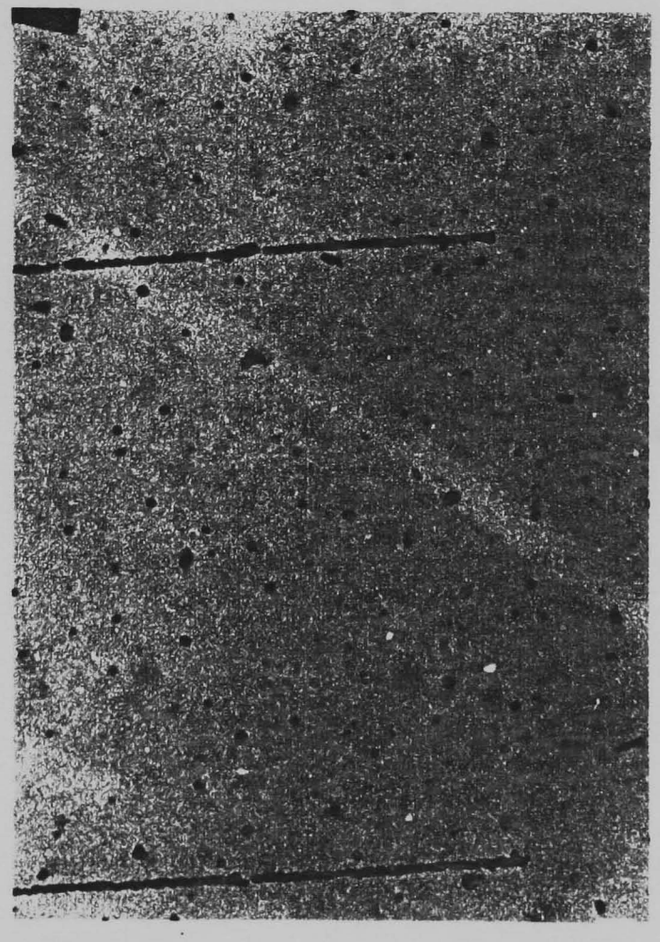


Fig. 4.8 Pattern used to find critical exposure and narrowest achievable linewidth. Exposure increased by 2% from group to group, from 3000-10000 microcoulombs/cm^2.





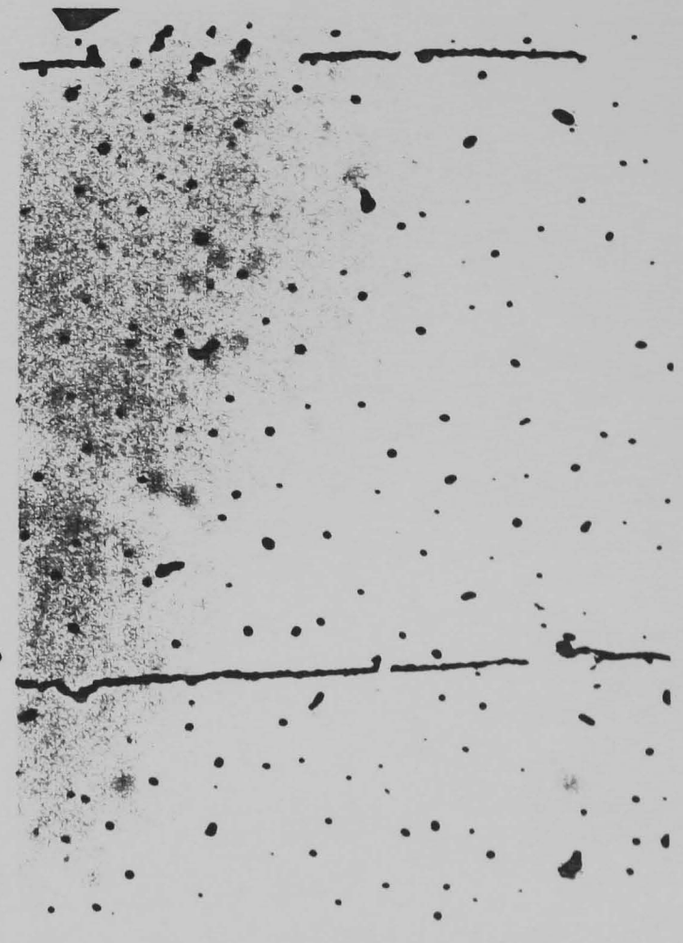


Fig. 4.9 TEM micrographs of lines exposed using pattern shown in Fig. 4.8. Critical exposure was taken to be 6330 microcoulombs/cm^2, as this was the lowest exposure for which the lines did not contain breaks or wiggles.

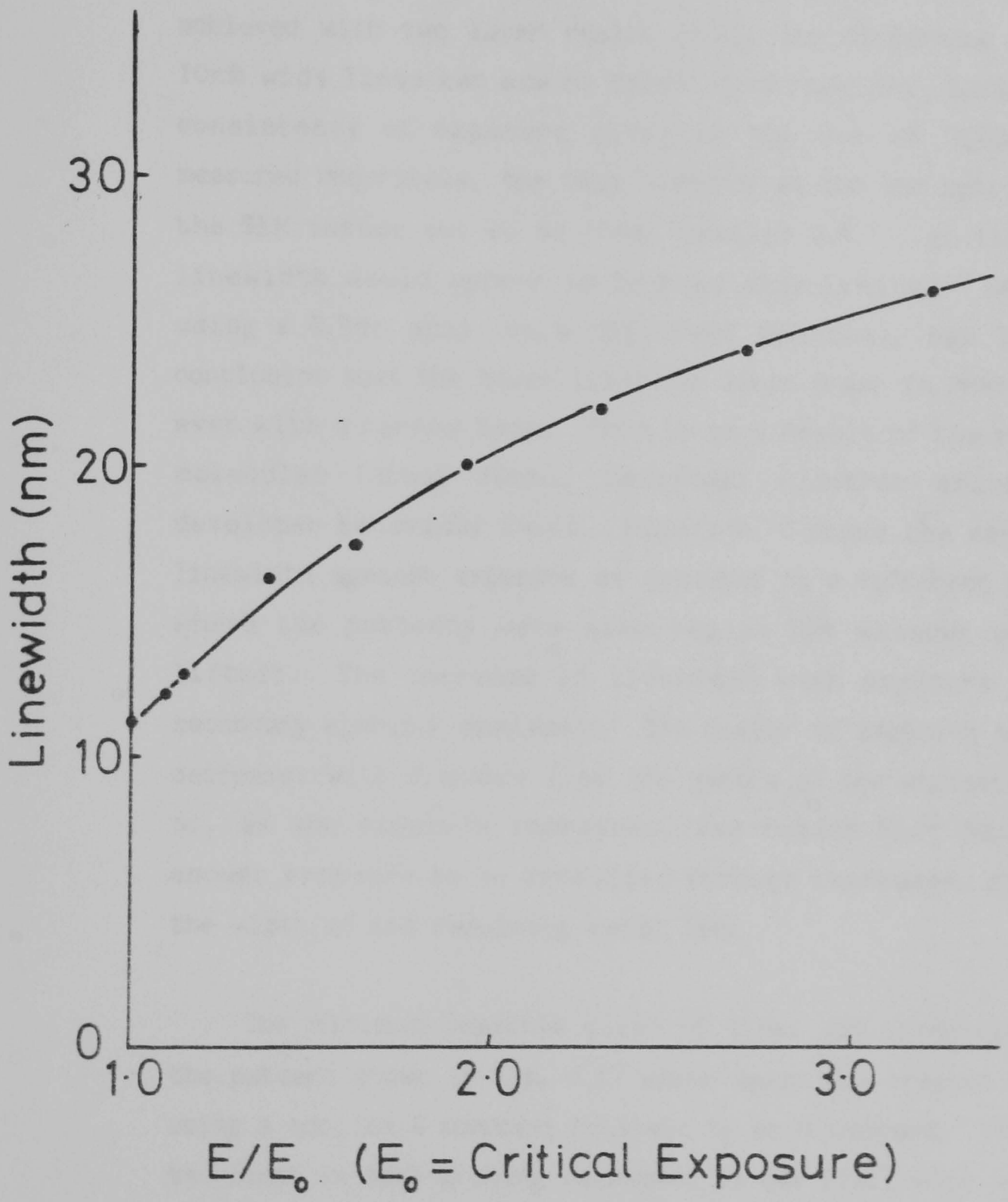


Fig. 4.10 Graph of linewidth vs. exposure for isolated lines examined in TEM without lifting off (After Lee, unpublished).

pattern. Fig. 4.9 shows lines exposed at and just below what was taken to be the critical exposure for liftoff. The critical exposure was taken to be the minimum which gave continuous lines, but as can be seen from Figure 4.9, the decision of what constitutes a continuous line is somewhat arbitrary. The minimum linewidth obtained was 10nm which is the same as previously achieved with two layer resist [4.2], the difference being that 10nm wide lines can now be fabricated routinely because of the consistency of exposure given by the use of SENDPAT. When measured accurately, the beam diameter at the 8nm spot setting on the SEM turned out to be 10nm (Section 2.4.1), so the minimum linewidth would appear to be beam size limited. Other work, using a 0.5nm spot (on a different machine), has led to the conclusion that the lower limit for lines drawn in PMMA is 9-10nm even with a narrow beam. This is as a result of the size of the molecules (about 10nm), secondary electron emission, and developer behaviour [4.3]. Figure 4.10 shows the variation of linewidth against exposure as measured in a different experiment where the patterns were examined in TEM without performing liftoff. The increase in linewidth with exposure is due to secondary electron emmission. The number of exposing secondaries decreases with distance from the centre of the exposed line, and so, as the exposure increases, the region that has received enough exposure to be developed through increases, and with it the width of the resulting metal line.

The minimum possible pitch of lines was investigated using the pattern shown in Fig. 4.11 which again was created on DESIGN, using a complex G command followed by an M command. The pitch of the lines in each grating decreases by one pixel with a period of three lines. The pattern was also exposed at 10000x magnification, so the pixel size was 3 x 2.5 nm. The exposure given to each group of lines was again increased in a geometric series, this time in such a way that the lowest exposure was just above the critical exposure. Fig. 4.12 shows TEM micrographs of some of the groups within one of these patterns. Fig. 4.13 is a graph of minimum pitch against exposure. The minimum pitch

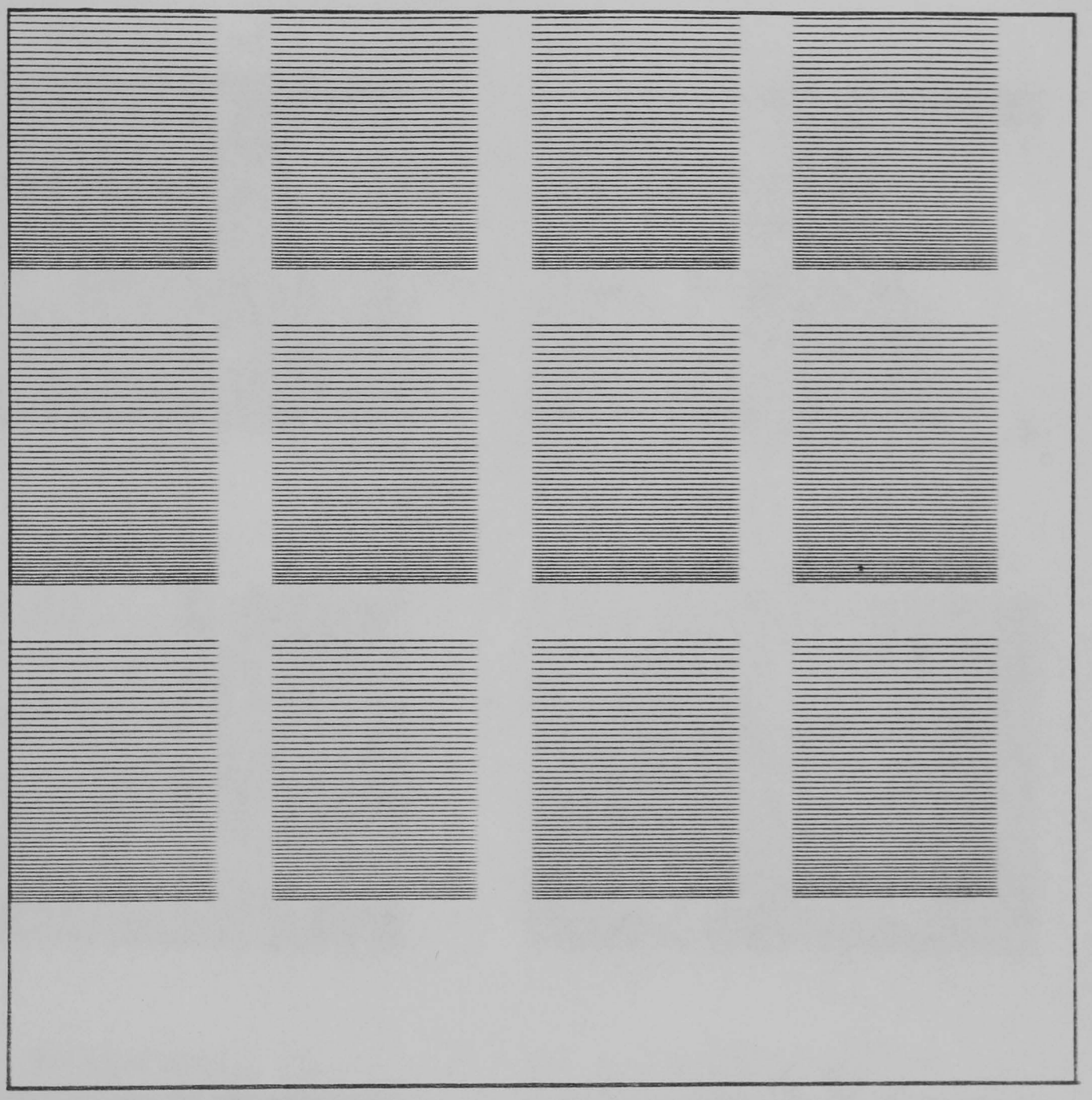


Fig. 4.11 Pattern to find minimum achievable pitch resolution of lifted off lines. Pitch of lines decreases by one pixel every third line down each group, and exposure increases by about 5% from group to group.

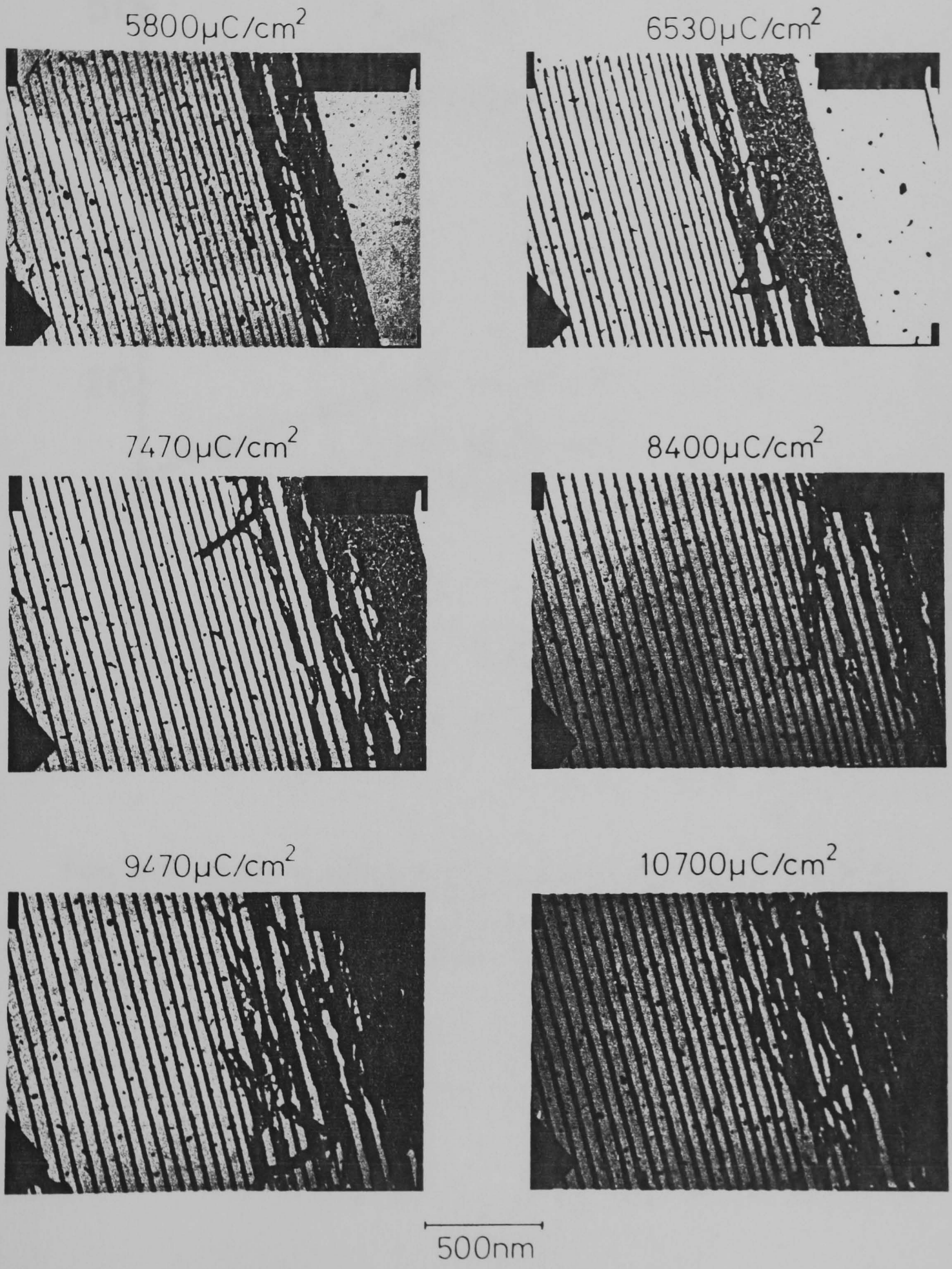


Fig. 4.12 TEM micrographs showing sections of pattern shown in Fig. 4.11. Narrowest pitch obtained was 40nm.

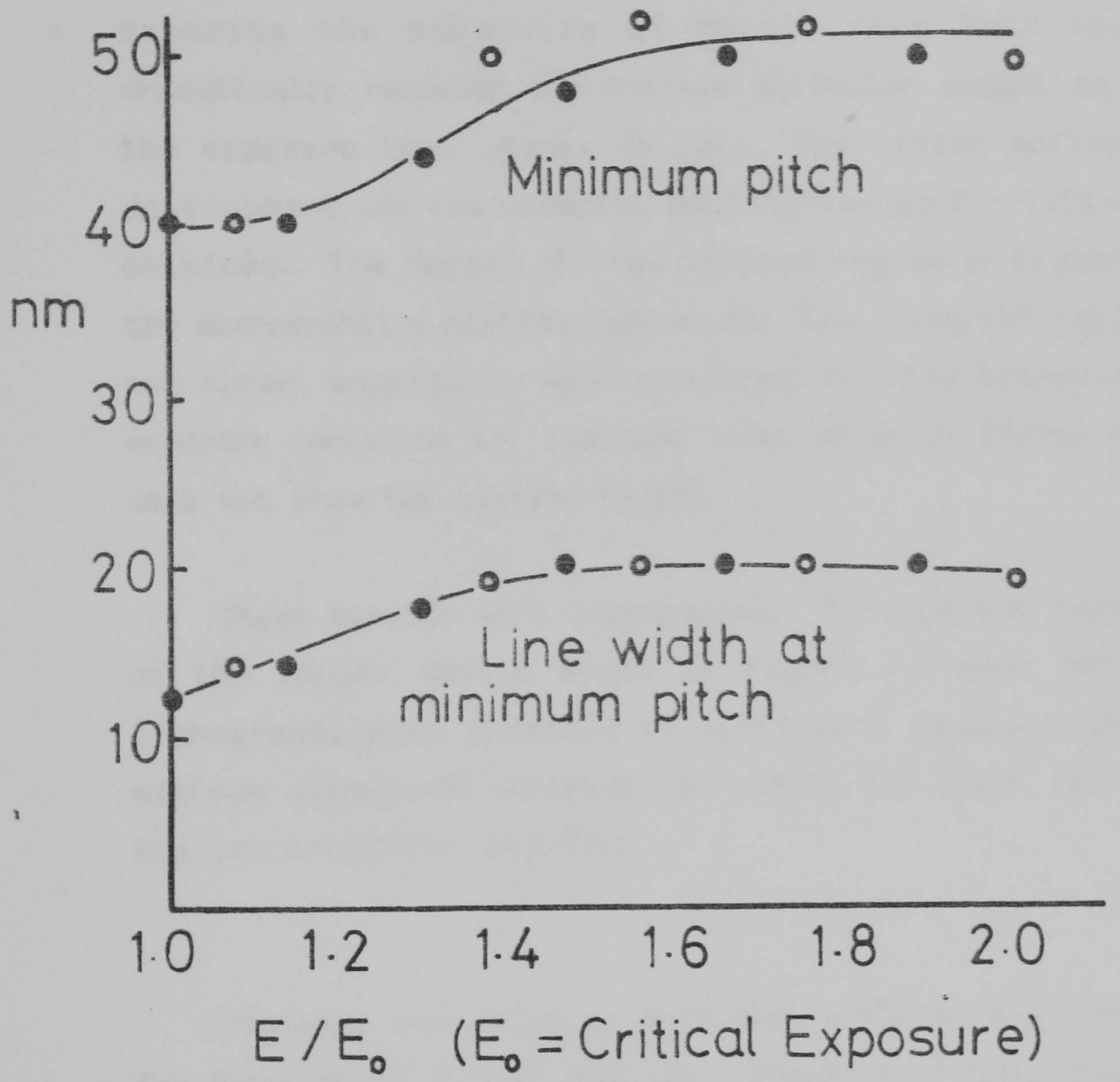


Fig. 4.13 Graph of minimum pitch of lifted off metal lines and linewidth at the minimum pitch vs. exposure. The gap between lines at the minimum pitch remains close to 30nm for all exposures, since the linewidth and minimum pitch increase in tandem.

achieved was 40nm. The interesting feature of the graph is that the minimum pitch rises from 40nm at the critical exposure to 50nm at 1.4 times the critical exposure, but does not increase further at greater exposures. Liftoff failure for closely spaced lines is probably due to the secondaries from neighbouring lines exposing the molecules of resist from both sides thus dramatically reducing the average molecular weight so that, at the exposure that causes failure, the resist softens during development and the undercut profile needed for liftoff is not obtained. The reason for the plateau region in Figure 4.13 and the corresponding plateau region for the linewith variation is not clear, especially when compared with the linewidth versus exposure variation for isolated lines shown in Figure 4.10 which does not show the plateau region.

These results were encouraging. The minimum feature sizes of the target device shown in Figure 1.6 were shown to be lithographically possible if sufficient care was taken. The minimum dimensions achieved for lifted off lines were state of the art or better [4.2,4.5].

Patterns consisting of more than gratings would be required for fabrication of real devices. Figure 4.14 shows two patterns which were made in order to demonstrate that such patterns could be made. The upper picture is of a pattern designed to measure the resistance of the central wire by the four point probe method in order to study the phenomenon of 'localisation' [4.6-4.8].

The lower picture is of a pattern that shows how curved lines become discontinuous close to the critical dose because the curves are in fact made up of discrete rectangles.

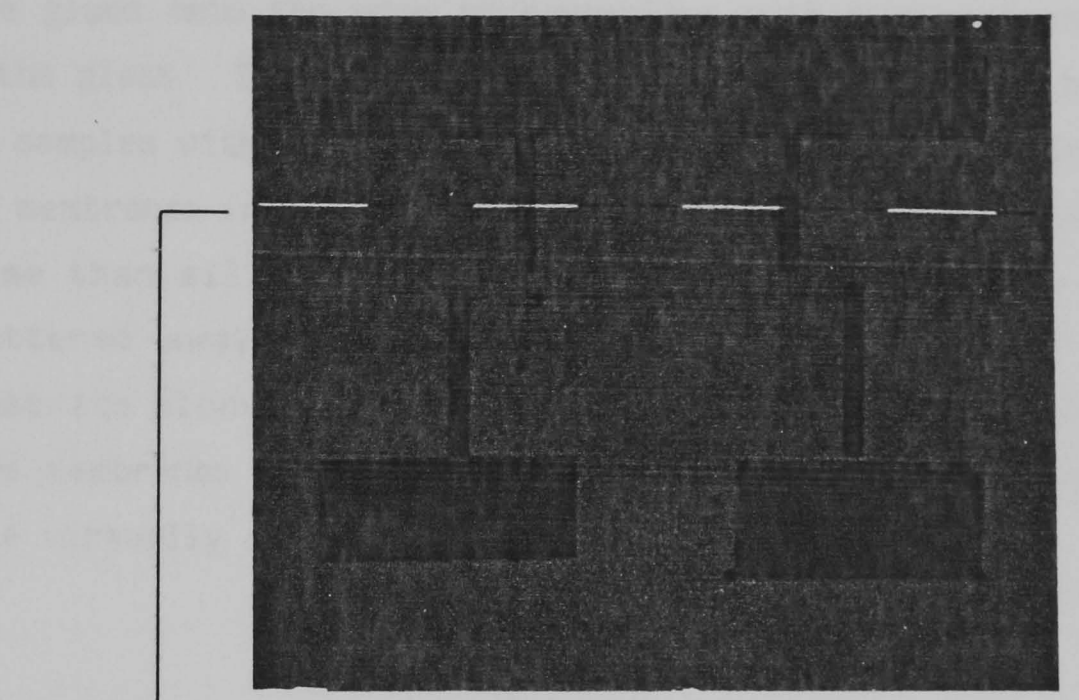


Fig. 4.14 (Upper) Four point probe pattern for localisation studies.

(Lower) Pattern showing break up of curves made from rectangles when close to the critical exposure.

4.2 LITHOGRAPHY ON GAAS MEMBRANES

The lithographic process on GaAs membranes was similar to that on silicon nitride membranes, except that the GaAs samples were glued onto the edge of coverslips with the membranes clear of the glass. This arrangement meant that it was easy to handle the samples without breaking them, while still being able to view the membranes in STEM for lithography. GaAs is considerably more dense than silicon nitride, and much of the primary beam is scattered away from the the transmission detector by the substrate alone. It was found that, on the machine used here, 50nm membranes were just usable in STEM, while 100nm membranes were virtually opaque.

4.2.1 Results

The standard two layer resist (Section 4.1.1.1) was spun onto a few membranes. The pattern used to find the minimum pitch resolution on silicon nitride membranes was used at 5000x magnification instead of 10000x so that it was less difficult to fabricate. Fig. 4.14 shows part of the pattern after liftoff of about 10nm of Au/Pd. The linewidth was about 15nm and the minimum pitch was 70nm, which although not as high resolution as was obtained on silicon nitride, was as good as could be expected from the pattern that was used.

After this first demonstration of lithography on GaAs membranes, it was decided not to make any more membranes for the time being even when more material was available; lithography had been achieved on a GaAs membrane with feature sizes comparable to that achieved on silicon nitride membranes. The absolute limits of lithography on GaAs membranes were not important, for the time being at least. It was more important to move on and try some of the other procedures that would be required for a membrane transistor. As it turned out, the other parts of this project took so much time that no more membranes were made, and no more

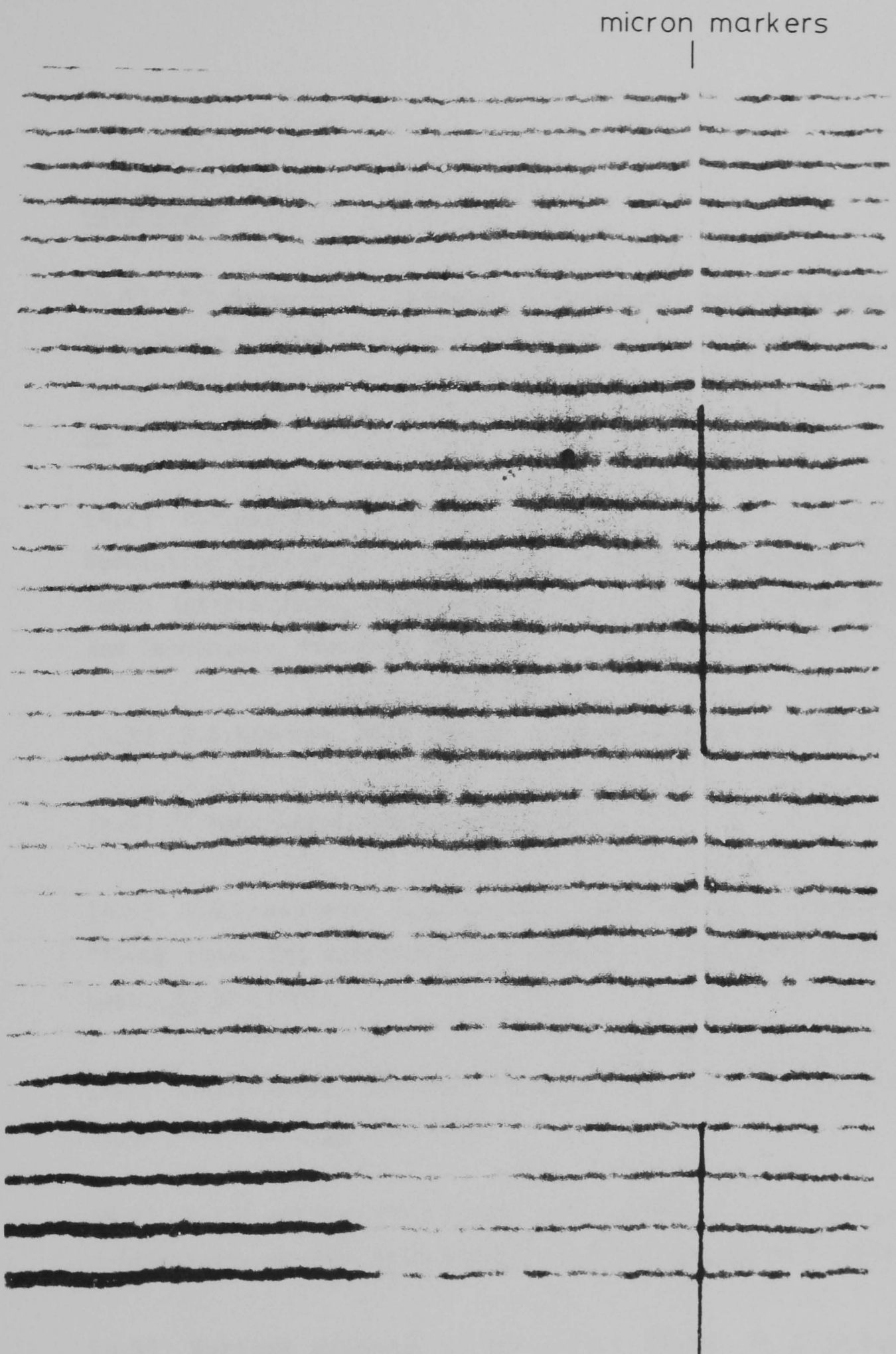


Fig. 4.15 STEM micrograph of metal lines lifted off on GaAs membrane. The line width is about 15nm and the minimum pitch in the pattern was 70nm.

lithography carried out on thin GaAs.

References

[4.1] 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(methyl methacrylate) films and liftoff", Appl. Phys. Lett. 38 436 (1981).

[4.2] S.P.Beaumont, B.Singh, and C.D.W.Wilkinson, "Very High Resolution Lithography - Thin Films or Solid Substrates", Proc. Tenth International Conference on Electron and Ion Beam Science and Technology, Montreal 1982.

[4.3] S.A.Rishton, Ph.D thesis, Glasgow University (1984).

[4.4] C.Macgreggor, unpublished (1981).

[4.5] H.G.Craighead, R.E.Howard, L.D.Jackel, and P.M.Mankiewich, "10-nm linewidth electron beam lithography on GaAs", Appl. Phys. Lett. 38 38 (1983).

[4.6] D.J.Thouless, "Maximum Metallic Resistance in Thin Wires", Phys. Rev. Lett. 39 1167 (1977).

[4.7] D.J.Thouless, "The Effect of Inelastic Scattering on the Conductivity of Very Thin Wires", Sol. St. Com. 34 683 (1980).

[4.8] Various authors, J. Non-Cryst. Solid. 35 & 36 Part 1, pp. 3-52.

CHAPTER 5 VERY HIGH RESOLUTION ALIGNMENT

5.1 INTRODUCTION

Alignment is necessary when a one pattern is required to be positioned very accurately with respect to another previously written pattern. The frame of the second exposure must coincide as nearly as possible with exposure frame of the first pattern when it was exposed. When the sample is placed in the SEM for the second pattern to be exposed, the new frame will be misaligned with respect to the old frame in terms of; X- and Y-displacement due to inaccuracies in positioning, rotation due to misplacement in the holder, and X and Y dimensions due to drifts in the scan amplifiers (Figure 5.1). These errors have to be corrected. This is normally done by exposing special features as part of the first pattern which act as reference marks for positioning the frame before exposing the second pattern. The beam is passed over the alignment marks under computer control, the marks are detected automatically and the position of the marks in the new exposure frame is found. By comparing the coordinates as found by the computer with the coordinates that the marks had when exposed, appropriate changes may be calculated and made so that the new frame is aligned with the old frame. The approach to alignment outlined above is used in many commercial electron beam writers, but it requires a fairly sophisticated computer to perform the calculations and control the electron column and stage [5.1,5.2,5.3]. The computing power of the KIM microprocessor is minimal, to say the least, and so the methods of alignment used in this work (both on thin and solid substrates (Section 6.5.2.2)) were designed to make errors in alignment visible on the monitor screen of the SEM, so that corrections could be made manually by the operator.

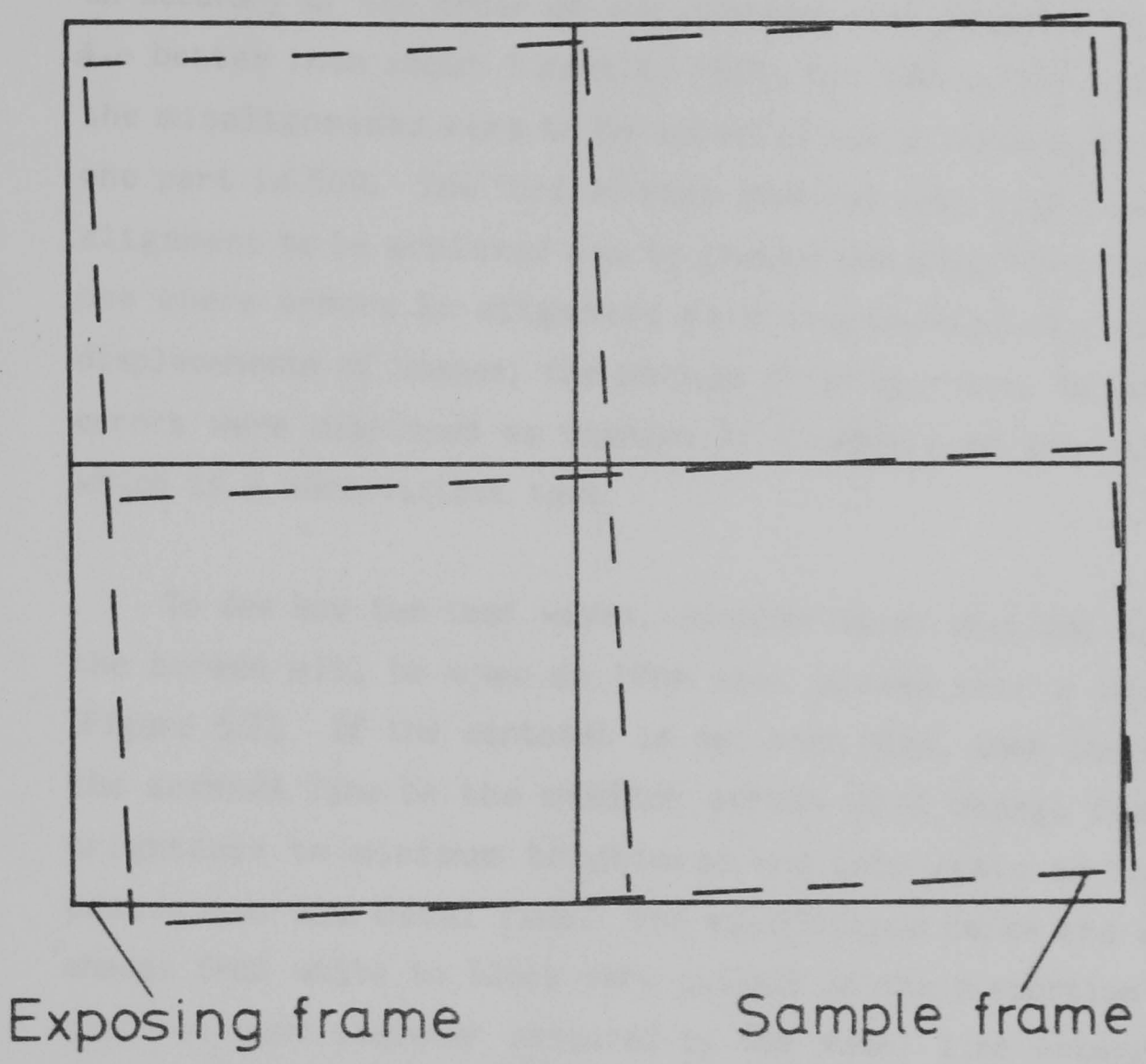
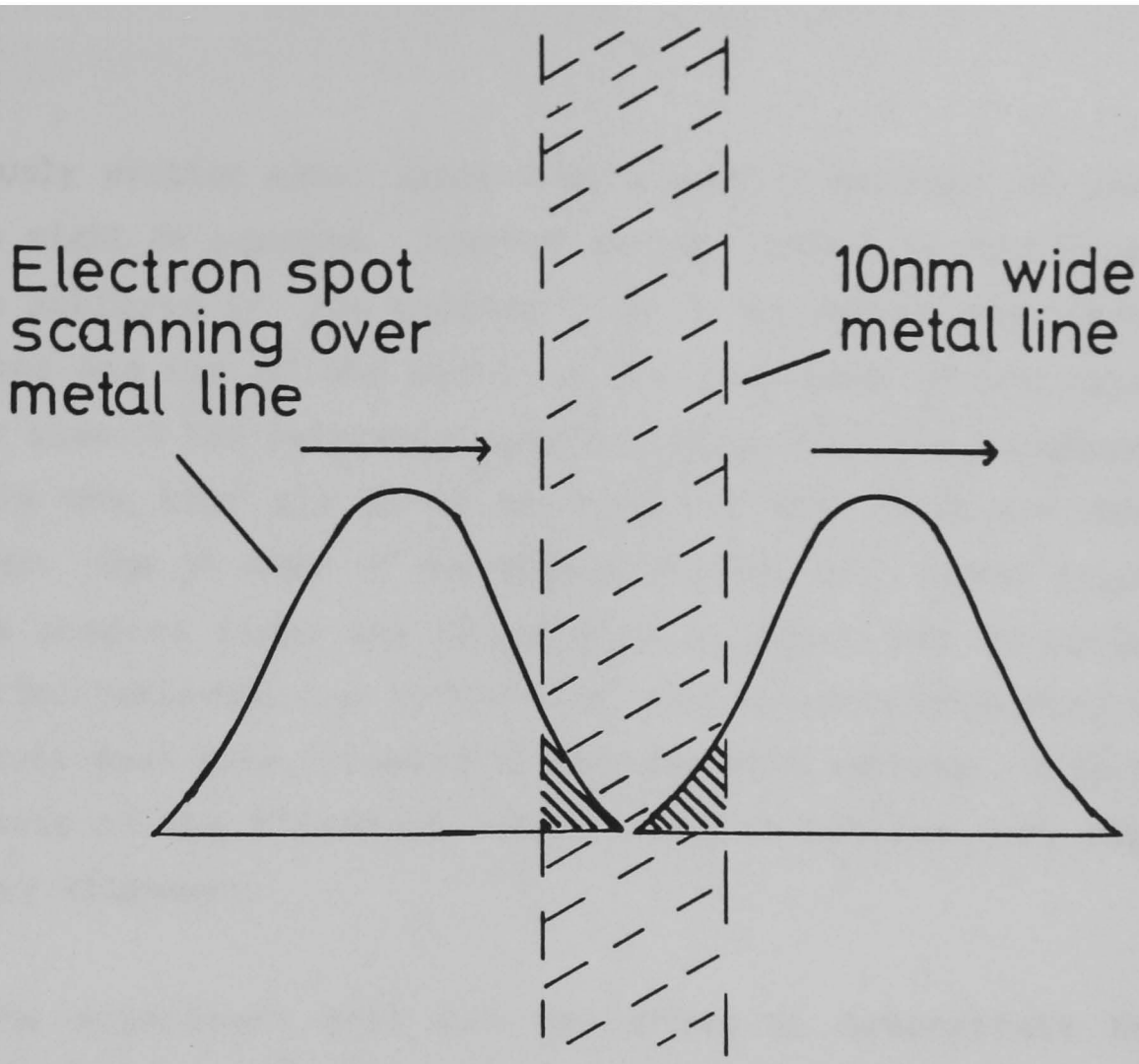


Fig. 5.1 When a sample is placed in the SEM the frame of a previously exposed pattern is misaligned with respect to the current exposing frame. Misalignments occur in position, rotation, and frame shape (X- and Y magnification). The alignment process involves the correction of these errors so that the two frames coincide.

5.2 ALIGNMENT METHOD

The problem was this; it was desired to perform alignment to an accuracy of the order of the electron spot diameter or better, i.e better than about 1 part in 2500, but the screen with which the misalignments were to be detected has a resolution of only one part in 500. The 'trick' that enabled very high resolution alignment to be achieved was to change the alignment test from one where errors in alignment were displayed on the screen as displacements of images, the obvious first approach, to one where errors were displayed as changes in intensity of scanned lines, which is a more visible test.

To see how the test works, consider first what the output to the screen will be when an 10nm spot passes over a 10nm line (Figure 5.2). If the contrast is set very high, then the tone of the scanned line on the monitor screen will change from full brightness to minimum brightness and back again as the beam passes over the metal line. The electronics cause the spot to change from white to black very quickly as the proportion of the electron beam diameter obscured by the metal line changes from about 15% to 25% i.e. the change from white to black takes place over a distance much smaller than the spot diameter. The resulting image on the screen is a bright line with a dark section in it, which if it could be measured would have a width corresponding to 18nm on the sample. Positional information to +/-9nm is therefore available if it can be made visible. If a pattern is being aligned at 5000x magnification, the dark region of a line scanned across a 10nm wide metal line would be about 0.1mm long, and since this is somewhat smaller than the spot size of the monitor, it would be very difficult to tell when the dark portion is in the position that corresponds to alignment. If however the scanned line were parallel to the metal line, then the alignment condition would be much more visible, either the whole line will be bright, or the the whole line will be dark. If the alignment test consists of aligning scanned lines with



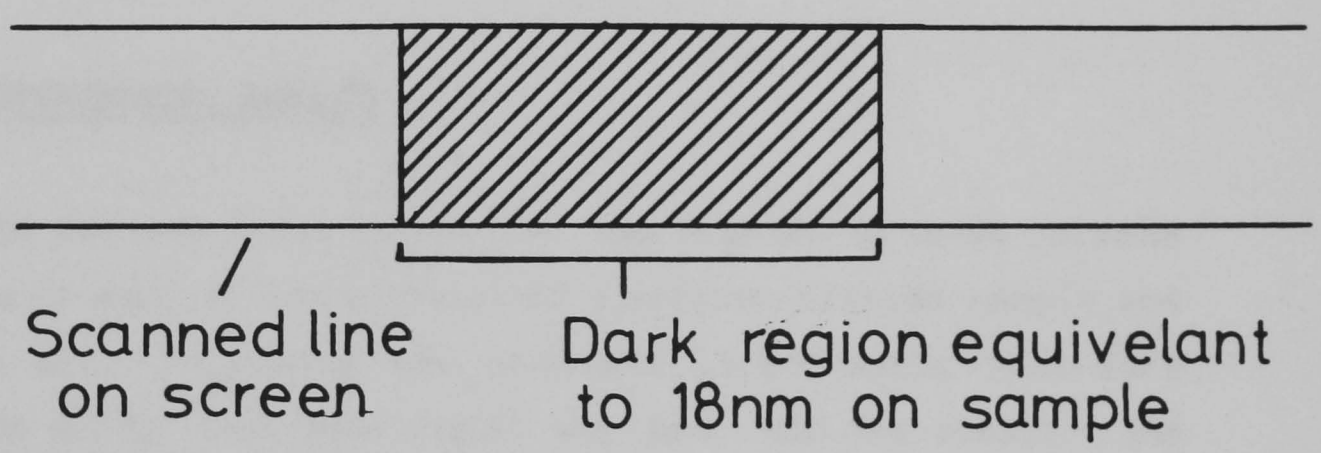


Fig. 5.2 When the 10nm electron spot of the SEM is passed over a 10nm wide line on a thin substrate, the resulting dark region in the scanned line on the monitor screen is equivalent to 18nm on the sample. This is because the electronics of the SEM, when set for high contrast, cause the displayed spot to change from bright to dark as the proportion of the electron spot obscured by the line goes from 15-25%. The change from bright to dark occurs in a distance much smaller than the spot diameter, and this fact is used in the alignment method to perform alignment to less than the spot diameter.

previously written metal lines then a overlay accuracy of about +/-9nm might be expected. However an improvement to this figure can be achieved if the scanned line is segmented into three sections and two of the sections are displaced by one pixel either side of the reference position (Fig. 5.3). The alignment test is now that all three sections of the lines are dark together. One or other of the displaced lines will appear bright if the scanned lines are misaligned by about 4nm in either direction perpendicular to the line, the distance depending on the exact spot size, linewidth, and contrast setting. This is the basis of the alignment method used to achieve very high accuracy alignment.

The experiment that was performed to demonstrate the capability of the method is described in the next section.

5.3 EXPERIMENTAL RESULTS

The pattern shown in Fig 5.4 was exposed in three columns of five on each of the windows of a silicon nitride sample and lifted off, following the standard lithographic procedure (Section 4.1.1); then more resist was spun onto the sample. The alignment marks can be seen in each corner of the pattern, the blocks are for coarse alignment and the T bars are for fine alignment. The sample was placed in the SEM once more and the frame positioned approximately over the first pattern using POSITION. In order to align the writing frame with the previously written pattern, the pattern shown in Fig. 5.5 was scanned repeatedly over the metal alignment marks. It consists of a horizontal and a vertical raster with the lines spaced 20 pixels apart so that the intensity of each line on the monitor can be seen without being affected by neighbouring lines. When the output from the transmission detector is viewed on the monitor, one of the blocks of each alignment mark was normally seen in each raster. Coarse alignment was performed by using the

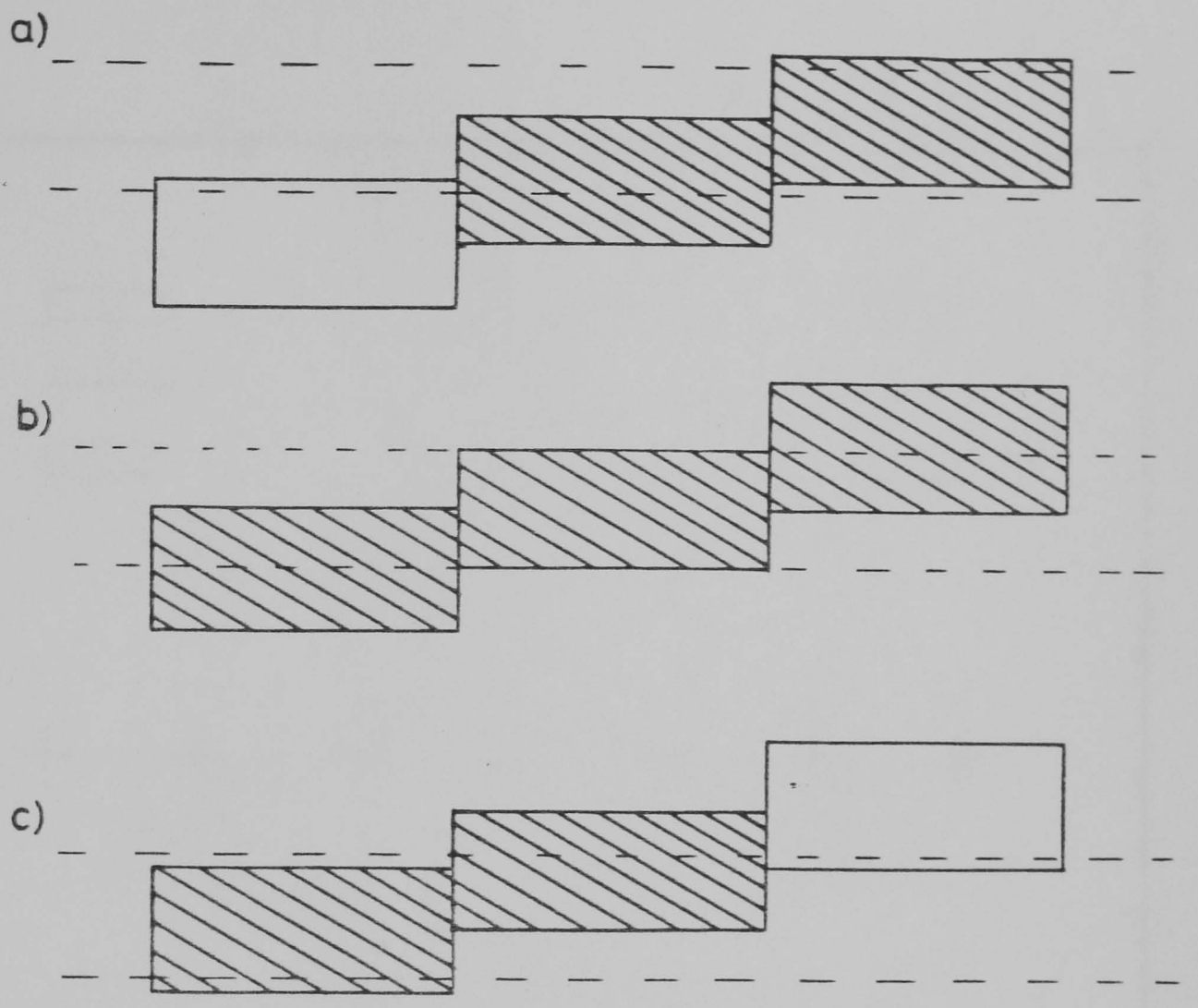


Fig. 5.3 Scanned reference lines for alignment consist of three sections. The centre section has the same position in the frame as the alignment mark had whan it was exposed, the other sections are displaced by one pixel either side of the reference position. A misalignment of only about 4nm will be enough to cause one of the offset lines to appear dark.

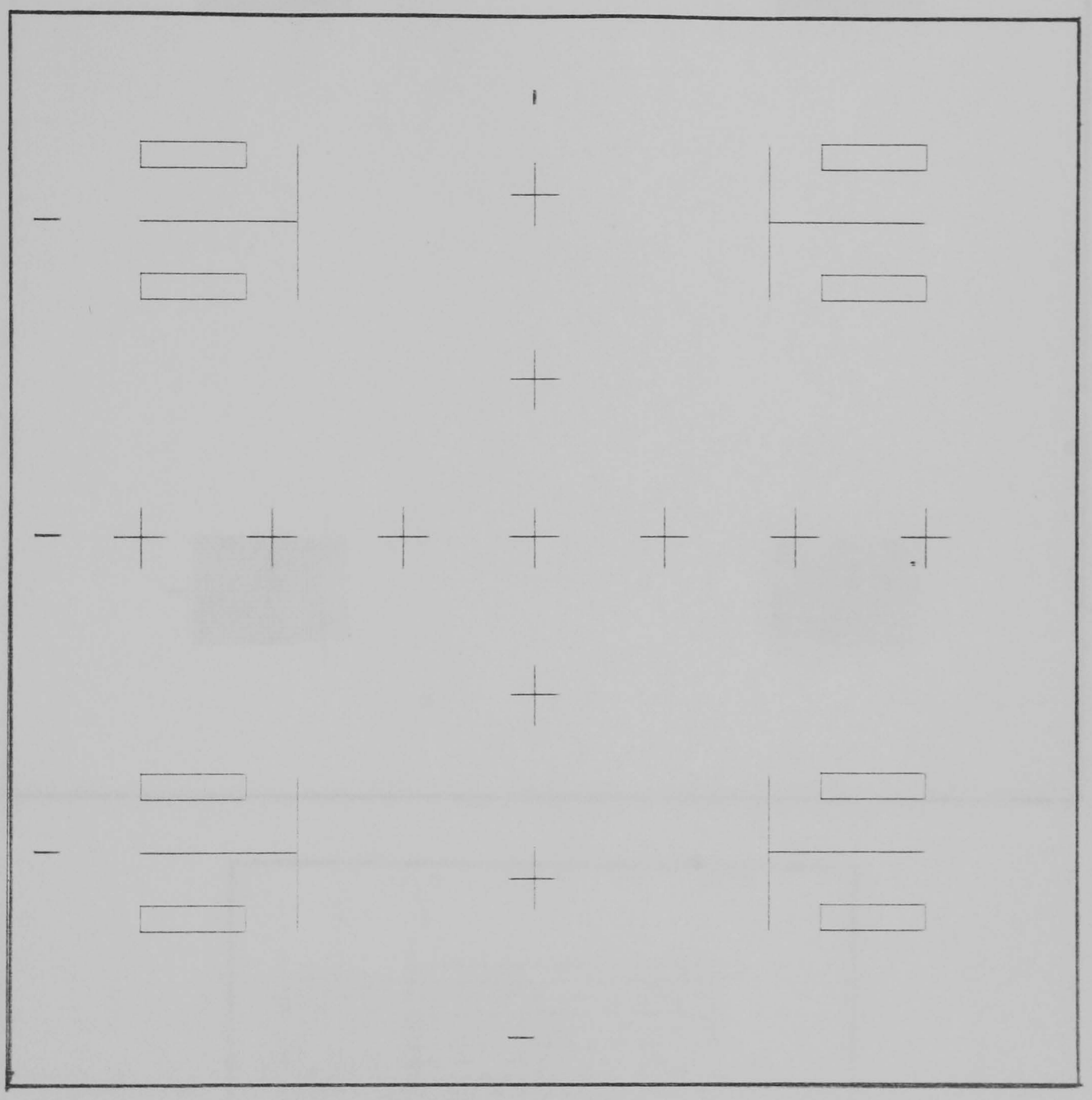
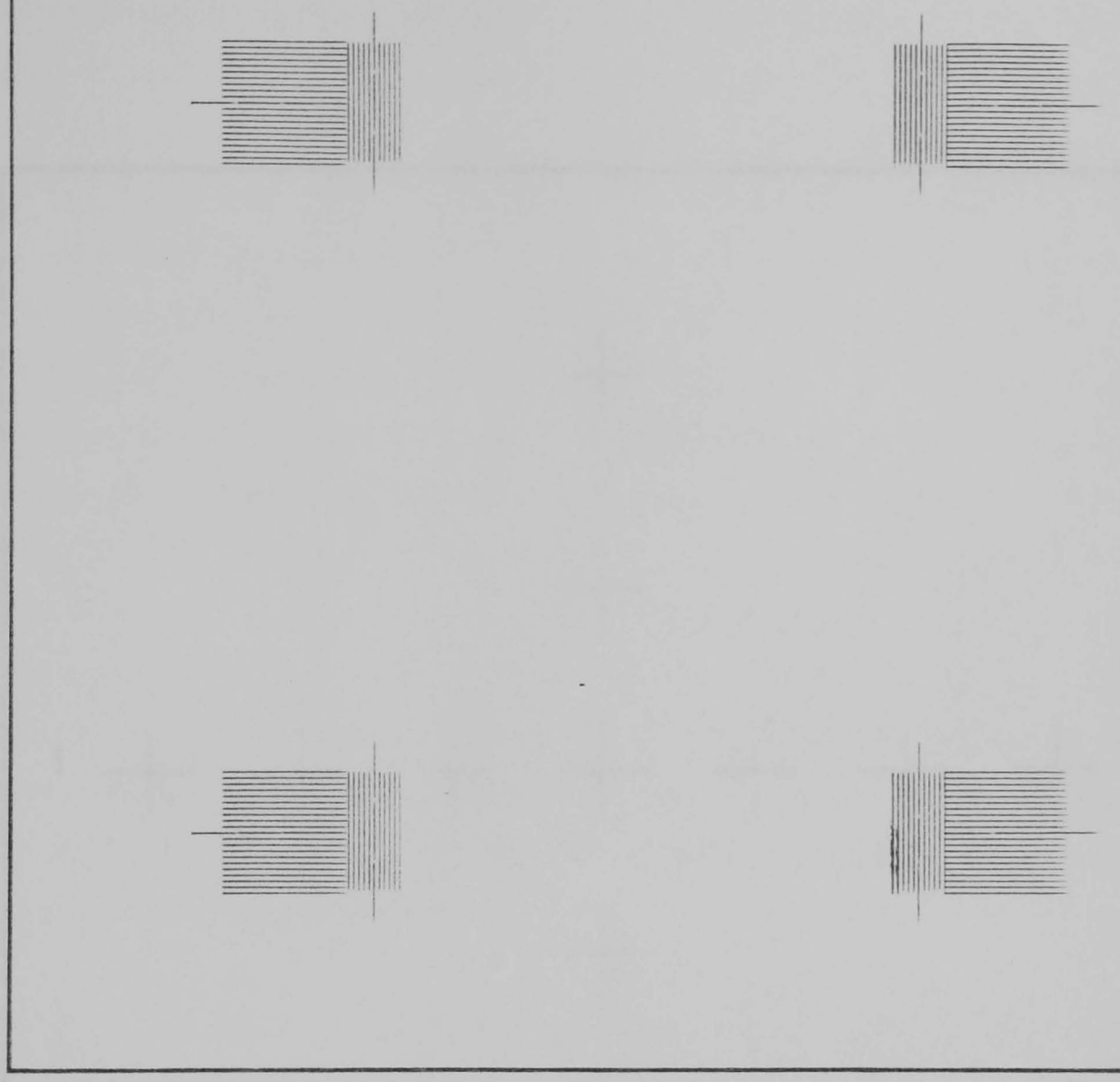


Fig. 5.4 Pattern used for first exposure. Alignment marks can be seen in each corner. The blocks are for coarse alignment, and the T-bars are for fine alignment.



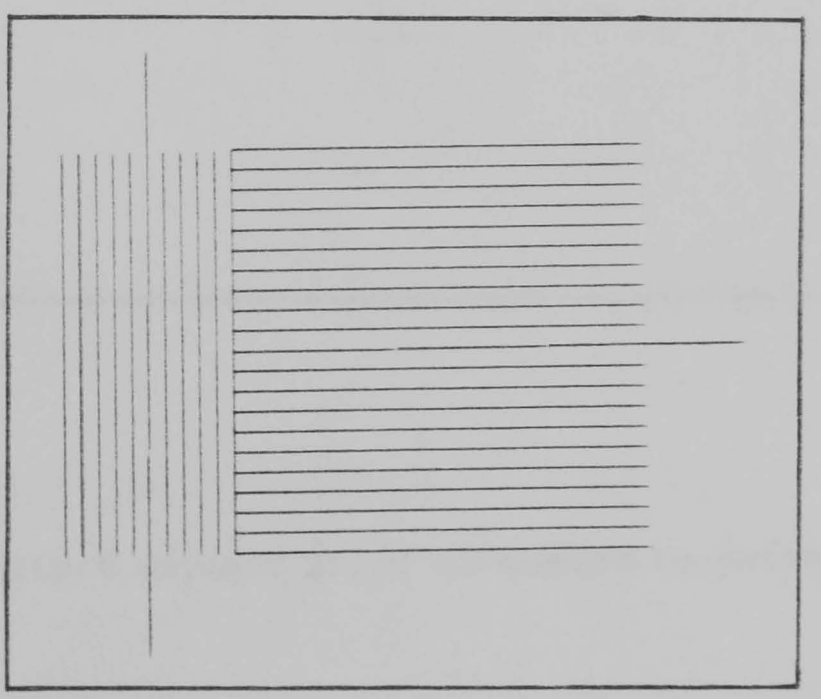


Fig. 5.5 Scanned alignment pattern. The extended line in each raster is the reference line. In the magnified view of one of the sections the three segments of the refernce lines may be seen.

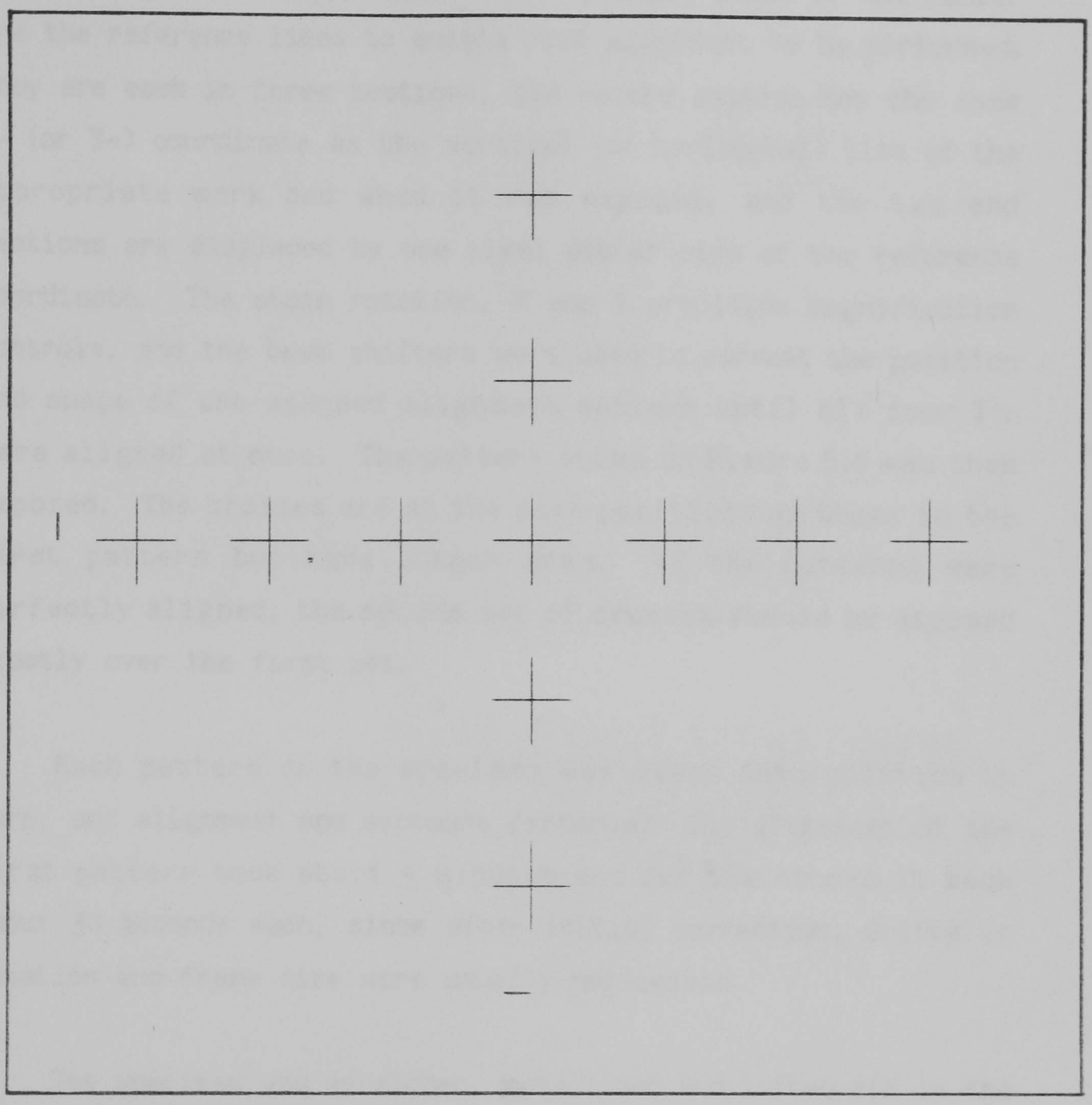


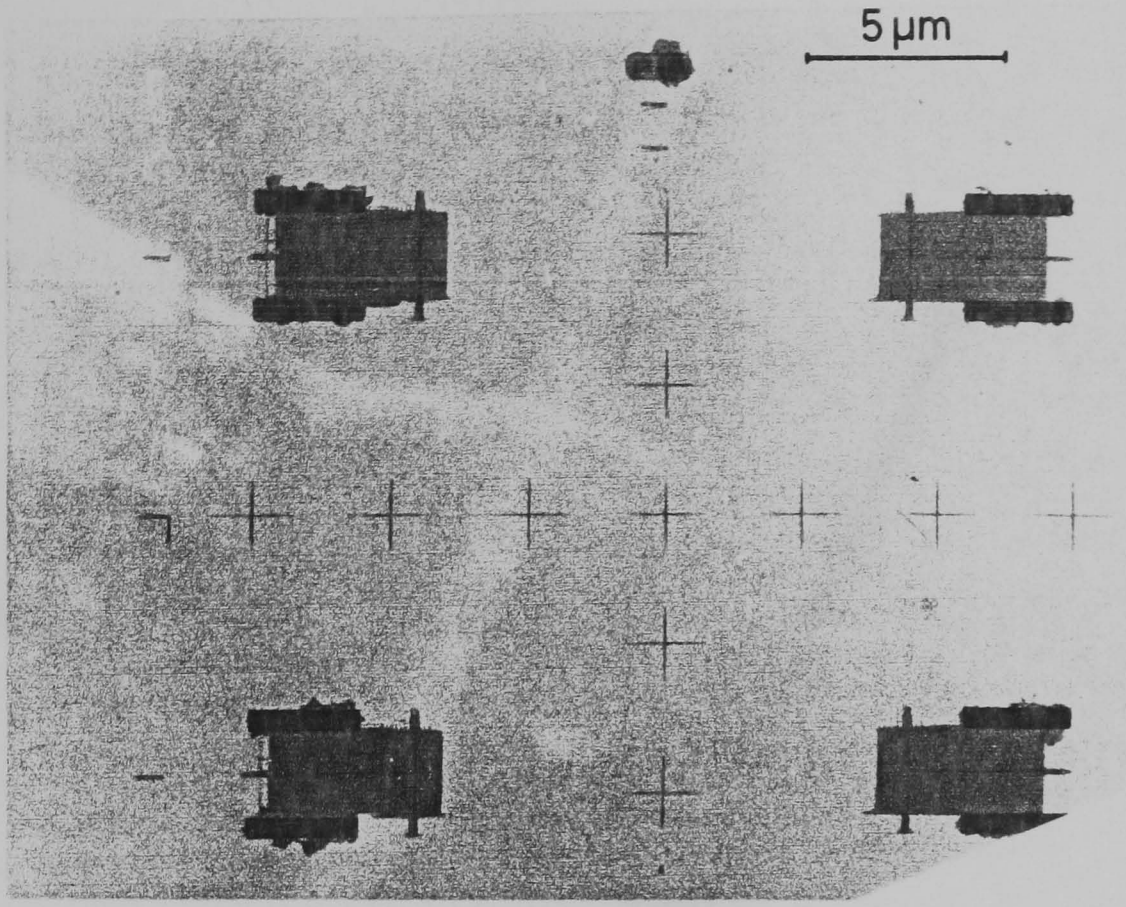
Fig. 5.6 Pattern exposed after alignment to pattern shown in Fig. 5.4.

beam shifters to place the blocks of each mark on either side of the corresponding horizontal scanned raster. This put the T of the mark close to the centre of each scanned raster. Individual lines of the raster could be seen to darken as the metal lines of the marks passed beneath them. The extended lines of the raster are the reference lines to enable fine alignment to be performed. They are each in three sections, the centre section has the same X- (or Y-) coordinate as the vertical (or horizontal) line of the appropriate mark had when it was exposed, and the two end sections are displaced by one pixel either side of the reference coordinate. The stage rotation, X and Y precision magnification controls, and the beam shifters were used to correct the position and shape of the scanned alignment pattern until all four T's were aligned at once. The pattern shown in Figure 5.6 was then exposed. The crosses are at the same positions as those in the first pattern but have longer arms. If the patterns were perfectly aligned, the second set of crosses should be exposed exactly over the first set.

Each pattern on the specimen was moved into position in turn, and alignment and exposure performed. The alignment of the first pattern took about 4 minutes and for the others it took about 30 seconds each, since after initial correction, drifts in rotation and frame size were usually negligible.

The specimen was developed, metallised and lifted off in the same way as for the first set of patterns.

Examination was conducted in a TEM. Figures 5.7 and 5.8 show typical results. In Figure 5.7 (upper) shows a complete exposure, all the crosses lie entirely on top of one another. Figure 5.7 (lower) shows a magnified view of a different exposure on the same sample. In Figure 5.8 one aligned cross is shown. The horizontal misalignment is too small to be accurately measured, the vertical misalignment is about 5nm. Accuracy of this order was maintained for all 15 patterns on the specimen.



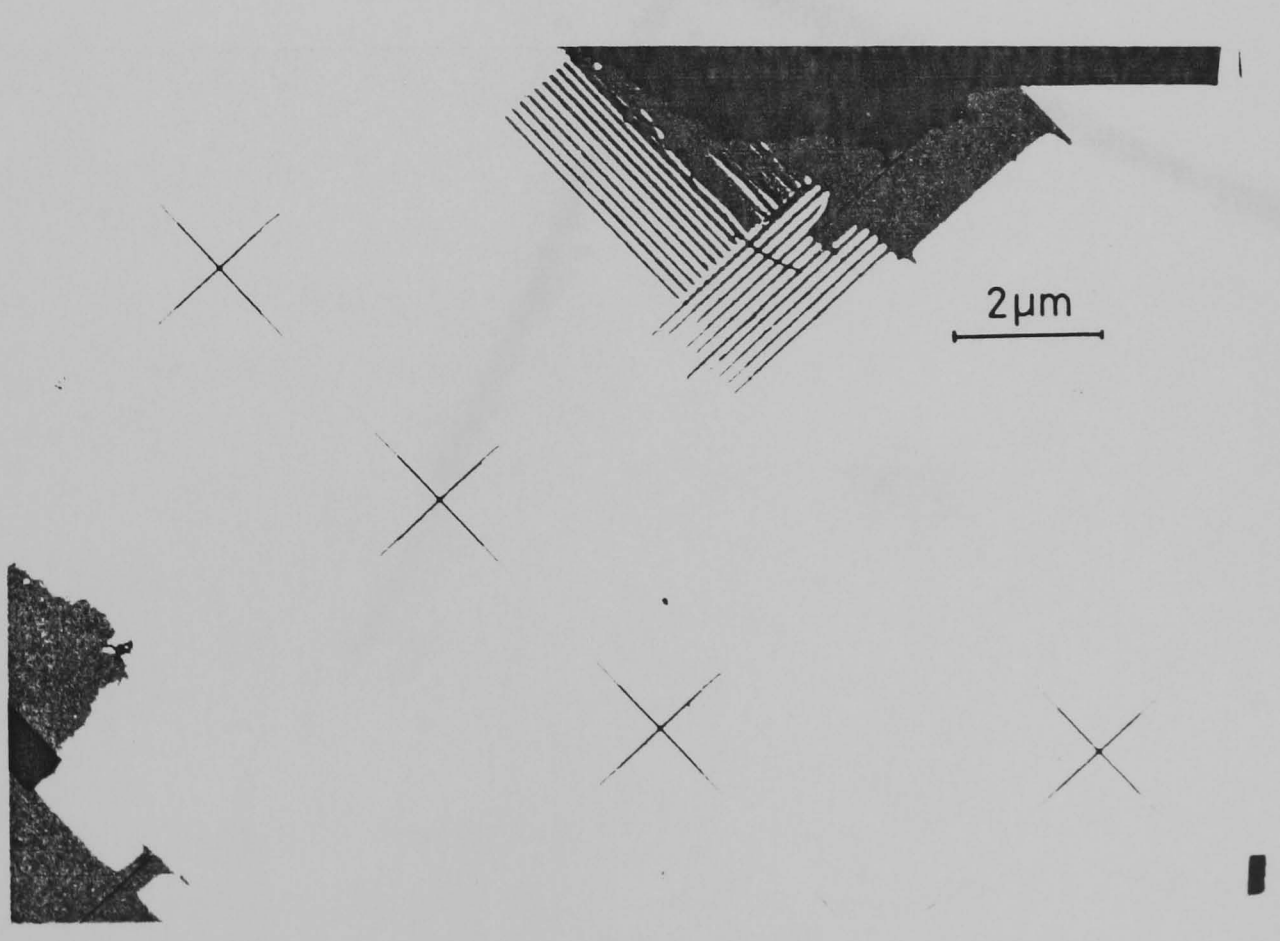


Fig. 5.7 (Top) Overall view of an aligned pattern.
(Bottom) Magnified view of a different exposure on the same sample.

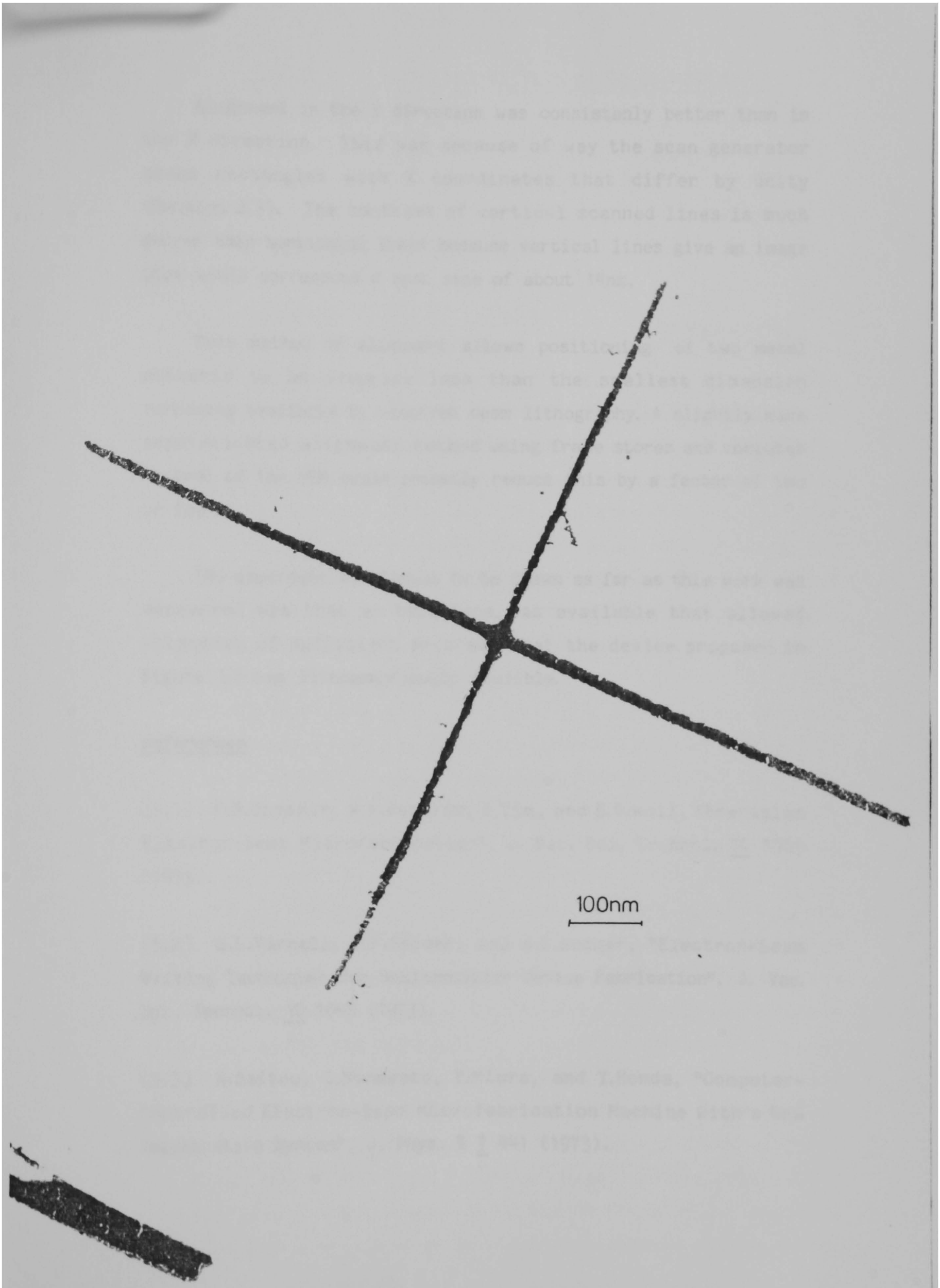


Fig. 5.8 A single pair of aligned crosses. Misalignment is about 5nm vertically, and too small to be measured horizontally.

Alignment in the Y direction was consistanly better than in the X direction. This was because of way the scan generator scans rectangles with X coordinates that differ by unity (Section 2.3). The contrast of vertical scanned lines is much poorer than horizontal lines because vertical lines give an image that would correspond a spot size of about 14nm.

This method of alignment allows positioning of two metal patterns to an accuracy less than the smallest dimension currently available by electron beam lithography. A slightly more sophisticated alignment method using frame stores and computer control of the SEM could probably reduce this by a factor of two or four.

The important conclusion to be drawn as far as this work was concerned was that an technique was available that allowed alignment of sufficient accuracy that the device proposed in Figure 1.6 was lithographically possible.

references

[5.1] F.S.Ozdemir, W.E.Perkins, R.Yim, and E.D.Wolf, "Precision Electron-Beam Microfabrication", J. Vac. Sci. Technol. 10 1008 (1973).

[5.2] G.L.Varnell, D.F.Spicer, and A.C.Rodger, "Electron-Beam Writing Techniques for Semiconductor Device Fabrication", J. Vac. Sci. Technol. 10 1048 (1973).

[5.3] N.Saitou, C.Munakata, Y.Miura, and Y.Honda, "Computer-Controlled Electron-Beam Microfabrication Machine with a New Registration System", J. Phys. E 7 441 (1973).

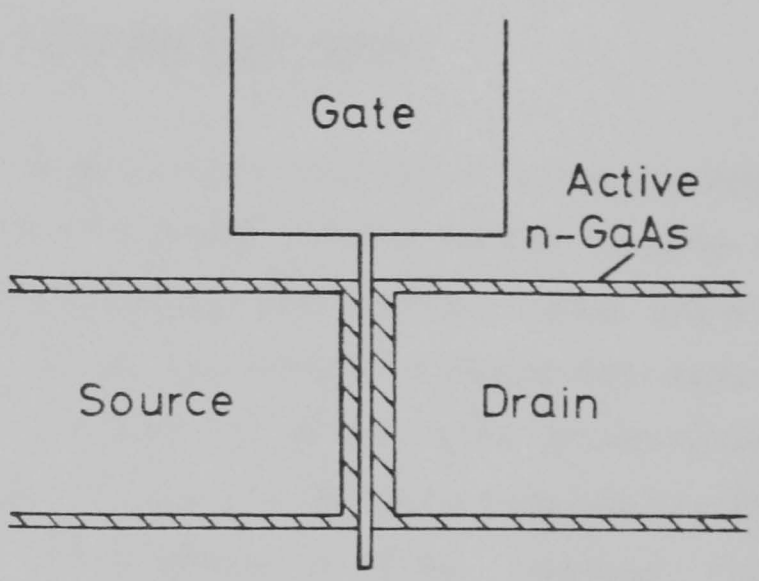
CHAPTER 6 FABRICATION OF FETS ON SOLID SUBSTRATES

6.1 THE SCHOTTKY GATE TRANSISTOR

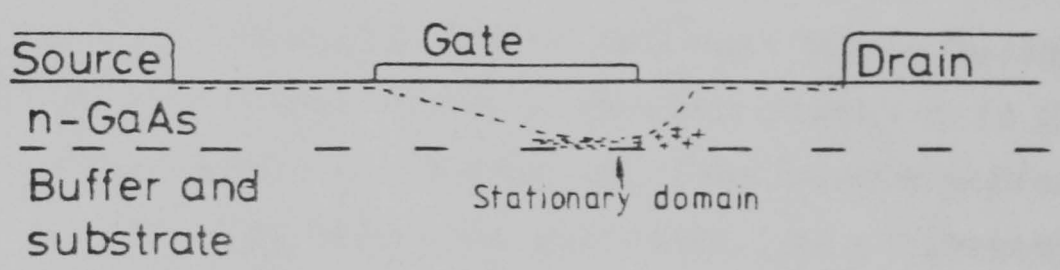
6.1.1 Structure of a Typical Device

A conventional solid substrate Schottky gate FET (or MEtal on Semiconductor FET (MESFET)) on GaAs is shown schematically in Fig. 6.1. Current flows through a thin n-type conducting layer between the two contacts called the source and drain, and is controlled by applying a voltage to the gate contact. The gate is a Schottky contact under reverse bias; the depth of the depletion region under the gate controls the current. Devices can be of two types a) enhancement type (normally-off), in which the built-in voltage of the gate fully depletes the channel under the gate and positive voltages are applied to the gate to reduce the depletion depth and allow current to flow, and b) depletion type (normally-on), in which the built-in voltage of the gate does not fully deplete the channel and current will flow from source to drain unless sufficient negative bias is applied to the gate to fully deplete the channel. The gate voltage for which current starts to flow in an enhancement device is called the threshold voltage, in a depletion device the equivelant is the gate voltage above which no channel current flows. Depletion devices have slightly greater fabrication tolerances than enhancement devices, so it was decided to attempt to fabricate depletion devices in this work. The conducting channel may be formed by ion implantation into a semi-insulating substrate [6.1,6.2], or by epitaxial growth using Liquid Phase Epitaxy (LPE), Vapour Phase Epitaxy (VPE), Metal Organic Chemical Vapour Deposition (MOCVD), or Molecular Beam Epitaxy (MBE) [6.3]. Ideally, the substrate material for the solid substrate transistors in this work should have been made from material

a)



b) Planar



c) Recessed gate

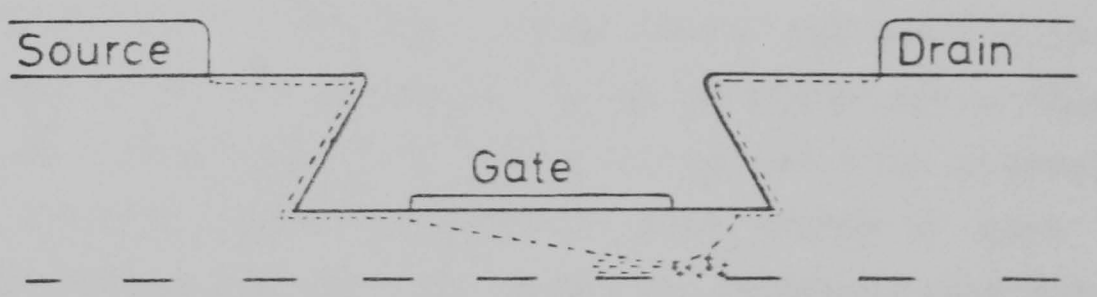


Fig. 6.1 a) Plan of a typical FET.

b) Planar device. The approximate position of the depletion region under the gate is shown as a dashed line. Note the finite depletion depth away from gate due to surface states. Conventially, the width of the gate stripe and the depth of the channel should be in a ratio of at least two for the gate to have good control.

c) Recessed gate device. The threshold voltage and saturation current can be controlled by the depth of the recess. Also, the parasitic resistance of the GaAs between the gate and the ohmic contacts is reduced when a recess is used. The thicker epi-layer may contain a thin highly doped layer at the surface to ease current flow to the region controlled by the gate still further.

similar to that which was to be used for the thin substrate transistors, i.e. grown by MOCVD or MBE, but due to the easier availability of VPE material (grown at Plessey (Caswell)), this was used instead.

6.1.2 Characteristics and Performance

Large MESFETs (with gate lengths of a few microns) can be understood in terms of a model similar to that used to describe a J-FET [6.4]. In this model, the electrons pass under the gate through two regions as the electric field increases, one of constant mobility, where the velocity is proportional to the electric field, and, after the velocity reaches the saturation velocity, a region of constant velocity. However, for smaller gates (say, less than 2 microns), the region of saturated velocity dominates, and the fact that at 3kV/cm the drift velocity in GaAs goes through a peak of $2x10^7$cm/s before falling to half this value at higher values of electric field has to be taken into account [6.5]. In equilibrium, a region of electron accumulation forms just before the narrowest part of channel followed by a region of electron depletion. These two space charge regions form a stationary Gunn domain and most of the drain voltage is dropped across it. It is the properties of this domain that cause saturation [6.6,6.7], rather than channel pinchoff. In even smaller devices, with gate lengths of about 1 micron, the effect of 'velocity overshoot' becomes important. Velocity overshoot occurs when electrons in GaAs enter a high-field region to be accelerated to the peak velocity before relaxing to the equilibrium velocity, so that for a short time the electrons are moving faster than the equilibrium velocity [6.8]. As device dimensions decrease, the velocity relaxation region occupies a larger fraction of the channel length. The effect improves the high-frequency response of small GaAs FETs. At even smaller dimensions (less than 1 micron), the dimensions of the device may be sufficently small that electrons encounter so few collisions in passing from one contact to another that

they are never reach an equilibrium condition. A device of this sort would be said to be exhibiting 'ballistic' characteristics. In a true ballistic device, the electrons do not collide at all with lattice atoms [6.9,6.10]. The question of whether a ballistic MESFET can be made is still open, but ballistic effects appear to have been seen in vertical FET-like devices when low operating voltages were used [6.11]. The regime that will apply for particular device dimensions will depend on the geometry and doping levels of the device in question, which control the mobility and electric fields in the device.

The devices to be made on solid substrates as part of this work were expected to be small enough to show the effect of velocity overshoot, and perhaps ballistic effect at the smaller sizes.

The first example of a GaAs MESFET that showed high frequency performance that was significantly superior to a silicon device of the same dimensions was demonstrated by Drangeid et al. [6.12]. It had a 1 micron gate fabricated by optical projection printing. Since then much effort has gone into producing devices with better noise figures using recessed gates in particular [6.13,6.14,6.15], and more recently into producing devices with significantly lower propogation delay times. The fastest devices for which figures have been published have been from Sadler et al. [6.16] (15-16ps with 0.6 micron gates) and Kato et al. [6.17] (about 17ps using 0.3 micron gates). Both these results are for devices that use a 'T' shaped gate as a mask for ion-implantation of high conductivity regions that reach very close to the gate contact. Interestingly, these devices are slightly faster than devices employing modulation-doped GaAs/GaAlAs heterostructures to give a two-dimensional electron gas with high mobility at the interface [6.18]

6.1.3 Requirements for Device Fabrication

6.1.3.1 Ohmic Contacts

The drain and source contacts have to be ohmic, that is the metal-semiconductor contact should have negligible resistance compared with the bulk or spreading resistances of the semiconductor in the device. The criterion to be satisfied for ohmic contacts in a device is that the voltage drop across the contacts should be small compared to the voltage drop across the active part of the device, i.e. the effect of the contacts should not be an important part of device performance. The ohmic contacts used in this work were formed out of Au/Ge or Au/Ge-Ni-Au. The contacts are Schottky type as-deposited and have to be alloyed at a temperature of between 300°C and 450°C. The fabrication of ohmic contacts is described more fully in Section 6.2.1.

6.1.3.2 Isolation

The function of the isolation step in the fabrication procedure of a GaAs FET is to force all the current between the source and drain to flow under the gate and not round about it (Fig. 6.2). Two methods of isolation were examined in this work: mesa etching and proton isolation. While both gave adequate isolation, processing difficulties (described below) prevented the use of proton isolation in any devices.

6.1.3.2.1 Mesa Etching

The simplest method of isolation (in terms of resources required, at least) is to etch away the conducting epi-layer through a resist mask so that each device is on an 'island' of conducting material [6.19,6.20]. On solid substrates the main disadvantage of the method is that the gate stripe has to climb a

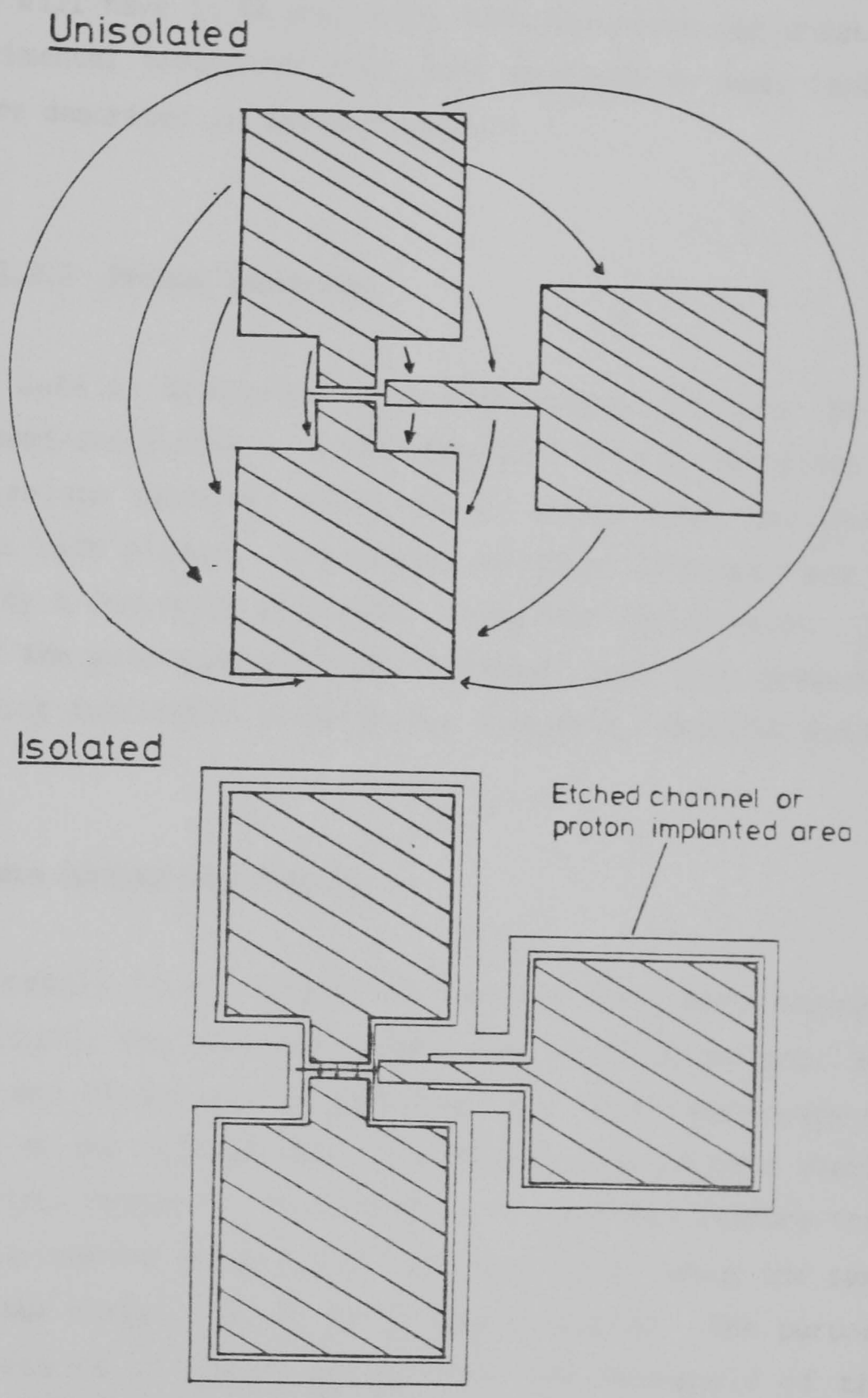


Fig. 6.2 Device isolation.

(top) If the pads are not isolated from each other, current will flow between them and not be controlled by the gate.

(Bottom) If the pads are isolated from each other by surrounding them with high resistivity material, either by etching away the conducting epi-layer or by implanting it with protons, then the current has no option but to flow under the gate through the only conducting material left between the drain and source.

step onto the mesa, and this can sometimes make liftoff difficult. All the devices fabricated in this work were isolated by a mesa etch step. The method will obviously be unsuitable for fabrication of devices on thin substrates, and a planar method of isolation will have to be employed, such as proton implantation. The experimental procedures that were employed to mesa isolate devices are described in Section 6.5.2.3.

6.1.3.2.2 Proton Isolation

When GaAs is bombarded with high energy protons, it is rendered semi-insulating ($> 10^6$ ohm-cm). This property can be used to isolate devices, with the advantage that the wafer surface is left planar. The active areas of devices have to protected by a removable gold mask during the implantation. The removal of the gold mask was the stumbling block that prevented devices being fabricated using proton isolation. (Section 6.2.2)

6.1.3.3 Gate contact and recess

Most metals form Schottky contacts on GaAs when vacuum-deposited [6.21], the ones most often used for gate contacts are aluminium and titanium, because they are relatively easy to process and do not diffuse into GaAs readily, (Ti is more stable than Al in this respect). The recess is an optional feature that is etched just before deposition of the gate metal using the same opening in the resist, and it is thus self-aligned. The purpose of the recess is to lower and control the threshold of the device, and also to lower the relative resistance of the parts of the channel not controlled by the gate. Recessing of gates might be possible on membrane devices if a very slow etch was used, and great care was taken.

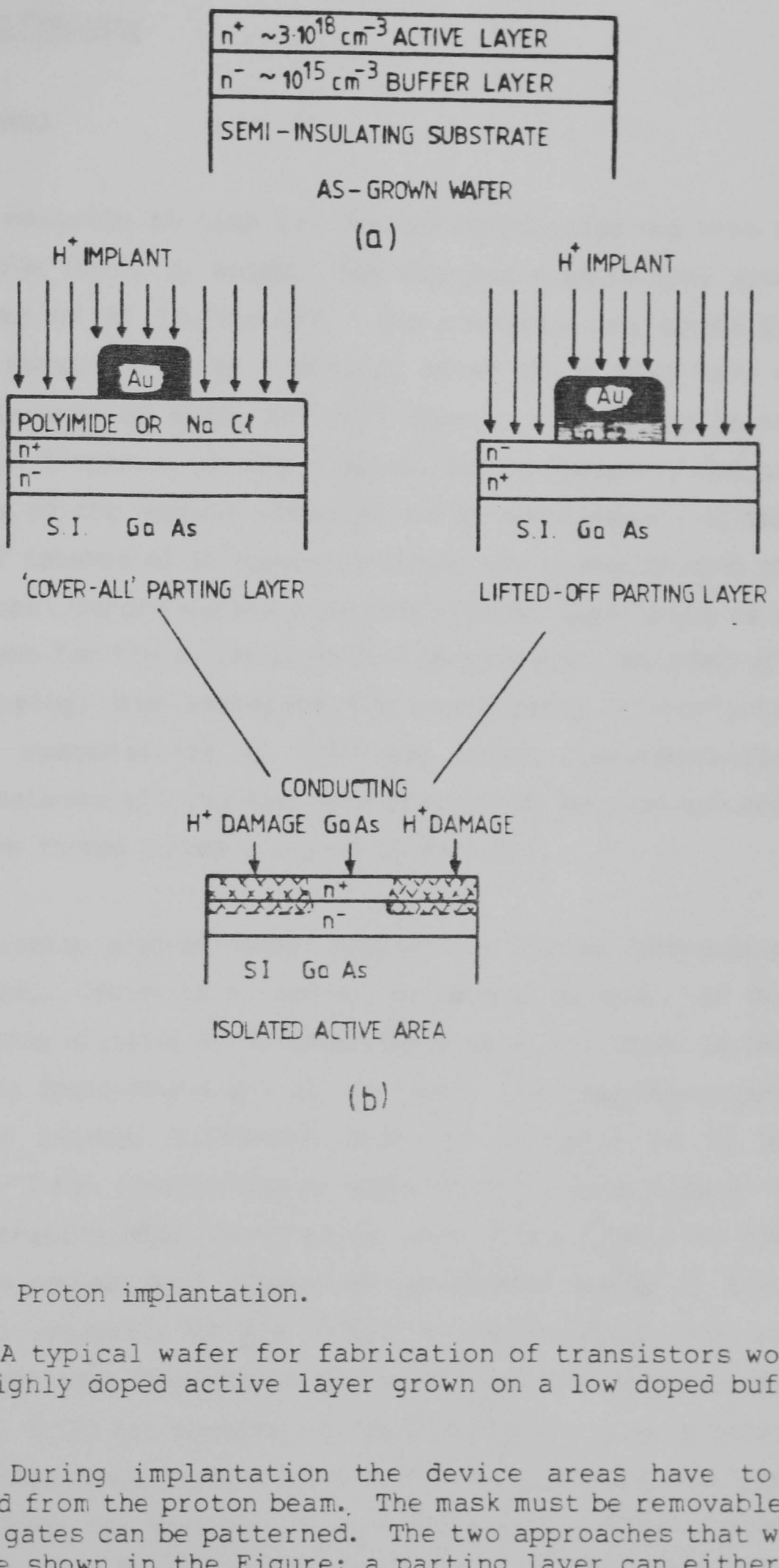


Fig. 6.3 Proton implantation.

a) A typical wafer for fabrication of transistors would have a highly doped active layer grown on a low doped buffer layer.

b) During implantation the device areas have to be protected from the proton beam. The mask must be removable so that the gates can be patterned. The two approaches that were tried are shown in the Figure; a parting layer can either be deposited over the whole sample, and the mask patterned on top of it, or the parting layer evaporated in the same vaccuum cycle as the gold of the mask and be lifted off with it. After removal of the mask by dissolving the parting layer, a conducting channel is formed, bounded by semi-insulating material or buffer material.

6.2 PRELIMINARY EXPERIMENTS

6.2.1 Ohmic Contacts

6.2.1.1 General

Ohmic contacts to GaAs are conventionally formed from a layer of Au/Ge (88:12 by weight, the eutectic combination) with an overlayer of Ni [6.22,6.23]. The contacts are normally alloyed at about 400°C for 1 minute, after which thay have a contact resistance of about 10^{-6}-10^{-5} ohm-cm. The mechanism of ohmic contact formation is very complex, but is basically due to the lowering of the contact potential to the GaAs because of the formation of intermetallic compounds during the interdiffusion of materials that occurs during alloying. The Ni layer acts as a wetting agent for the Au/Ge layer and so prevents 'balling up' during alloying, and increases the homogeneity of contacts alloyed at temperatures of 400° and above. Intermetallic compounds between all the elements present in the contact and substrate are formed during alloying [6.24,6.25].

The alloying process causes problems as far as fabrication of a very small device is concerned, as lateral movement of the contacts during alloying could cause shorting of the drain to the source if the drain-source gap is too small. Two approaches are possible if lateral diffusion from the contacts is to be minimised; 1) the contact can be annealed for a long time at a lower temperature than is normally used in the hope that the atoms in the contact will acquire enough thermal energy to form the required compounds for the contact to become ohmic, but not so much energy that they can diffuse away from the contact edge [6.26,6.27], or 2) the contact can be annealed for a very short time (a few seconds), using a powerful lamp, laser, or electron beam to provide the heating, so that the atoms in the contact simply do not have time to diffuse into the surrounding GaAs.

Very little rigourous experimentation was done to optimise fabrication of ohmic contacts in terms of minimising contact resistance, edge roughness and lateral diffusion of the contact materials during alloying. The aim of the experiments that were performed was to establish whether low temperature annealing of Au/Ge contacts without the overlayer of Ni could produce ohmic contacts with negligible lateral diffusion. Contacts without Ni were examined since as-deposited contacts with a Ni overlayer have slightly larger grain size than those without, and this was expected to be important in a very small device.

6.2.1.2 Experimental Procedure and Results

The experiment that was performed consisted of laying down a pattern of 100 micron squares of Au/Ge (88:12) by conventional photolithography and liftoff. The layer was 100nm thick. The substrate was part of a wafer that had been grown by MOCVD for membrane fabrication with a structure similar to that shown in Fig. 3.13. The 5 x 5 mm patterned sample was divided into chips that measured 0.8 x 0.7 mm. Each was alloyed for 1 minute at temperatures ranging from 250°C to 400°C. The heating was performed by a strip heater in a reducing atmosphere of H_2 (5% in Ar). After annealing, adjacent pads were probed and the I-V charactaristic examined on a curve tracer. If the characteristic was a straight line then the contact was deemed ohmic, otherwise it was still Schottky type. It was found that ohmic behaviour resulted for a range of alloying temperatures from 270-370°C, which was consistant with other published work on the subject [6.26]. Later experiments by others in this Department appear to show that the thickness of the active layer plays an important part in determining the minimum temperature for ohmic contact formation. It was found to be much easier to form ohmic contacts on a layer 80nm thick than on one 600nm thick. When quantitative experiments were performed to find values for the contact resistivity for various alloying conditions, it was found that the lowest that could be obtained for Au/Ge contacts was 10^{-4}ohm-

cm compared with published figures of less than 10^{-5} ohm-cm for Au/Ge-Ni contacts alloyed at 400°C [6.22,6.23,6.28]. It was therefore decided to abandon Au/Ge contacts for the time being at least and to use a conventional contact of 100nm Au/Ge, 20nm Ni, and 100nm Au. Solid substrate transistors were made using contacts alloyed at 330°C and at 400°C. The higher temperature contacts gave much improved device performance (7.2).

6.2.2 Proton Isolation

It is desirable to keep the surface of a wafer planar during processing. Etching away unwanted conducting material between the active areas of devices has been a useful and easily achieved method of isolating transistors fabricated on the same chip from each other, but in creating the isolated islands of material, undesirable steps are etched with a depth slightly greater than the thickness of the epitaxial layer. At the edges of the mesas local resist thinning occurs. This can cause difficulties when trying to lift off the gate of a MESFET which has to cross the mesa edge.

Proton implantation offers a means of rendering the areas of material between active areas highly resistive while maintaining a planar surface.

When high energy ions enter a semiconductor they cause damage as they lose energy through inelastic collisions. In GaAs the damage occurs in such a way as to cause many efficient deep traps. The effect of this is that if the correct proton dose is used, the sheet resistivity of an epitaxial layer can be increased from a few ohm-cm up to the order of 10^6 ohm-cm, which is sufficiently high that leakage currents are reduced to an acceptably low level [6.29,6.30].

During the implantation, regions that are to remain active have to be protected from the beam by a gold mask. The gold has

to be at least as thick as the layer to be implanted, since the penetration of protons in Au is only slightly less than for GaAs.

After implantation, the gold mask has to be removed from the surface of the GaAs in order that the gates can be laid down. If the gold mask is deposited directly onto the GaAs then it will have to be removed by a fairly lengthy etch step. GaAs is itself etched by gold etches, so an alternative approach is required. The gold is deposited onto an intermediate parting layer of a material that can be dissolved in a solution that does not attack GaAs. The gold should then be able to be lifted off by etching or dissolving the parting layer.

Two approaches to the parting layer are possible (Fig. 6.3); a) a 'coverall' parting layer that is deposited over the whole surface of the sample, and upon which the gold mask is patterned, or b) a parting layer that is patterned by the same resist stencil as the gold mask, and lifted off with it. Only data on a lifted off mask has been published [6.31]

Experiments were performed, first to find the correct ion dose to render samples semi-insulating, and then to test some parting layers for suitability. All samples were implanted at the SERC Implantation Facility at Surrey University.

6.2.2.1 Experiments with No Mask

Samples were sent for implantation with ohmic contacts on them but no mask. The proton energy was set to 650kV and the beam passed through an 18 microns of aluminium in order to reduce the average energy of the beam and also to increase the energy spread. The nominal dose at the foil was 4.10^{14} ions/cm^2. The dose at the target was not measured. The ohmic contacts were probed before and after the proton bombardment. The current flowing between adjacent contacts with an applied votage of 5V was 10mA before implantation and less than 1 microamp (the

resolution of the curve tracer used) after implantation. This indicated that implantation did work as an isolation method.

6.2.2.2 Experiments with Different Parting Layers

Three parting layer materials using both the 'coverall' and 'lifted off' approaches were tried to find one that worked. Polyimide and sodium chloride were tried as 'coverall' parting layers and calcium fluoride was tried as a 'lifted off' parting layer.

6.2.2.2.1 Polyimide

It had been shown by Grabbe et al. [6.32] that metal stencils on top of a layer of polyimide could be lifted off by dissolving the polyimide in acetophenone. The polyimide used was Du Pont XU218, a pre-imidised polyimide that does not cross-link if it is baked at less than 200^{o}C and remains soluble in certain organic solvents.

An experiment was performed to assess the suitability of the polyimide as a parting layer for a proton implantation mask.

-A GaAs sample was patterned with an array of 100 micron diameter gold/germanium dots by photolithography and liftoff. The gold/germanium was alloyed at 320^{o}C for 1 minute to give ohmic contacts to the GaAs.

-The sample was spun with a layer of XU218 polyimide about 0.5 microns thick and baked at 180^{o}C for 20 minutes.

-The sample was scribed into four parts. One piece was used as a control to check that the polyimide dissolved in acetophenone when no further processing had been done.

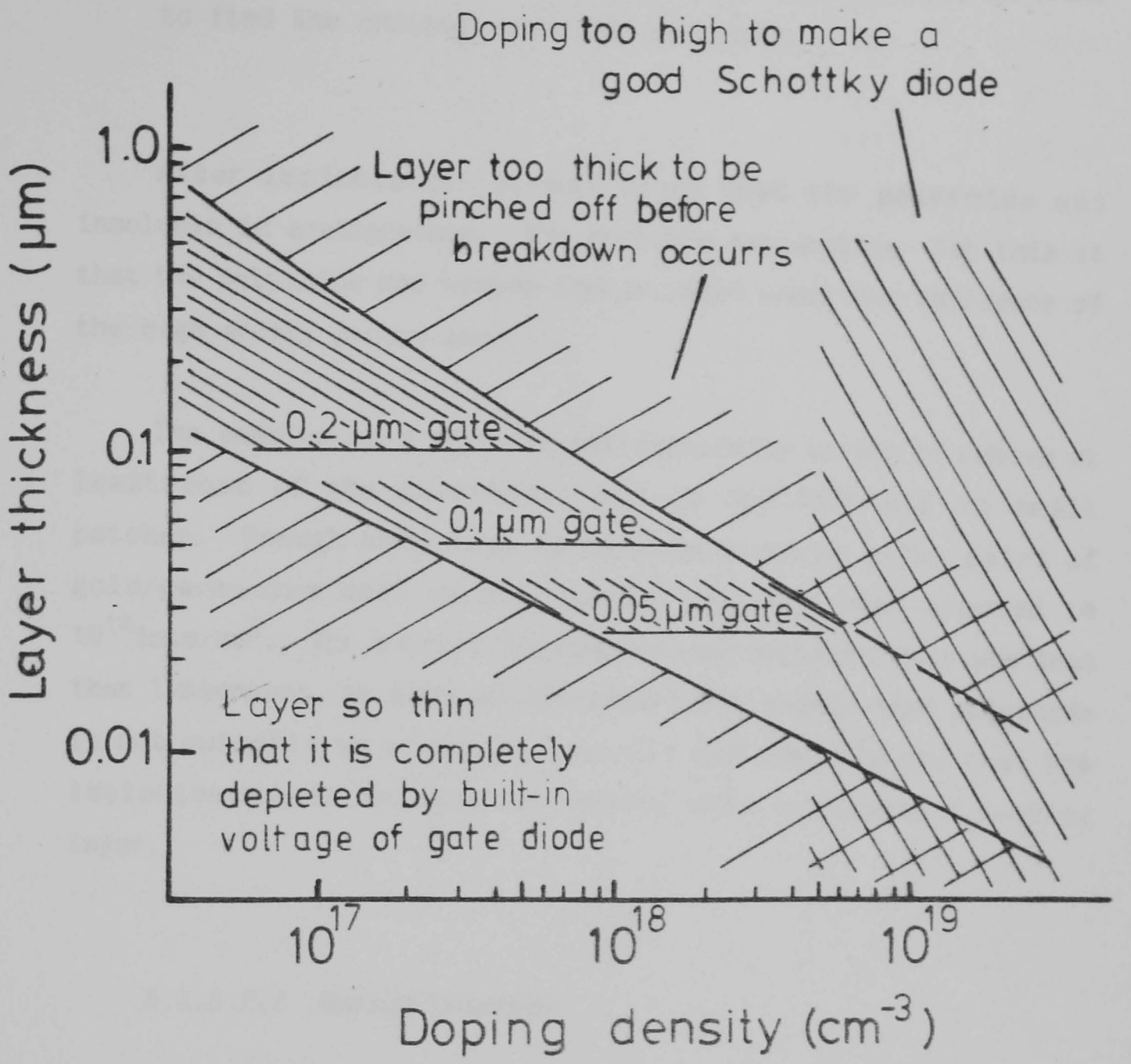


Fig. 6.4 Plot showing allowed values for thickness and doping of active layer. If the layer is too thin it will be completely depleted by surface states (lower boundary), if it is too thick, the device will not pinch off before the reversed-biased diode forming the gate breaks down (upper boudary), if the doping is too high, the diode of the gate will have 'soft' charactaristics. The final constraint is that that the ratio of gate length to channel depth should be greater than two. The upper bound of layer thickness is shown for three values of gate length. In reality, the doping is chosen to be as low as possible, in order to maximise the breakdown voltage of the gate.

-Three samples were implanted at the SERC ion implantation facility at Surrey University. The doses were 10^{13}, 5.10^{13}, and 10^{14} ions/cm^2, all less than the previous dose, in order to find the critical dose for isolation.

After implantation it was found that the polyimide was insoluble in acetophenone. The probable explanation for this is that the polyimide had become cross-linked under the influence of the high energy proton beam.

The samples were agitated ultrasonically to try to remove at least some of the polyimide, and it did lift off in small patches. Enough came away to be able to probe a few pairs of gold/germanium dots on the sample that had been exposed to 10^{14}ions/cm^2. The current flowing between adjacent dots was less than 1 microamp, so although the experiment showed that polyimide is not suitable as a parting layer it did demonstrate that the isolation itself had been successful with a 'coverall' parting layer.

6.2.2.2.2 Sodium Chloride

Sodium chloride was thermally evaporated from a tungsten boat onto a GaAs sample. A layer of PMMA was spun onto the sample and baked. A test pattern was exposed and developed and 0.2 microns of aluminium was lifted off. The aluminium pattern could be removed easily by dipping the sample into water and agitating in an ultrasonic cleaner. When the experiment was repeated in more humid conditions the sodium chloride absorbed enough water from the atmosphere during the time between evaporation and resist spinning, and between resist development and metal evaporation, that during lift off the metal lost all adhesion and none remained on the surface. Sodium chloride was therefore abandoned as a suitable parting layer because of its sensitivity to atmospheric conditions.

6.2.2.2.3 Calcium Fluoride

The process attempted was similar to that used by D'Avanzo [6.31]. Calcium fluoride is virtually insoluble in water, but is slightly soluble in acids, notably hydrochloric acid. Instead of covering the whole sample with the parting layer as before, the parting layer is evaporated immediately before the gold absorbing layer during the same vaccuum cycle. The parting layer is thus patterned together with the gold by liftoff.

Two major problems were found. The first occurred when thick gold (>0.1 microns) was deposited for lift-off. Any part of the exposed pattern with dimensions greater than 50 microns lifted off completely leaving no gold at all in those regions. This was probably due to the fairly high stress in the evaporated gold film combined with its relatively poor adhesion to the calcium fluoride even though nichrome was used as a sticking layer. The stress in evaporated gold films can be reduced by reducing the evaporation rate. The films used in these experiments were evporated at a rate of 2nm/sec as measured with a quartz crystal thickness monitor. Slower evaporation rates were not attempted because the process failed at the next hurdle even for small features that did lift off properly.

The gold masking layer is decribed by D'Avanzo being "easily removable" in a weak solution of HCl, but this was not found to be the case here. When the removal process was attempted it was found that the gold could only be removed by placing the sample into a beaker of concentrated (30%) hydrochloric acid which itself was being agitated in an ultrasonic cleaner. The gold was removed after about five minutes.

A GaAs epi-layer sample with a gold absorber pattern on it was sent for proton implatation. It was implanted in the manner described in Section 6.2.2.1 with a nominal dose of 10^{14} ion/cm^2. The gold could not be removed even after 10 minutes of etching.

Two reasons for the failure could be that the beam affected the calcium fluoride in some way to render it insoluble in hydrochloric acid, or perhaps that the calcium fluoride had reacted with something in the atmosphere during the four weeks between the proton mask being fabricated and the attempt at its removal after implantation.

Some differences in the procedure used here compared with D'Avanzo were that the thicknesses of the gold and of the calcium fluoride were less than half D'Avanzo's values (since a thinner active layer was being implanted), and nichrome was used in place of silicon dioxide. This second difference may account for the peeling off of large gold features if the SiO_2 used by D'Avanzo was deposited as a stress relieving film, as it might have been if it were deposited under conditions that lead it to be under compressive stress. However, the failure to be able to remove the CaF_2 in HCl is more puzzling, and would appear to indicate some difference in the CaF_2 used here and by D'Avanzo, perhaps in the water content of the crystals.

Failure to achieve a removable proton implantation mask was the main reason that a thin membrane transistor was not fabricated.

6.3 DESIGN OF TRANSISTOR PATTERNS

6.3.1 Design Requirements

A rather simple empirical approach was used in the design of the devices used in this work. The device design had to meet a number of criteria. The pinch-off voltage must be less than the breakdown voltage of the Schottky diode that forms the gate. Also the device must not be pinched off at zero gate bias due to the built in voltage of the diode. The doping of the active layer must not be higher than about 5.10^{18} cm^{-3} or the diode formed on such a layer will be very 'soft', i.e. the leakage current on

reverse bias will be large. These conditions are met by choosing a correct combination of active layer thickness and doping density. The allowed values for doping density and layer thickness for 3 gate lengths are illustrated in Fig. 6.4. The lower limit to the thickness of the layer occurs (for a depletion device) where the device would be pinched off by the built-in potential of the gate (about 0.8V) [6.33]; the upper limit is where the device cannot be pinched off before breakdown of the gate occurs [6.34]. The constraint that forces layers to be thin and with high doping density is that the ratio of the gate length to the channel thickness under the gate should ideally be greater than 2:1 to give good control of the current flowing under it. According to this, gate lengths of 0.1 microns would require a layer thickness of only 50nm, rather hard to grow by any technique other than MBE or MOCVD. In reality, substrate requirements have to be tempered by the material that is actually available, and the layer on which transistors were in fact made in this work was 80nm thick with a doping density of $3.10^{18}cm^{-1}$ which according to Figure 6.4 should be too thick to pinch off. This was found to be the case when planar transistors were made (Section 7.2). The layer was grown by VPE, rather than MOCVD or MBE, due to the easier availability of such material. The disadvantage of such material is that the interface between the undoped buffer layer and the doped active layer is about 20nm thick, which is relatively thick compared to the active layer, particularily if a gate recess is etched. Conduction through the interface layer might have a large effect on the charactaristics of a recessed-gate device (see Section 7.3).

6.3.2 Exposure Frame Sizes

A trade off always has to made when using electron beam lithography. Ideally one would like to be able to expose high resolution features over a large field, but one is limited by the number of picture points available, in this case 4096 x 4096. Generally the field size that is chosen is the largest one that

is consistent with resolving the smallest feature required in the pattern. As is seen in Sections 6.5.2.2, the spot size may, in certain circumstances, require to be smaller than the pixel size, if alignment scans are not to affect the exposed pattern. This may not matter if, as the case of the mesa level, the edge quality is not particularily important, or, as in the case of the gate level, critical features are sufficiently overexposed to remove edge roughness caused by the finite size of pixels.

The first requirement of the layout of the transistor pattern was that it should contain as many transistors as possible, in order that the average time taken in the fabrication of a transistor was minimised, and in order that the GaAs substrate material was used efficiently. These two goals were realised by maximising the exposing frame size (and with it the exposing spot size) and minimising the area taken by a single transistor.

It was reasoned that the resolution (narrowest line or space) required for the ohmic contacts and the isolation step would not be required to be less than 1 micron even for 2nd generation devices (devices designed for rf testing for instance). Therefore a spot size of 0.5 microns was used at an indicated magnification of 80x for the ohmic contact level giving a frame size of 1.5 x 1.2mm and a pixel size of 380 x 290nm. A spot size of 0.25 microns was used when exposing the isolation level and the gate level when all the gates were exposed together to give more time for alignment (Section 6.5.2.2). Narrow gate stripes were exposed individually using an indicated magnification of 640x which gives a frame size of 195 x 150 microns and a pixel size of 48 x 36 nm. A spot size of 32nm was used in order to be able to align to pattern features without exposing them significantly.

6.3.3 General Layout of Transistor Pattern

Once the frame size for the transistor pattern had been chosen, the next objective was to design a layout that would maximise the number of transistors per exposure. The main constraint acting on the number of devices within a frame is the size of the probing pads through which each device is tested. These have to be at least 100 x 100 microns if they are to be reliably probed or bonded to. The design that was eventually chosen is shown in Fig 6.5. The drain and source contacts of adjacent devices have been commoned to save space. Transistors are arranged in five columns of eight with the device width increasing from column to column from left to right. The channels are 30,20,10,5, and 2 microns wide. Devices of various widths were included in the design to find out if the resistance of very narrow gate stripes had a significant effect on device performance. The length of the gates could be varied over the eight transistors in a column so that information on the effects of channel width and gate length can be gained from each chip.

6.3.4 Design of Ohmic Contact Level

Two decisions had to be made concerning the dimensions of the drain and source contacts; how wide should the drain-source gap be, and how much wider should the drain and source contacts be than the conducting channel defined by the mesa etch.

Figure 6.6 shows the pattern used for the ohmic contact layer. Ideally the drain-source gap should have been as short as posssible (say 2-3 microns) so that the series resistance of the channel was minimised, but in order to make aligning the gates easier, the drain source gap was initially set to 5 microns and later to 10 microns because of problems in aligning the gates when all were exposed at 80x. The metal of the drain and source contacts was 10 microns wider than the mesa that forms the channel for each column except for the 2 micron wide channel

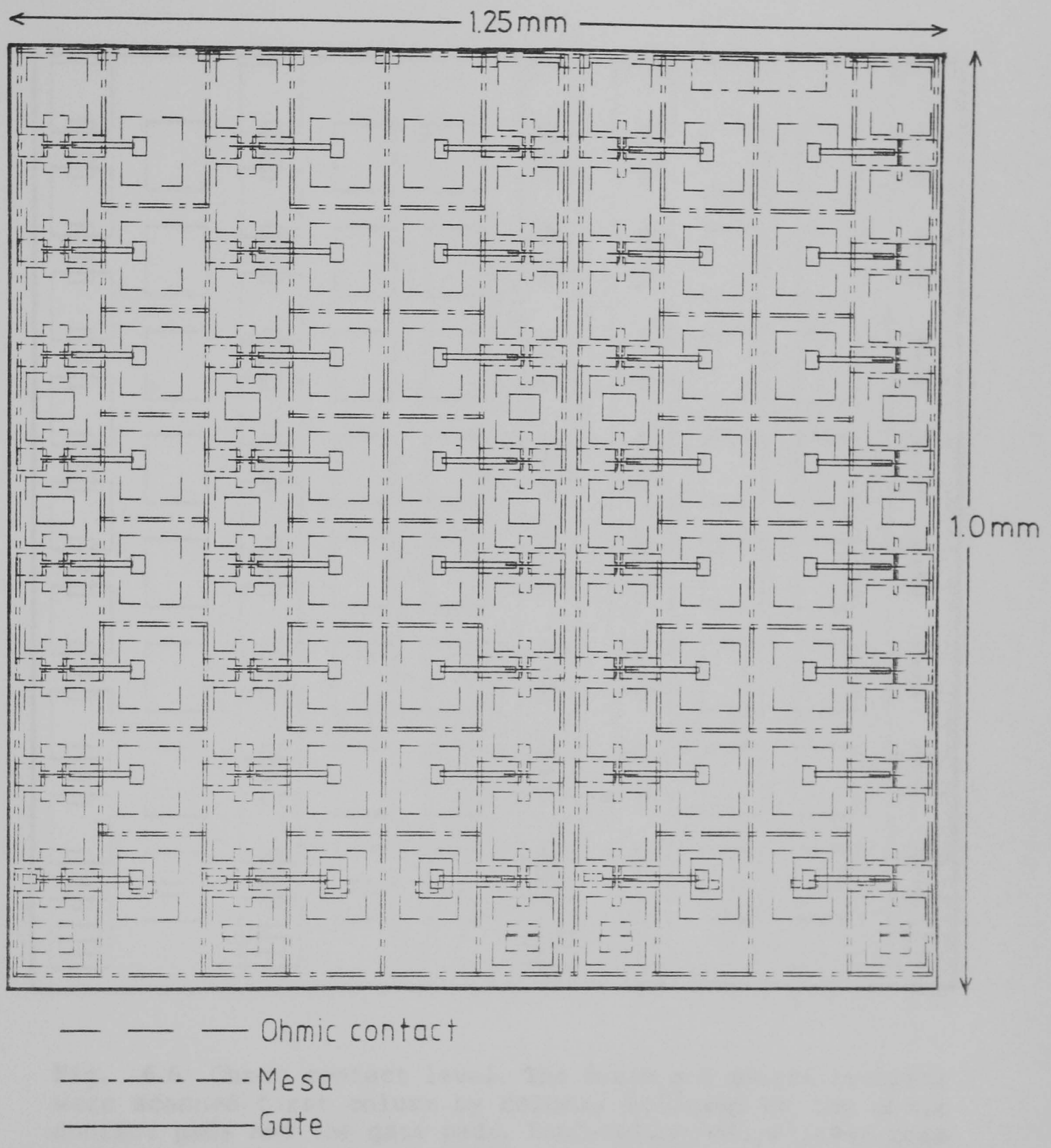


Fig. 6.5 Layout of the transistor pattern. The device widths change from column to column, and are, from left to right, 30,20,10,5, and 2 microns wide.

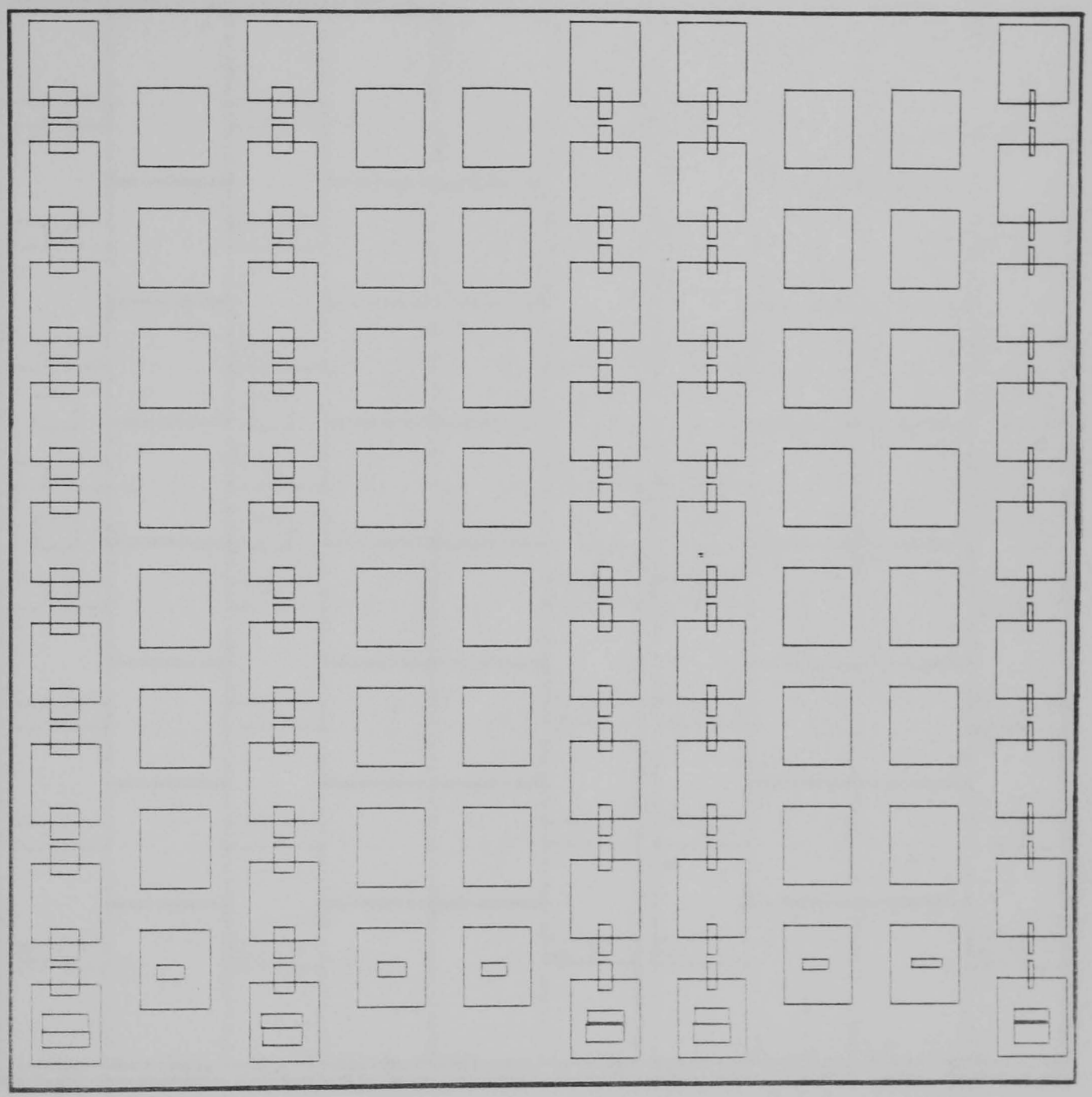


Fig. 6.6 Ohmic contact level. The drain and source contacts were scanned first column by column, followed by the ohmic contact pads and the gate pads. Each column was written from bottom to top to minimise hysteresis effects. The nested rectangles drawn within the bottom pad of each column were drawn immediately before the pad or drain contact above it, and prevented skewing of the beam at the top edge of the pad or contact. Gate pads are included at this level to save writing time later on.

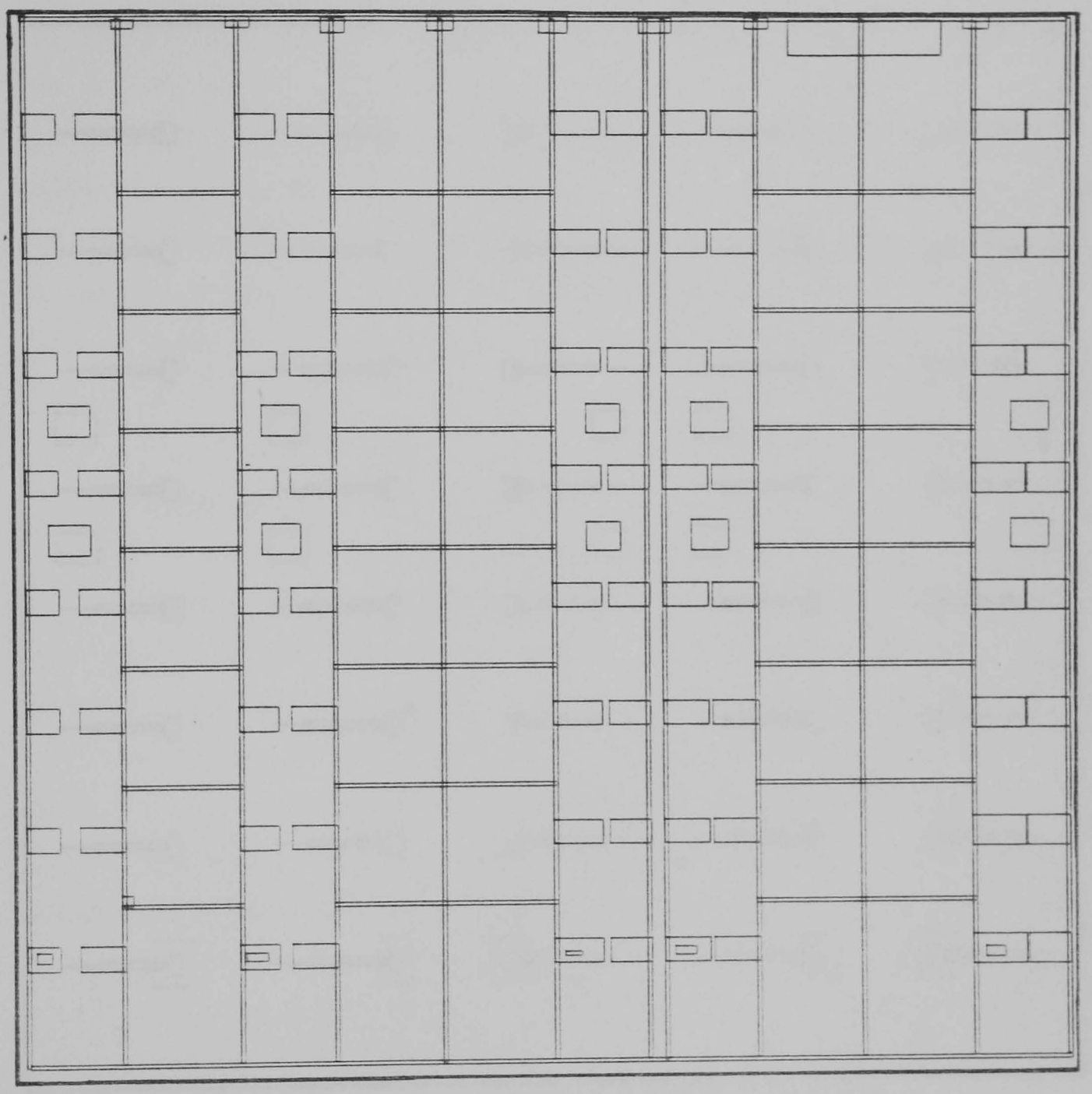


Fig. 6.7 Mesa level. The 6 micron wide channels isolate the pads from each other, and the conducting channels are defined by the pairs of rectangles that overlap the ohmic contacts. The windows exposed on the two middle rows of ohmic contacts enable the leakage current to be checked during the etch.

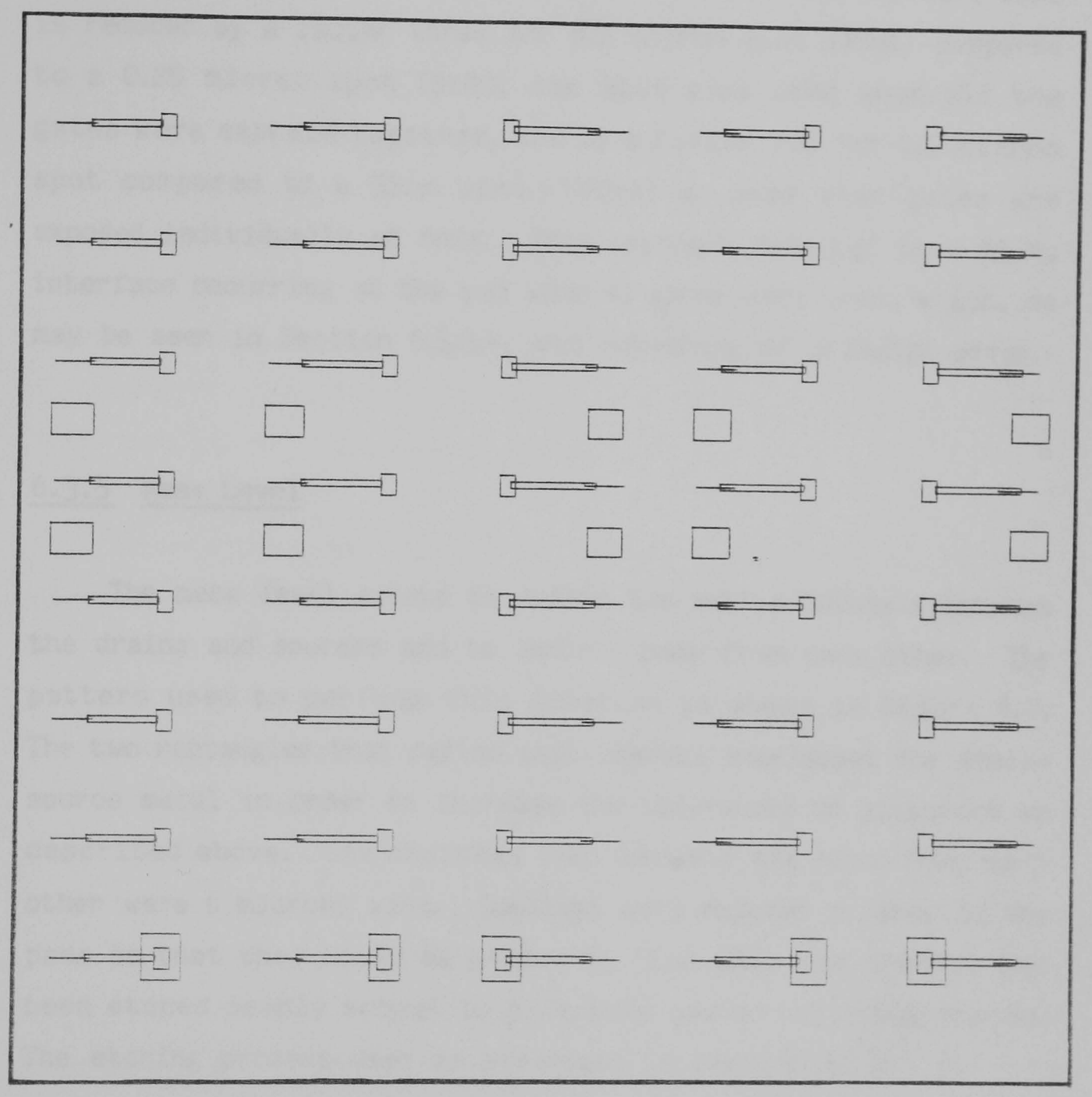


Fig. 6.8 Gate level (80x). The narrow gate stripe is connected to its pad by a wide connecting track. The two rows of rectangles form windows in the resist to enable probing of pads to measure saturation currents if a gate recess is being etched. Several versions of this pattern were written, with different pixel widths and exposure values for the gate stripe, in order to vary the linewidth on the chip.

where the overlap was 2 microns on each side. Conventional devices normally have the mesa slightly wider than the ohmic contact metallisation, but the alignment accuracies required for the narrow channels would have been very strict so this format was not adopted. The gate pads were exposed as part of the ohmic contact pattern in order to save time later on: the exposure time is reduced by a factor three for 0.5 micron spot (15nA) compared to a 0.25 micron spot (5nA), the spot size used when all the gates were exposed together, and by a factor 100 for 0.5 micron spot compared to a 32nm spot (150pA) as used when gates are exposed individually at 640x. This approach resulted in a Al/Au interface occurring at the pad when Al gates were used, which, as may be seen in Section 6.5.3.4, was something of a design error.

6.3.5 Mesa Level

The mesa level exists to define the active channels between the drains and sources and to isolate pads from each other. The pattern used to perform this function is shown in Figure 6.7. The two rectangles that define each channel overlapped the drain-source metal in order to increase the tolerances on alignment as described above. The channels that isolate the pads from each other were 6 microns wide. Openings were exposed on some of the pads so that they could be probed to find when the channel had been etched deeply enough to give good isolation between pads. The etching process used is described in Section 6.5.2.3.

6.3.6 Gate Level

The gates on the first devices were exposed at 80x magnification, all 40 being exposed at once. The gates themselves were lines, one,two, or four pixels wide, the width of the resulting metal line was varied by changing the degree of over-exposure of the line. The pattern used to pattern 40 gates is shown in Figure 6.8.

In order to be able to fabricate sub-0.1 micron gates, it was decided to expose the gates on later devices individually using a magnification of 640x, i.e. a frame size of 195 x 150 microns and a pixel size of 48 x 36nm. A computer program (TRANSF) was written to convert pattern data from one magnification to another. The program was used to put the ohmic contacts from a 30 micron wide device into the centre of a 640x frame. Using the ohmic contacts as a reference it was then a simple matter to use DESIGN to compose gate patterns with different widths of gate stripe, and alignment patterns registering on the gap itself and the right-hand edge of the drain-source metallisation. Figure 6.9 shows the high resolution gate pattern together with the other levels. Similar patterns were created for right-handed gates.

The exposure was done in two parts; one pattern exposed the gate and another exposed a stripe 10 microns wide that connects the gate to a 50 micron square pad exposed on the gate pad. In this way the area of contact between the gate metallisation and the probing pad was increased, thereby reducing any associated interfacial resistances. The stripe narrowed to 3 microns in order to pass between the drain and source contacts and get very close to or just onto the mesa where it connected with the gate. A window was exposed on the the ohmic contact pad above the gate being exposed. When all the gate patterns in a column were exposed, all the pads except the bottom one had an opening in the resist over them. These openings were used to facilitate probing the pads to measure the channel current if a gate recess was being etched (Section 6.5.3.3). The width of the stripe was varied by changing the width of the exposed line (in terms of pixels) and the exposure of the line. Thus a separate gate pattern was required for most gate lengths (Section 6.5.3.2). This was the reason for exposing the gate pattern in two parts, as it requires less memory to store 7 patterns containing single rectangles of different widths (the gates) plus one pattern containing 6 rectangles than to store 10 complete patterns

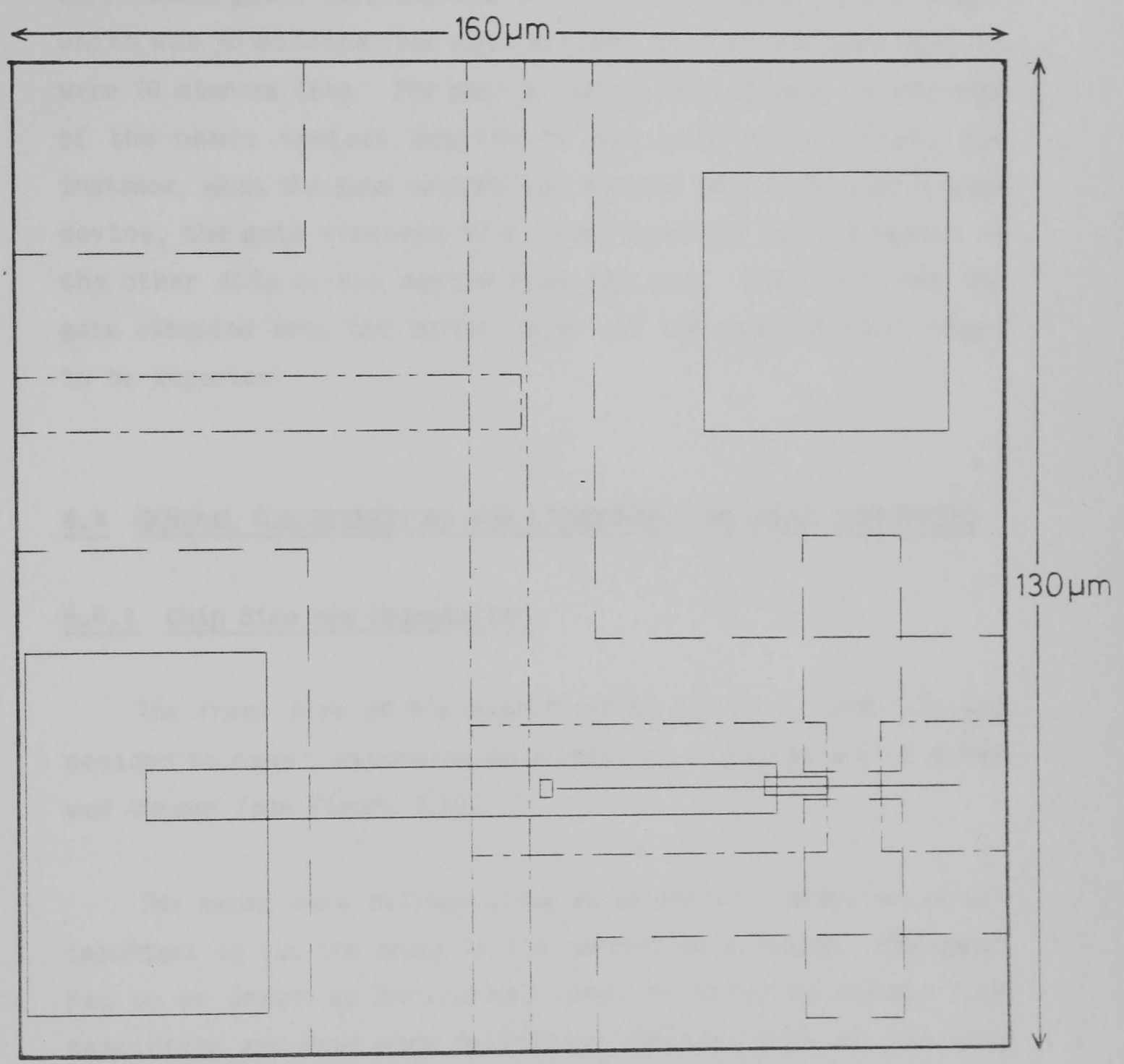


Fig. 6.9 Gate level (640x). Each gate and its connecting track is exposed individually using two patterns, one for the gate alone, and one for the connecting track. Several versions of the gate pattern were written with different pixel widths in order to enable different widths of gate stripe to be made. When all the gates in a column had been exposed and developed, the windows exposed on the pad above each gate enabled saturation currents to be measured during recess etching. This figure shows a right-handed gate, an equivelant set of patterns was written for left-handed gates.

containing 7 rectangles, 6 of which do not affect the width of the gate stripe. The small rectangle at the end of the gate stripe was exposed immediately before the gate and was to prevent distortion of the gate due to hysteresis (Section 2.4.3). All left-handed gates were exposed using the same gate stripe length which was 30 microns, the gate stripes on right-handed devices were 10 microns long. The gate patterns were aligned to the edge of the ohmic contact nearest to the gate pad so that, for instance, when the gate pattern was exposed onto a 20 micron wide device, the gate extended 10 microns into the etched region on the other side of the device from the pad. The fact that the gate extended onto the buffer layer off the mesa was not judged to be important.

6.4 GENERAL CONSIDERATIONS FOR LITHOGRAPHY ON SOLID SUBSTRATES

6.4.1 Chip Size and Orientation

The frame size at 80x magnification is 1.5 x 1.2mm. It was decided to make 9 exposures on a chip, so a chip size of 6 x 5mm was chosen (see Figure 6.10).

The mesas were defined using an anisotropic etch, so it was important to cut the chips in the correct orientation. The gates had to be drawn as horizontal lines in order to obtain high resolution and good edge definition (Section 2.3), so the mesa edges that the gates run up had to be vertical relative to the monitor of the SEM. The mesa edge had to be parallel to the $(01\bar{1})$ direction if the correct profile was to be obtained. For small etch steps, the etch profile for mask edges parallel to the (011) direction are undercut and are therefore difficult to climb with a lifted off metal line. The orientation of the VPE wafer was found by etching a scribed cross on a small piece of the wafer as described in Section 3.3.6.2.

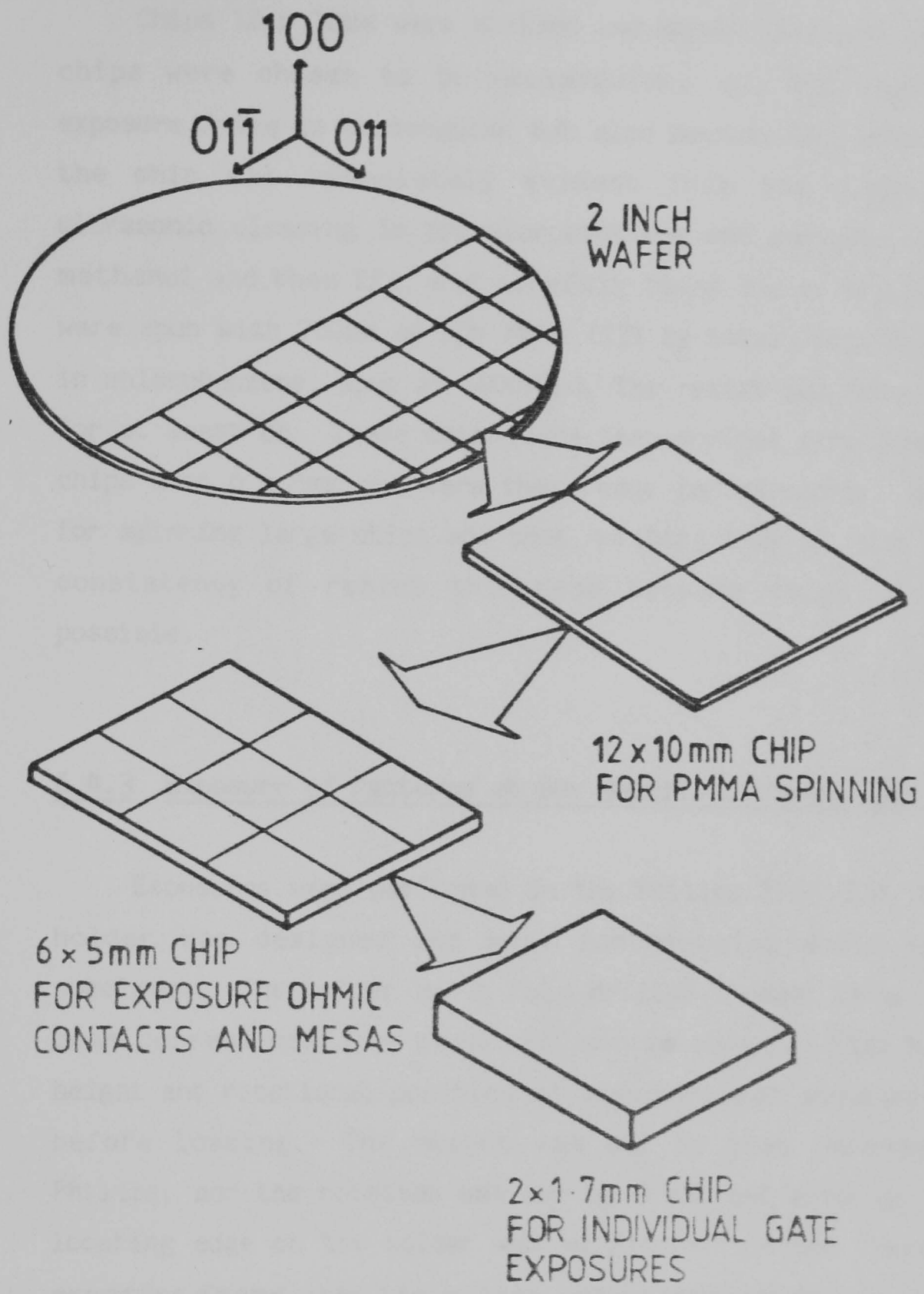


Fig. 6.10 The orientation of a wafer was first found by the etch test described in Section 3.3.6.2, it was then cut into 12 x 10 mm chips for spinning of the first layer of the resist to ensure consistancy between the four chips that this was then cut into. After patterning the ohmic contact and mesa levels, the chip is spun with resist and then separated into nine dies each containing one 80x exposure site for patterning of the gates.

6.4.2 Preparation of Chips for Exposure

Chips 12 x 10mm were scribed and broken from the wafer. The chips were chosen to be rectangular, not only because the exposure frame is rectangular but also because the orientation of the chip was immediately evident from its shape. After ultrasonic cleaning in trichloroethylene and acetone, rinsing in methanol and then IPA, and carefuly being blown dry, the chips were spun with 760nm of BDH PMMA (13% by total weight dissolved in chlorobenzene, spun at 6000rpm). The resist was baked at 180^{o}C for at least 8h. These chips were then scribed into four smaller chips each 6 x 5mm and were then ready for exposure. The reason for spinning large chips and then scribing them up, was to ensure consistency of resist thickness between chips as much as possible.

6.4.3 Exposure of Patterns at 80x Magnification on GaAs Chips

Exposures were performed in the Philips PSEM 500. A special holder was designed and made for exposing solid substrate specimens. The holder had a hole drilled through it so that the beam current could be measured, and is shown in Fig. 6.11. The height and rotational position of the specimen were adjustable before loading. The height was set to that recommended by Philips, and the rotation was set by trial and error so that the locating edge on the holder was within 1-2^{o} of the Y-axis of the exposing frame when the holder was loaded and the tilt was set at 0^{o}. The path of the centre of the exposure frame and the approximate size of the frame at various points during exposure of a chip is shown in Fig. 6.12.

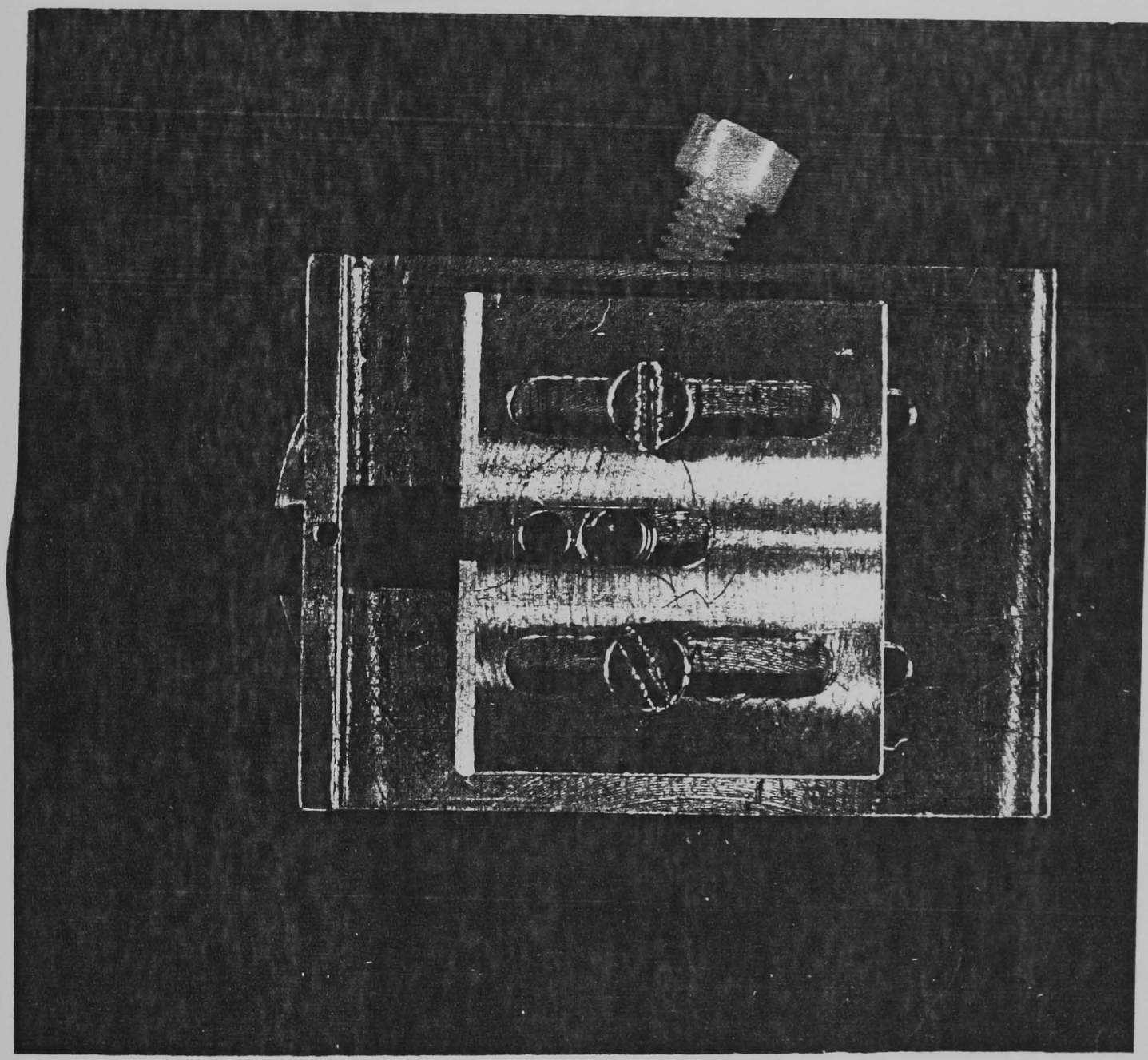


Fig. 6.11 The holder used during the exposure of solid substrate samples has a hole drilled through it to enable the beam current to be measured before and after exposure. A faraday cup is a recent modification and the aperture and its retaining circlip may seen to the right of the drilled hole.

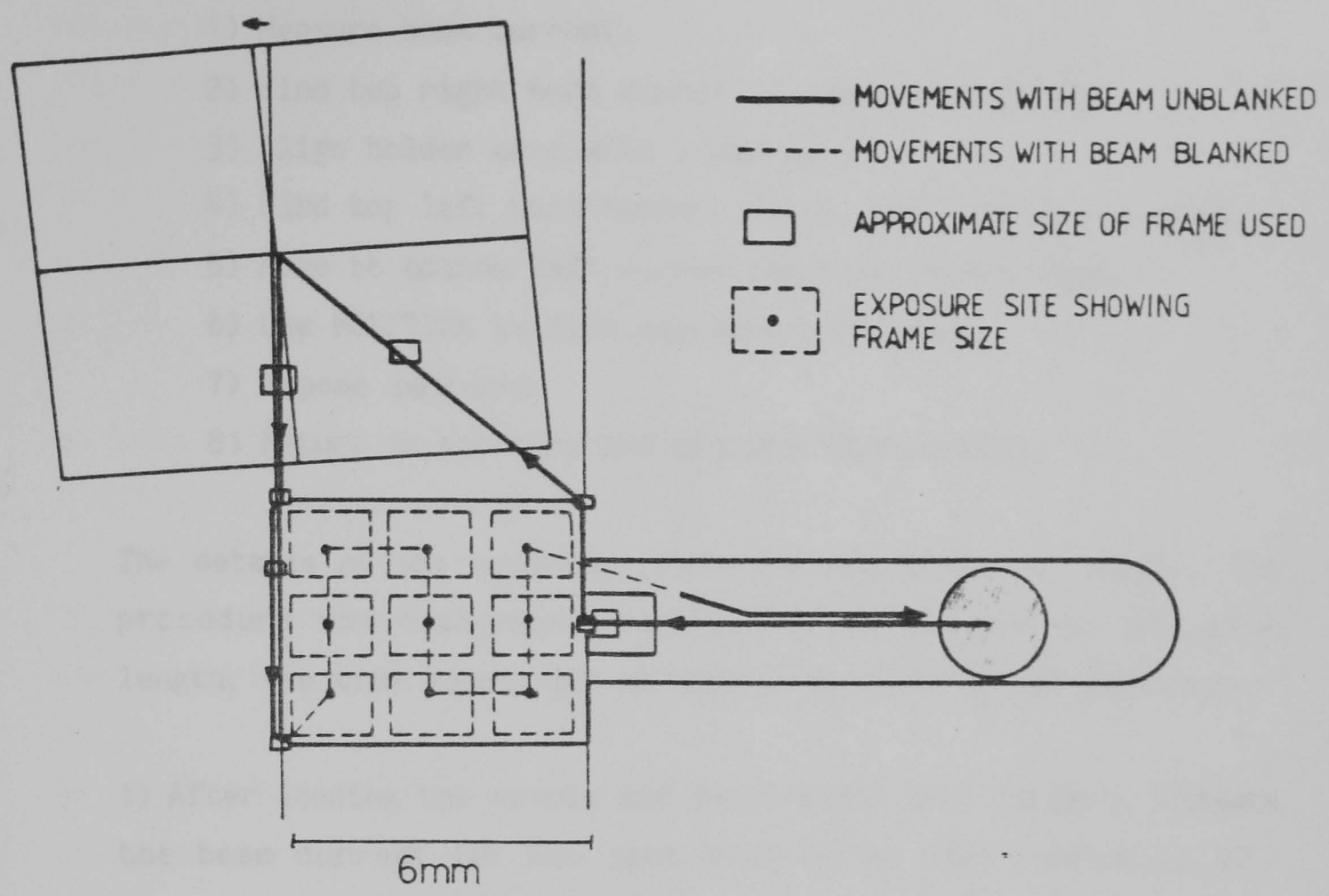


Fig. 6.12 The path taken by the centre of the frame and the approximate size of the frame is shown during exposure of a chip.

The main features of the path were;

1) Measure beam current.
2) Find top right hand corner of sample and focus.
3) Align holder axes with frame axes.
4) Find top left hand corner, focus, and note coordinates.
5) Move to bottom left corner and note coordinates.
6) Use POSITION to find exposing positions.
7) Expose patterns.
8) Return to aperture and measure beam current.

The details of the exposing procedure are described below. The procedure was designed to minimise the following; the path length, the time taken, and undesired exposure on the specimen.

1) After loading the sample and setting the tilt to zero, measure the beam current for the spot size to be used (normally 0.5 micron or 0.25 micron at 50kV).

2) Using the 1/16 frame scan at 20x magnification, move towards the specimen. As soon as the specimen appears in the left edge of the frame, increase the magnification so that the frame size decreases and continue moving towards the edge. Increase the magnification each time the edge of the specimen appears in the frame until the magnification reaches 640x. Move to the top left corner, keeping the sample in view at the edge of the frame. The frame size at this point is 195 x 150 microns, so the exposure along the edge of the sample does not encroach onto the device area. Focus at the corner. The reason for focussing before aligning the axes is that the act of focussing rotates the image.

3) Move diagonally by 2mm up and 6mm left to the locating step in the holder. Reduce the magnification to 40x, with full frame and cross hair cursor. Use the goniometer rotation to align the locating step with the vertical line of the cursor. The specimen and frame axes will now be approximately aligned.

4) Reduce to 1/16 frame at 40x magnification and move down the step in the holder to the top left corner of the specimen, increasing the magnification to 640x as the corner appears in the frame. Refocus, if necessary. Move the corner to the centre of the screen, and note the goniometer coordinates.

5) Move to the bottom left corner of the sample, keeping the edge of the specimen in view.

6) Move corner to centre of screen and note the coordinates. Use POSITION to find coordinates of exposing positions.

7) Expose the required pattern at each exposing position in turn, aligning to a previously laid down pattern if necessary.

8) After exposing the last pattern, move to the beam current measurement aperture and check the beam current. Normally the current will have changed by less than 5% from the start of the exposure procedure to the end.

6.4.4 Development

The sensistivity of resists depends, among other things, on the developer mixture, the developer temperature, and the developing time, as described in Section 1.6.

As development proceeds it is found that the value of the sensitivity of a layer of PMMA decreases quickly with time, until, after a time referred to here the critical development time, the sensitivity levels off. After the critical development time, the contrast will start to decrease because of the slow dissolution of unexposed resist. Weak developers will saturate more strongly than strong developers, and the contrast will be more insensitive to overdevelopment.

6.4.4.1 Developer Composition

Previous work in this Department had measured sensitivities and contrasts for various developer mixtures [6.35]. From these results it was decided to use a 1:1 mixture of IPA and MIBK, which gives about a factor 3 improvement in sensitivity over 3:1, without severe loss of contrast. Since pattern features were relatively coarse, some loss of contrast could be sustained while still achieving good resist profiles for liftoff.

6.4.4.2 Development Time

An experiment was performed for each resist (including changes in resist thickness), and developer (including changes in temperature) to find the critical development time. A chip was exposed with the pattern shown in Fig 6.13. The exposure is constant for each group of lines which have widths of 1,2,3,4,5,8,11,14,17,20 pixels while the exposures are increased from group to group by a constant factor. The range of exposures was chosen to bracket the critical (minimum) exposure for complete clearing of an exposed line during development. The exposures for the rectangles had to be optimised during a few passes through the experimental procedure.

The chip on which the pattern had been exposed was scribed and broken into pieces each containing one complete exposed pattern (the exposed patterns were faintly visible before development). The pieces were then developed for a range of times and examined in an optical microscope. It was easy to find the development time above which no gain in sensitivity occurred, simply by looking to see what the critical exposure for a given linewidth was for each development time in turn, and finding the development time above which little or no decrease in critical exposure occurred (Figure 6.14).

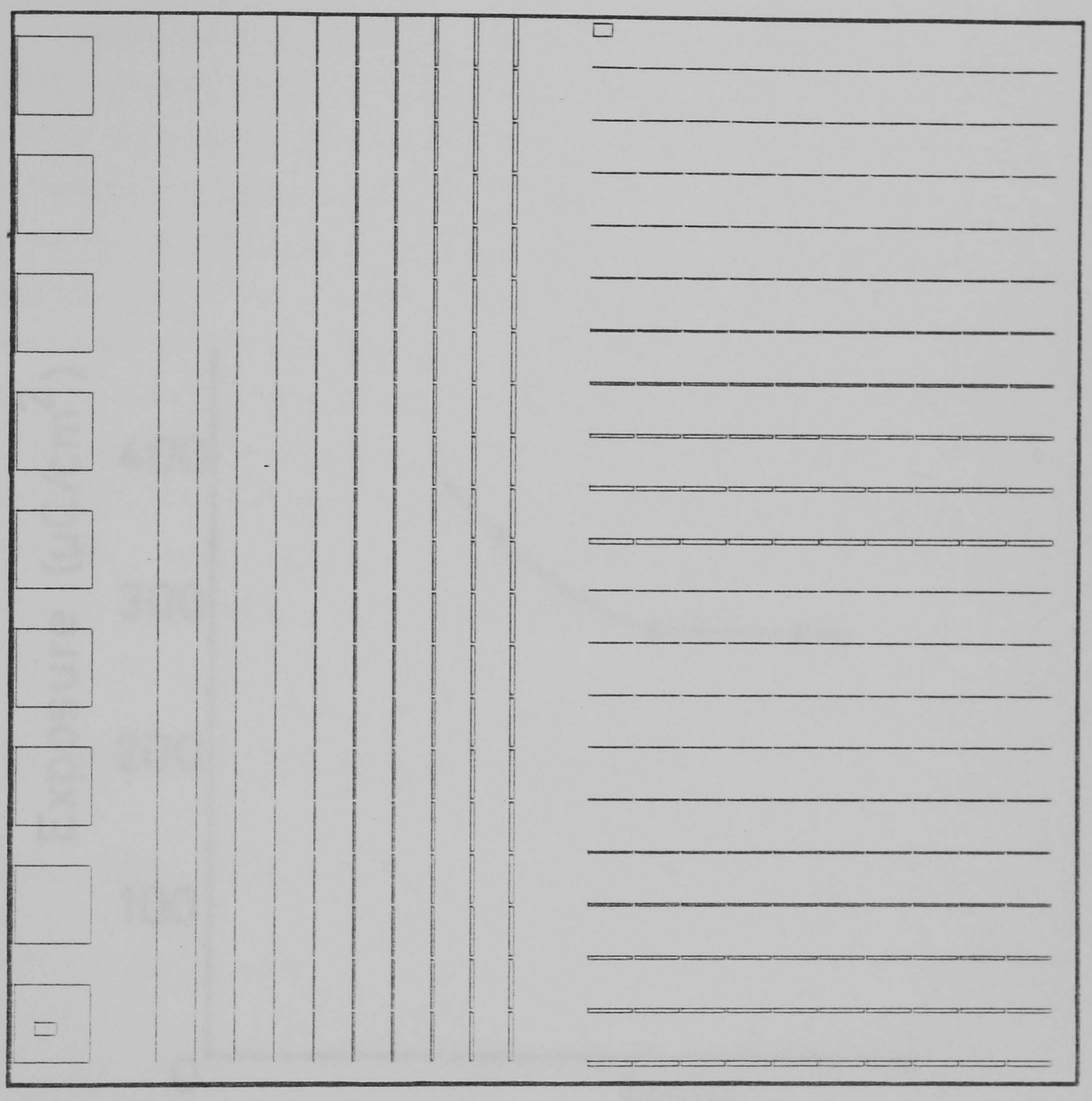


Fig. 6.13 Pattern used to find critical exposures for a range of pixel widths. Vertical lines are exposed in rows of increasing exposure and similarly for horizontal lines.

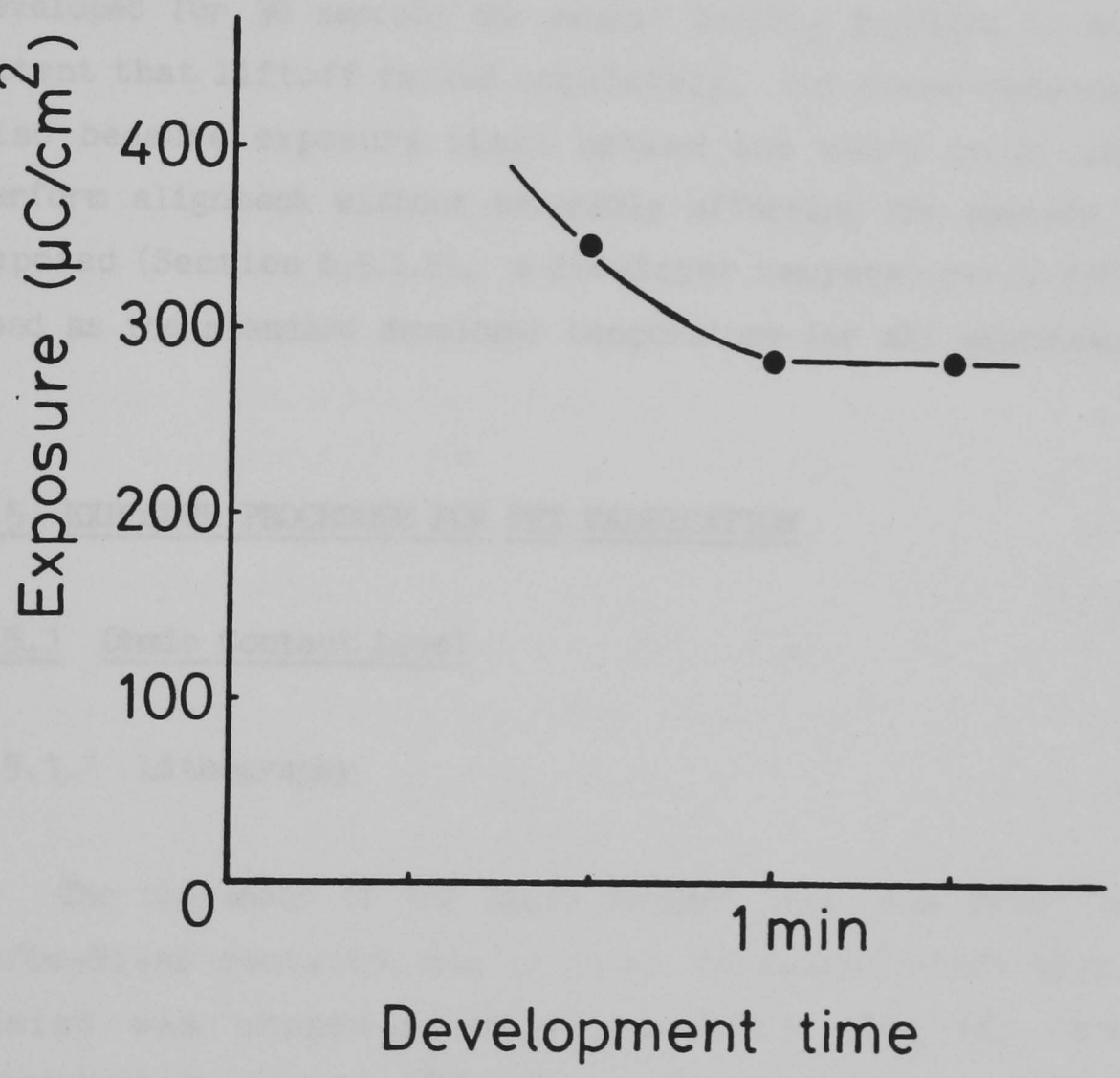


Fig. 6.14 Plot of critical exposure for 100 micron blocks versus development time. Resist was 1 micron of PMMA (mol. wt. 185,000), developer was 1:1 IPA:MiBK at 23.0°C. After sufficient development the critical exposure is independent of development time. For consistant results, a development time should be used that is slightly longer than the critical development time.

The development time must not be too short (say less than 15s) or else repeatability problems may occurr.

6.4.4.3 Developer Temperature

Experiments were done with developer temperatures of up to 38°C. It was found that the sensitivity of 600nm of BDH PMMA resist when developed in a 3:2 mixture of IPA and MIBK at 38° was a factor of three lower than when 1:1 at 23.0°C was used. However the critical development time was found to be around 10 seconds which is too short to be useful. When samples were developed for 30 seconds the resist profile degraded to such an extent that liftoff failed completely. For these reasons, and also because exposure times became too short to be able to perform alignment without adversely affecting the pattern to be exposed (Section 6.5.2.2), a developer temperature of 23°C was used as the standard developer temperature for all exposures.

6.5 EXPOSURE PROCEDURE FOR FET FABRICATION

6.5.1 Ohmic Contact Level

6.5.1.1 Lithography

The thickness of the ohmic contact layer was 250nm for the Au/Ge-Ni-Au contacts, and in order to make liftoff easy, the resist was chosen to be 760nm thick (13% BDH PMMA in chlorobenzene spun at 6000rpm). First, the correct development time was found by the procedure described in Sect 6.4.4.4. It was found to be 25s in 1:1 at 23.0°C. Next, the correct exposures for each part of the pattern was found. The critical exposures for isolated lines of any linewidth were already known from the experiment to find the critical development time.

Fine tuning of the exposure was done with an early pattern that contained the drain and source contacts with their pads but not the gate pads. Each column of pads was split into two groups of four. The ten groups of four pads were given a range of exposures that bracketted the critical exposure. The exposures for pairs of drain and source contacts in each column were similarily bracketed. The test pattern was exposed on another piece of scrap GaAs. The sample was developed, metallised with 100nm of aluminium (for ease of evaporation), and lifted off by soaking in acetone for 5 minutes, followed by shooting. The lifted off pattern was examined in an optical microscope. The exposures that were chosen as being "correct" were about 20% higher than the minimum required for clean liftoff with smooth edges at the critical parts of the pattern, .i.e. the drain and source contact edges. The slight overexposure guaranteed good liftoff even if the beam current drifted during exposure. The exposures used were 350 microcoulombs/cm^2 for the pads and wider drain and source contacts, and 400 microcoulombs/cm^2 for the column with 2 micron wide channels (6 micron wide drain and source contacts). The new exposures were entered using DESIGN. Once the pattern could be reliably lifted off on scrap material, it was exposed on chips from the VPE wafer with the same resist layer on them, using the exposure procedure described in Section 6.4.3. After development, chips were cleaned of surface oxide by dipping them into concentrated HCl for 30s followed by rinsing in deionised water. The ohmic contact metallisation was then evaporated and lifted off by soaking in acetone and shooting. The composition and alloying conditions of the contacts are described in the following section.

6.5.1.2 Contact Composition and Deposition

The ohmic contacts used in transistors were a composite of Au/Ge(12:88)-Ni-Au evaporated in the same vaccuum cycle to thicknesses of 100nm, 20nm and 100nm respectively. Early devices were alloyed at 330°C for 1 minute, later ones were alloyed at

400°C for 1 minute. It was found that devices fabricated with the higher alloying temperature had better characteristics than ones with low temperature contacts, (Section 7.2).

6.5.2 Mesa Level

6.5.2.1 Finding the Correct Exposure Values

After the ohmic contacts had been formed, the pads were isolated from each other by etching a channel between them, which also forms the active mesas. The pattern used to do this was shown in Fig. 6.7. A thicker layer of resist was used for this level just to be sure that the GaAs epi-layer received as much protection from the etch as possible. The resist was again BDH PMMA but was 1.1 microns thick (15% spun at 6000rpm). A development time of 1 minute in 1:1 at 23.0°C was used, and the correct exposures for the various features of the pattern were found in similar way as for the ohmic contact pattern (Section 6.5.1.1). The standard exposure procedure was used (Section 6.4.3), and each exposure was aligned to one of the 9 ohmic contact patterns on the chip by the alignment procedures described in the following sections.

6.5.2.2 Alignment

The two alignment methods that were employed use features of the transistor pattern for registration, but the techniques could easily be extended to suit any pattern.

As was described in Section 2.3, the output from the scan generator which drives the beam on the sample also drives the monitor screens. The intensity of the spot on the screen as it is moved by the scan generator from point to point is modulated by the output of the detector in use. Thus as a pattern is scanned over a sample a picture of the sample appears on the

screen as the sample is being viewed through a mask with holes in it corresponding to the pattern scanned out by the beam. It is the facility of being able to scan particular areas of a sample and being able to see the output from a detector during scanning that is used during alignment.

The particular areas of the sample that are scanned over during alignment are called alignment or registration marks. These may be specially written areas not part of any devices or they may be parts of devices themselves, as in this case. The criterion for good alignment marks is that the detected signal from them should give good contrast against the background and they should have good definition. The Au/Ge-Ni-Au marks used here were very good in both these respects.

The two stages of alignment were; initial alignment done using a large area scanned pattern giving alignment to within 3-4 microns, followed by a more precise method using a small area scanned alignment pattern that gave alignment to within 1 micron.

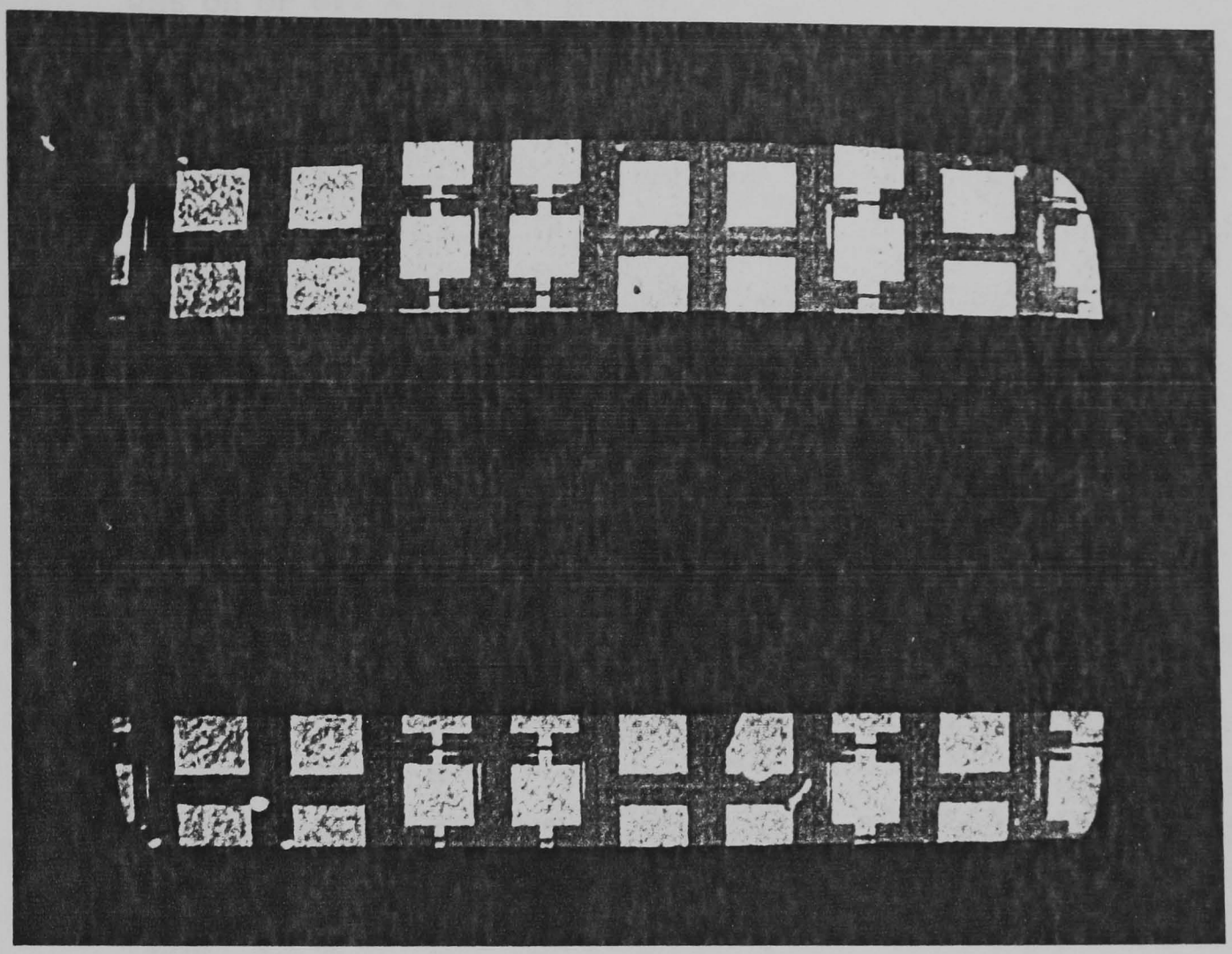
6.5.2.2.1 Coarse alignment

After positioning the frame roughly over the exposing site using POSITON, an alignment pattern was scanned once using a dwell time of 2.0 microseconds at each pixel (i.e. as fast as the DACs could be driven). The pattern consisted of two horizontal 20 pixel pitch gratings overlaid by 10 partial outlines of rectangles in two rows. These outlines were in the same positions as the edges of the pads in the second top and second bottom rows of the ohmic contacts. The output on the monitor screen given for moderate misalignment is shown in Figure 6.15 (upper). The object was to place each of the second top and second bottom probing pads such that its edges coincided with the scanned rectangle. The visibilty of this alignment test was not very good, one micron on the sample corresponded to about 0.2mm on the screen, roughly half the spot diameter. The intensity of

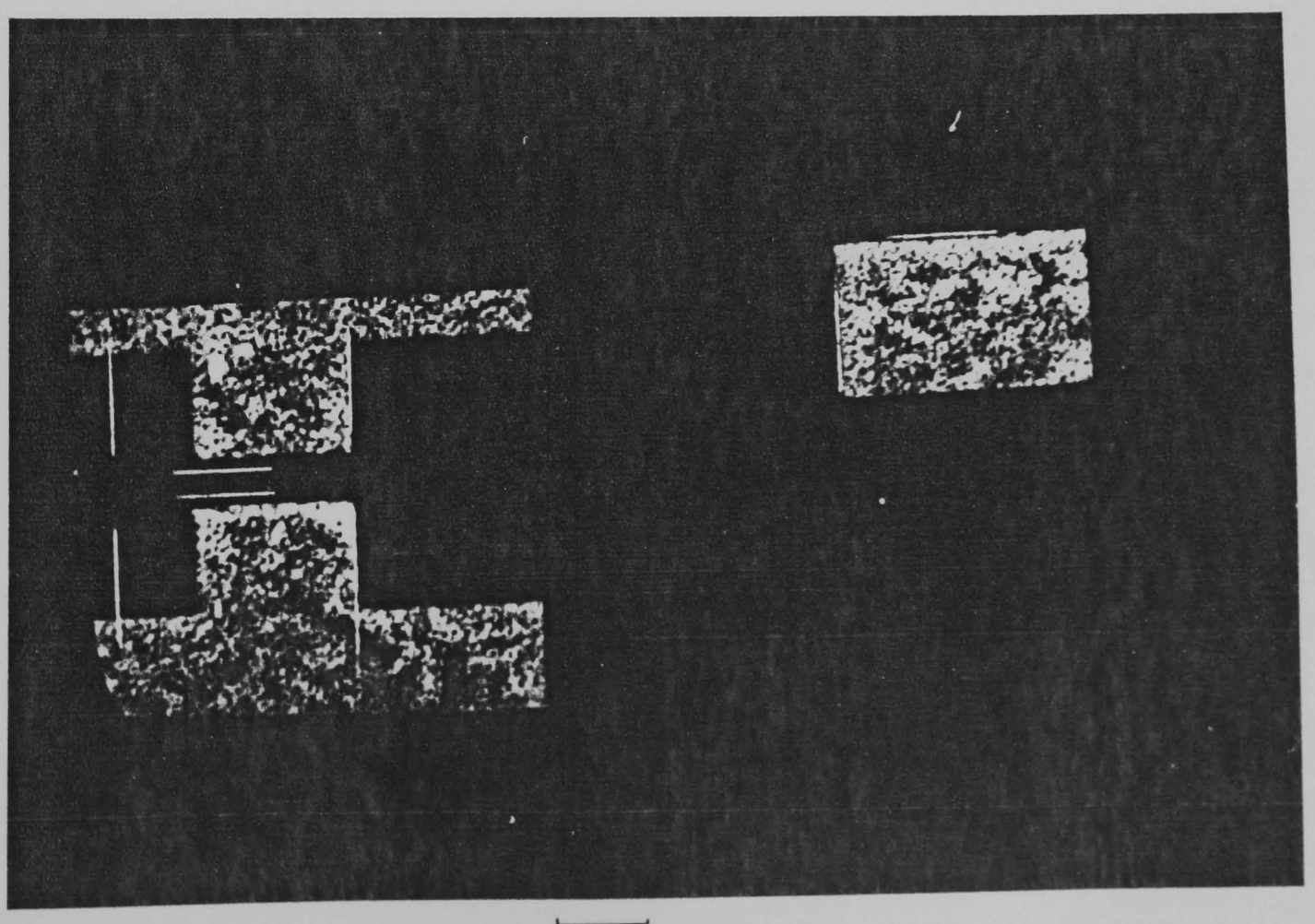
Fig. 6.15 Output from monitor screen during coarse alignment.

(Upper) 80x magnification alignment. Alignment marks on specimen are the ohmic contact probing pads. The scanned marks are the partial outlines of rectangles, which when the exposing frame is aligned with the sample frame, coincide with the edges of the appropriate pad. In the Figure, the X-magnification is too large, the Y-magnification is too low, and the frames are misaligned in the Y-direction by about 15 microns and by about 0.5^{o} in rotation.

(Lower) 640x magnification alignment of left-handed gate. Alignment marks are the right hand edge of the drain and source contacts, the two visible edges of the gate pad, and the drain source gap. Figure shows frames in good alignment. Scanned alignment pattern was originally designed for wider device with 5 micron gap rather than the 10 micron gap of this pattern.



100μm



20μm

the scanned line did not change very much according to whether it was on or off a pad, so the test could not be changed to one of testing intensity, as was done on thin substrates (Section 5.2). Alignment to within 3-4 microns could usually be achieved using 4-5 scans for the first exposure on a sample, when fairly large adjustments to the rotation and X and Y magnifications had to be made, and 2 scans for the others. Alignment to within 1 micron could be achieved using 7-8 scans, but the extra exposure given to the sample affected the quality of the pattern being exposed by causing increased linewidths and degraded resist profiles which sometimes prevented clean liftoff.

The 'squares' method is good for general location work because the scan covers a large area of the sample, but for the same reason it carries the danger of inadvertantly changing the exposure under the scanned area enough to alter or even destroy the resist image. It was found that if a 0.5 micron spot was used, the area under the alignment scans received enough exposure to prevent good liftoff. Changing to a spot size of 0.25 microns extended the 'safe' number of scans from 2 to about 7

6.5.2.2.2 Fine Alignment

It was originally thought that the squares method would provide good enough alignment to overlay patterns to an accuracy of 1 micron. However the limited time allowed for scanning reduced its effectiveness. Another more accurate method was developed. There were no alignment marks as such on the ohmic contact pattern, so a method was devised that used a feature of the ohmic contacts themselves. The pattern shown in Fig. 6.16 shows the scanned alignment pattern superimposed on the ohmic contact pattern. The pattern consisted of four crosses. The width of the vertical bar of each cross was the same as that of the corresponding ohmic contact when it was exposed, and the height of the horizontal bar was the same as that of the gap. When the frame was properly aligned, the crosses scanned the

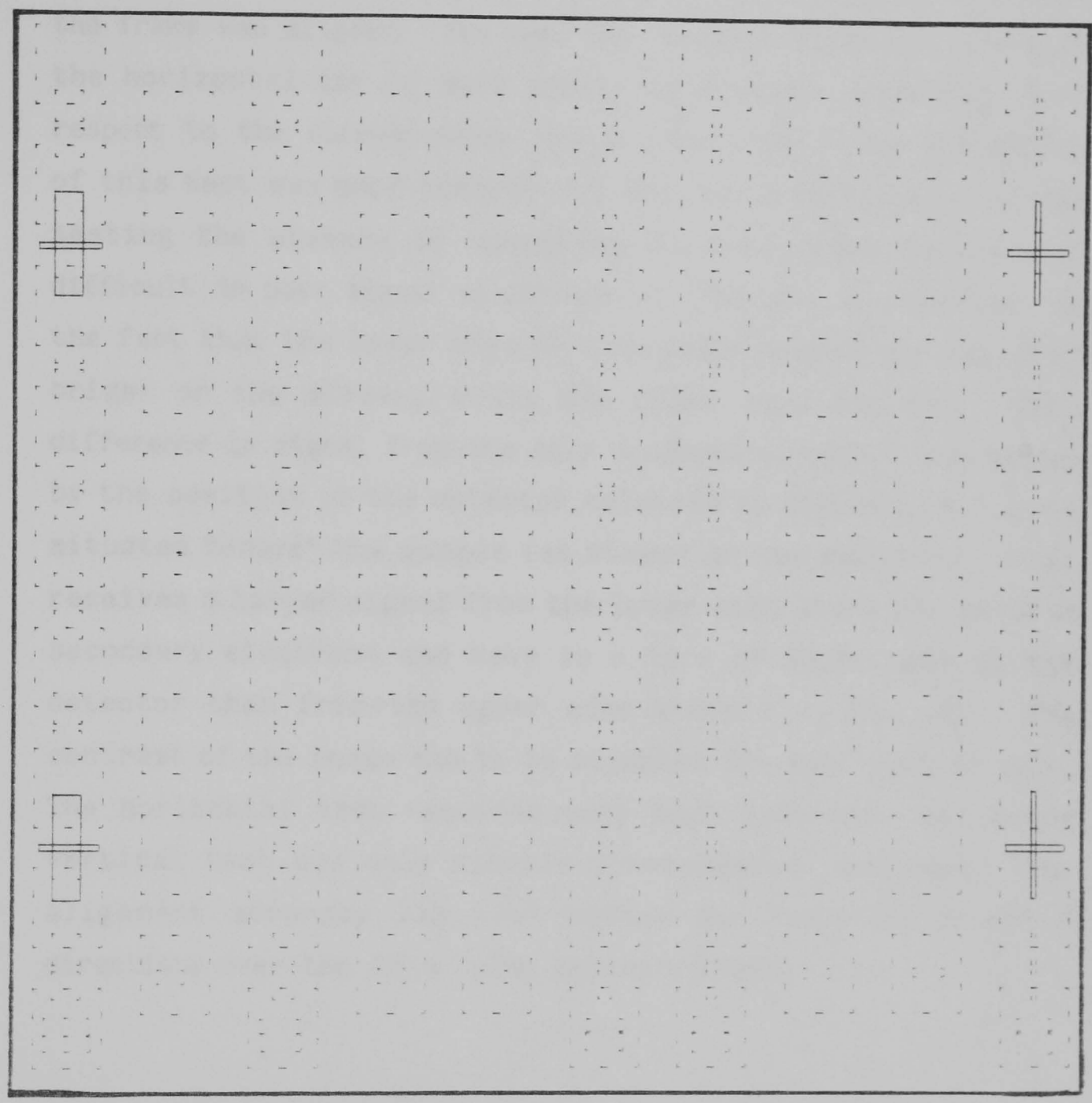


Fig. 6.16 Scanned alignment pattern superimposed on ohmic contact pattern. When frames are aligned, the vertical bars of the crosses coincide with the ohmic contacts, and the horizontal bars lie excactly in the drain-source gap.

second top and second bottom drain source gaps of the two outer most columns. The alignment tests in the X and Y directions were different. Alignment in X direction was achieved when the vertical bar of each cross lay exactly centrally with respect to its corresponding mark. The visibility of this test was good. A bright line down each side of the scanned bars could be seen when the frame was aligned. The test for vertical alignment was that the horizontal bar of each cross lay exactly centrally with respect to the corresponding drain- source gap. The visibility of this test was poor because the test was a null test, one was testing the absence of something on the screen (which was difficult in poor signal conditions). The test was confused by the fact that the lower edge of each drain source gap appeared bright on the screen, while the upper edge did not. This difference in signal from the same topological object was caused by the position of the detector relative to the sample. It is situated "above" the sample (as viewed in the monitor) so it receives a larger signal from the lower edge where the emitted secondary electrons can move in a line of sight path to the detector than from the upper edge where they can not. The contrast of the image had to be adjusted for each test in turn. The horizontal test required very high contrast, while the vertical test was only visible for moderate contrast. The alignment accuracy was +/-1 micron for both the X and Y directions over the 1.5 x 1.2mm exposure frame.

6.5.2.3 Etching Procedure

The samples were etched for periods of 5 seconds in a 1:8:200 mixture of H_2SO_4 :H_2O_2:H_2O (Section 3.3.2), rinsed in water, and blown dry. After each period of etching,the leakage current between adjacent pads was measured. When the current fell to 0.05mA at 25V from its initial value of 4.5mA at 1V applied voltage, etching was stopped and the devices were taken as being isolated. In a typical case, three periods of etching would be required, and it was noted that the current only dropped

by 0.5mA during the first period of etching, but but by 2.5mA during the second. A 'start up' time of this nature was common among slow etches and was also encountered during etching of gate recesses (Section 6.5.3.3).

6.5.3 Gate Level

After the gate level resist had been spun, chips were scribed and broken into 9 dies each containing one complete transistor pattern.

6.5.3.1 Large Frame Exposures

The gates on the first complete devices were exposed using the same frame size as was used for the ohmic contacts and isolation, all the gates being exposed together (Figure 6.8, Section 6.3.6). The resist used was 760nm of BDH PMMA, the same as for the ohmic contact layer. The pattern was aligned in the same way as for the mesa level. Development was for 25 seconds in 1:1 at 23.0°C. The surface oxide was cleaned off as before with a 30s dip in HCl prior to the deposition and liftoff of 100nm of Al plus 20nm of Au/Pd. The function of the Au/Pd was to enable the gates to be seen in the SEM so that accurate linewidth measurements could be made. Transistors with gate lengths from 1.1-1.8 microns were fabricated using this pattern. The electrical characteristics are described in Section 7.2.

6.5.3.2 Small Frame Exposures

The width of the narrowest gate stripe that could be fabricated using 80x magnification for exposure was 0.7 microns. Narrower gates had to be patterned individually using the pattern described in Section 6.3.6 and shown in Figure 6.9.

A two layer resist was used for narrow gates, 100nm of Aldrich PMMA on spun on top of 170nm of BDH PMMA, baked for an hour after deposition of the first layer, and for at least eight hours after deposition of the second, both at 180°C [6.36]. The critical development time and exposures were found by the methods described in Section 6.4.4. The widths of lifted off metal lines drawn with different widths (in pixels) and exposures were measured in the SEM. The results are shown in Fig. 6.17. From this, combinations of linewidth in pixels and exposure were chosen to give a standard series of metal linewidths. These are also shown in Fig. 6.17.

The exposure procedure for a die was similar to that of a chip. The only difference being that POSITION was not used to find the exposing positions. The step required between exposing sites was found by trial and error. The distortion in exposures done at 80x was such that the vertical distance between drain-source gaps was 155 microns at the bottom of a die, but only 135 microns at the top.

Devices with gate lengths from 700nm down to 75nm were made (Figure 6.18). The metallisation was 60nm of Al followed by 17nm Au/Pd. The electrical charactaristics of the devices are described in Chapter 7. Narrower gates could not be be made, probably because of the size of the pixels used.

6.5.3.3 Recessed Gates

The planar devices would not pinch off because the active layer was too thick for its doping (6.3.1,7.2). A technique that is often employed in FET manufacture is to etch a recess into the active layer through the opening in the resist that will be used to lift off the gate. Usually the recess is etched through a highly doped contact layer that carries the current to the edge of the recess, thus lowering the parastitic resistance of the channel and reducing noise in the device. In this work the

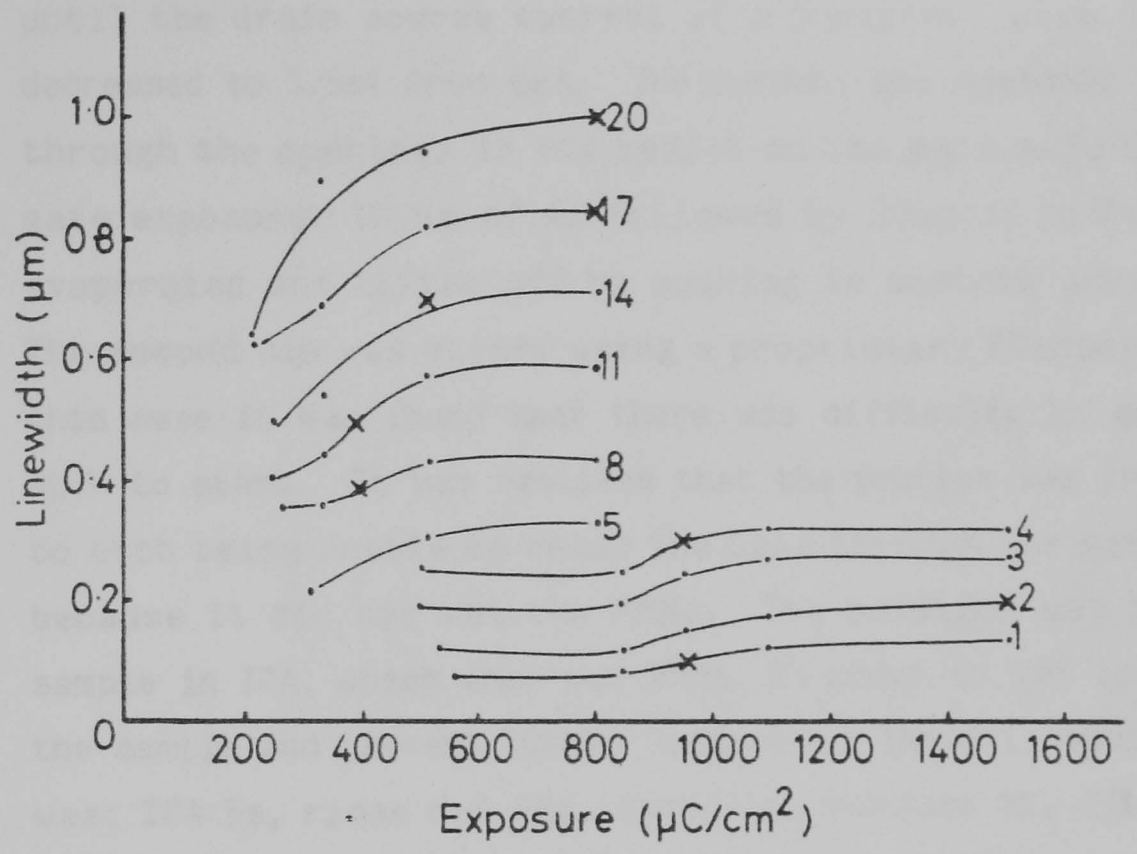


Fig. 6.17 Plot of width of lifted off lines versus exposure for differnt pixel widths exposed at 640x magnification using 32nm spot. The crosses show the exposures and pixel widths used to define gate stripes with widths of 0.1,0.2,0.3,0.4,0.5,0.6,0.85, and 1.0 microns. The apparent abrupt change in the critical dose (the left end of each line) between 4 and 5 pixel linewidths is not fully understood, but may be due to proximity effect, since the wider lines at critical doses lay near the centre of the pattern, while the narrow lines lay near or at the edge.

recess was used principally to lower the threshold of devices and also to enable them to be pinched off. Two dies had recesses etched before evaporation of the gate metallisation. The first die was etched in 1:8:1000 H_2SO_4:H_2O_2:H_2O for periods of 1 minute until the drain source current of a 5 micron wide device had decreased to 1.5mA from 6mA. The current was measured by probing through the openings in the resist on the pads made during the gate exposure. 100nm of Al followed by 20nm of Au/Pd was then evaporated and lifted off by soaking in acetone and shooting. The second die was etched using a proprietary Plessey etch. In this case it was found that there was difficulty in getting the etch to start. It was realised that the problem was probably due to etch being unable to reach the GaAs through the gate opening because it did not wet the PMMA. The solution was to dip the sample in IPA, which does wet PMMA, in order to wet the whole of the sample and prevent bubble formation. The full etch procedure was; IPA 5s, rinse H_2O 10s, deoxidise surface NH_3 (3%) 5s, etch recess 2-5s, rinse H_2O 10s, blow dry, measure drain-source current. The procedure was once more stopped when the current on 5 micron wide devices had dropped from 6mA to 1.5mA. The gate metallisation used in this case was 40nm of Ti followed by 80nm of Al, evaporated during the same vaccuum cycle.

Some undercutting of the resist was found to have occurred when these devices were examined in SEM. The gate stripe widths were on average 0.05 microns larger than expected from the pattern data that was used to expose them.

The characteristics of the recessed devices are described in Section 7.3.

6.5.3.4 'Purple Plague'

When gold and aluminium are heated in contact with each other they interdiffuse and the aluminium areas are eaten up by the "purple plague" [6.37]. A design fault when aluminium gates

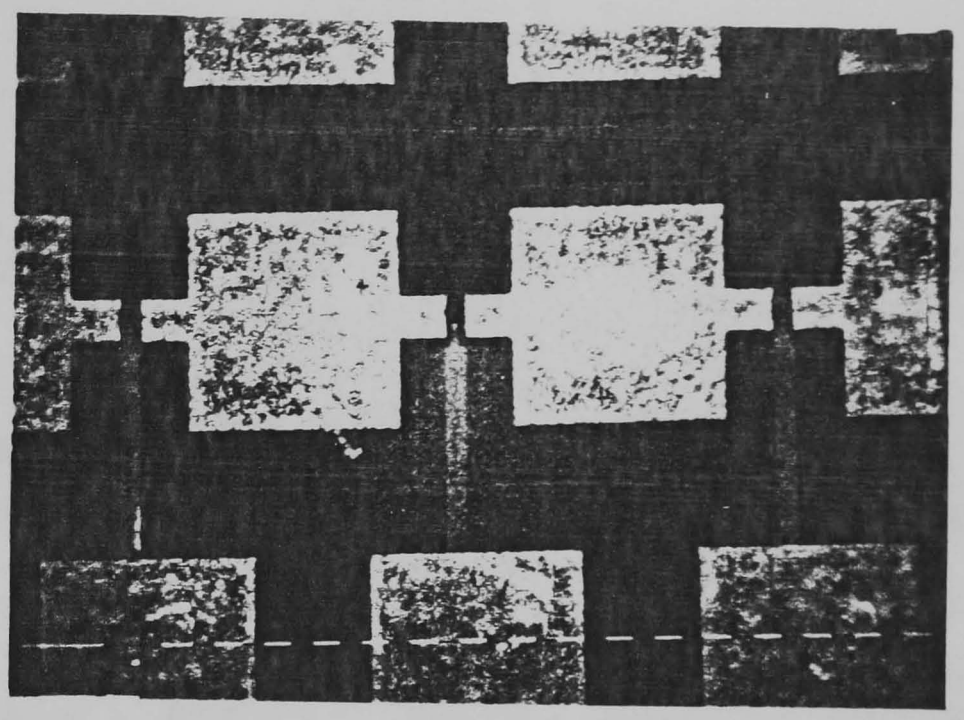
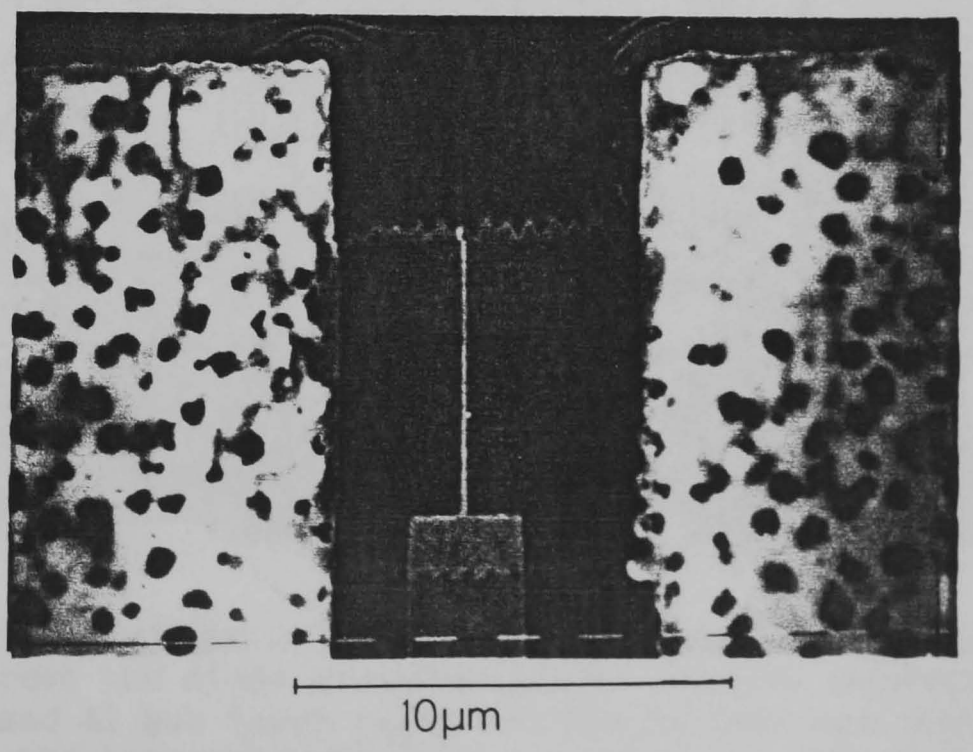


Fig. 6.18 Transistor with 0.075 micron gate.

(Upper) General view showing the probing pads and isolation channel. (10 micron markers)

(Lower) Magnified view showing connecting track overlapping onto mesa. Note edge roughness of mesa due to use of spot smaller than the pixel size and the working of the scan generator. Also there is some unetched GaAs close the the corners of the ohmic contacts where the resist did not develop through because of local thickening in the vicinity of the corners.



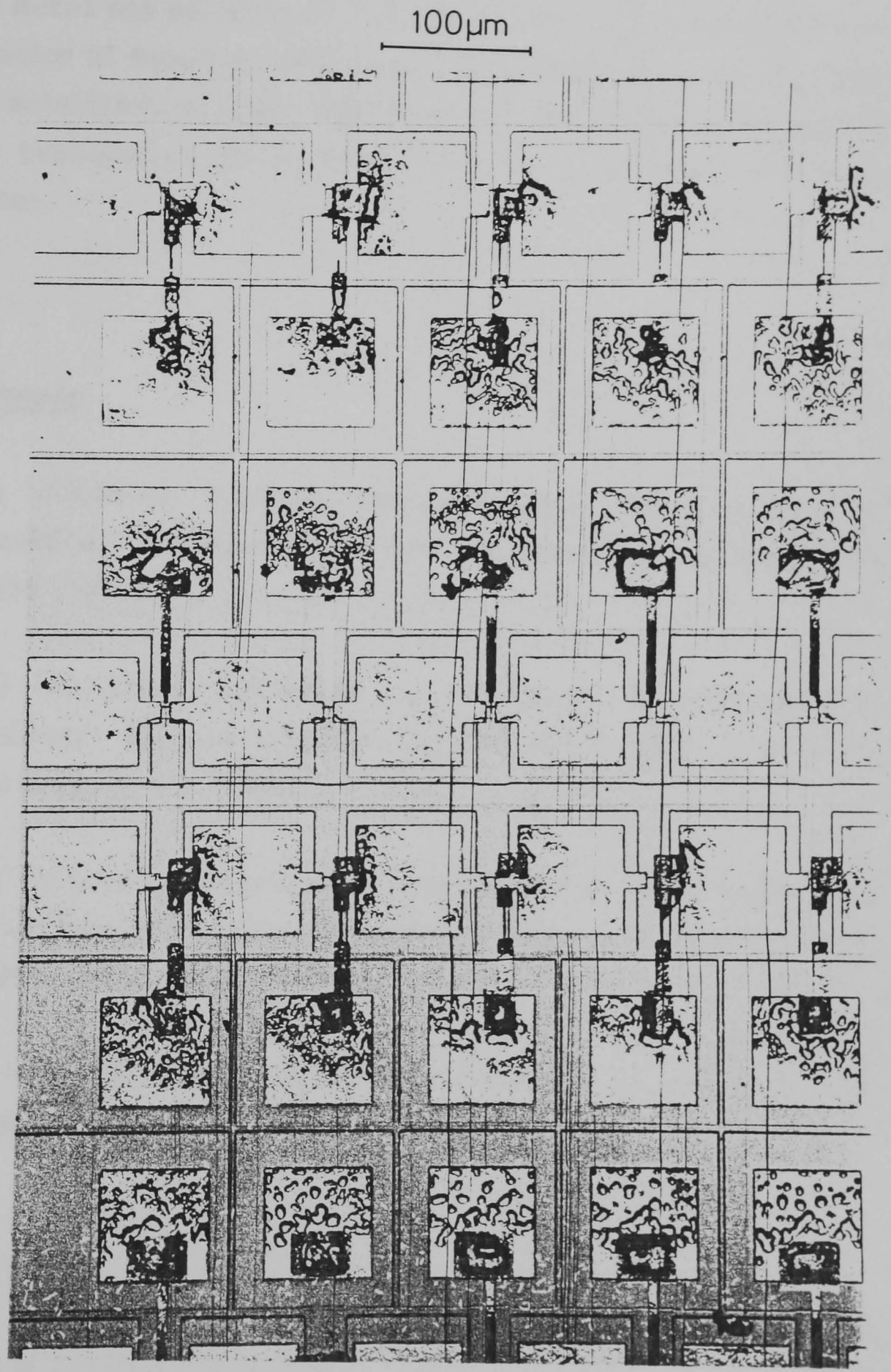


Fig 6.19 'Purple plague'. Aluminium gate contacts were heated to try to improve the diode characteristics. Massive interdiffusion of the Au and Al has taken place and the pattern was destroyed. As well as the gates on this chip, a pattern was exposed to try to measure gate resistances on the top and third rows of devices.

were used was that the gate metal goes onto a gold contact pad. So long as the device is not heated no apparent problems occur, but disaster happened when an anneal at 350°C was tried as a way of improving poor gate diode charactaristics (Fig. 6.19). The gate metal has not enjoyed the experience. It is possible that formation of eutectic compounds at the interface of the aluminium gate metallisation might degrade device performance, even without high temperature processing, so Al/Au interfaces should be avoided.

References

[6.1] B.M.Welch, F.H.Eisen,and J.A.Higgins, "Gallium Arsenide Field-Effect Transistors by Ion Implantation", J. App. Phys. 45 3685 (1974).

[6.2] R.C.Eden, B.M.Welch, and R.Zucca, "Planar GaAs IC Technology: Applications for Digital VLSI", IEEE J. Sol. St. Circuit. SC-13 419 (1978).

[6.3] M.Omori, T.J.Drummond, and H.Morkoc, "Low-noise GaAs Field-effect Transistors Prepared by Molecular Beam Epitaxy", App. Phys. Lett. 39 566 (1981).

[6.4] S.M.Sze, "Physics of Semiconductor Devices", 2nd Ed. p. 329, John Wiley and Sons, New York, 1981.

[6.5] Ref. 6.4 p.325.

[6.6] B.Himsworth, "A Two-Dimensional Analysis of Gallium Arsenide Junction-Field-Effect Transistors with Long and Short Gates", Sol. St. Electron. 15 1353 (1972).

[6.7] M.S.Shur, "Analytical Model of GaAs MESFETs", IEEE Trans. Electron Dev. ED-25 612 (1978).

[6.8] C.A.Liechti, "Microwave Field-Effect Transistors - 1976", IEEE Trans. Microwave Theory Tech., MTT-24 279 (1976).

[6.9] M.S.Shur and L.F.Eastman, "Ballistic Transport at Low Temparatures for Low-Power High-Speed Logic", IEEE Trans. Electron Dev. 26-26 1677 (1979).

[6.10] C.K.Williams, T.H.Glisson, M.A.Littlejohn, and J.R.Hauser, "Ballistic Transport in GaAs", IEEE Electron Dev. Lett. EDL-4 161 (1983).

[6.11] L.F.Eastman, R.Stall, D.Woodward, N.Dandekar, C.E.C.Wood, M.S.Shur, and K.Board, "Ballistic Electron Motion in GaAs at Room Temperature", El. Lett. 16 524 (1980).

[6.12] K.E.Drangeid, R.Sommerholder, and W.Walter, "High Speed Gallium-Arsenide Schottky-Barrier Field-Effect Transistors", El. Lett. 8 228 (1972).

[6.13] R.S.Butlin, A.J.Hughes, R.H.Bennet, D.Parker, and J.A.Turner, "J-Band Performance of 300nm Gate Length GaAs FETs", in 1978 Int. Electron Devices Meet. Dig. Tech. Papers, p136.

[6.14] H.Fukui, J.V.Dilorenzo, B.S.Hewitt, J.R.Velebir, H.M.Cox, L.C.Luther, and J.A.Seman, "Optimisation of Low-Noise GaAs FETs", IEEE Trans. Electron Devices, ED-27 1034 (1980).

[6.15] K.Ohata, H.Itoh, F.Hasegawa, and Y.Fujiki, "Super Low-Noise GaAs MESFETs with a Deep-Recess Structure", IEEE Trans. Electron Devices, ED-27 1029 (1980).

[6.16] R.A.Sadler, and L.F.Eastman, "High-Speed Logic at 300K with Self-Aligned Submicrometer-Gate GaAs MESFETs", IEEE Electron Dev. Lett. EDL-4 215 (1983).

[6.17] N.Kato, K.Yamasaki, K.Asai, and K.Ohwada, "Electron-Beam Lithography in n^+ Self-Aligned GaAs MESFET Fabrication", IEEE Trans. Electron Devices, ED-30 663 (1983).

[6.18] P.Tung, D.Delagebeaudeuf, M.Laviron, P.Decleuse, J.Chaplart, and N.T.Linh, "High-Speed Two-Dimensional Electron-Gas FET Logic", El. Lett. 18 109 (1982).

[6.19] R.L.Van Tuyl and C.A.Liechti, "High Speed Integrated Logic with GaAs", IEEE Sol. St. Circuit. SC-9 269 (1974).

[6.20] R.L.Van Tuyl, C.A.Liechti, R.E.Lee, and E.Gowen, "GaAs MESFET Logic with 4-GHz Clock", IEEE Sol. St. Circuit. SC-12 485 (1977).

[6.21] E.H.Rhoderick, "Metal-Semiconductor Contacts", p.19, Oxford University Press (1978).

[6.22] N.Braslau, J.B.Gunn, and J.L.Staples, "Metal-Semiconductor Contacts for GaAs Bulk Effect Devices", Sol. St. Electron. 10 381 (1967).

[6.23] J.S.Harris, Y.Nannichi, and G.L.Pearson, "Ohmic Contacts to Solution-grown Gallium Arsenide", J. App. Phys. 40 4575 (1969).

[6.24] G.Y.Robinson, "Variation of Schottky-barrier energy with interdiffusion in Au and Ni/Au-Ge films on GaAs", J. Vac. Sci. Technol. 13 884 (1976).

[6.25] A.Illeadis, "Metallurgical Behaviour of Ni/Au-Ge ohmic Contacts to GaAs", Sol. St. Comm. 49 99 (1984).

[6.26] J.C.Werthen, and D.R.Scifres, "Ohmic Contacts to n-GaAs using Low-temperature Anneal", J. App. Phys. 52 1127 (1981).

[6.27] O.Aina, W.Katz, B.J.Baliga, and K.Rose, "Low Temperature Sintered AuGe/GaAs Ohmic Contacts", J. App. Phys. 53 777 (1982).

[6.28] K.Heime, U.Konig, E.Kohn, and A.Wortmann, "Very Low Resistance Contacts to n-GaAs", Sol. St. Electron. 17 835 (1974).

[6.29] J.P.Donnelly and F.J.Leonberger, "Multiple-Energy Proton Bombardment in n^+-GaAs", Sol. St. Electron. 20 183 (1977).

[6.30] J.C.Dyment, J.C.North, and L.A.D'Asaro, "Optical and Electrical Propeties of Proton-Bombarded p-type GaAs", J. App. Phys. 44 207 (1973).

[6.31] D.C.D'Avanzo, "Proton Isolation for GaAs Integrated Circuits", IEEE Electron Devices ED-29 1051 (1982).

[6.32] P.Grabbe, R.E.Howard, E.L.Hu, and D.M.Tennant, "Metal-on-Polymer Masks for Reactive Ion Etching", Proc. Tenth International Conference on Electron and Ion Beam Science and Technology", Montreal 1982.

[6.33] Ref. 6.4 p. 248.

[6.34] Ref. 6.4 pp. 103 and 76.

[6.35] G.Curry, unpublished (1982).

[6.36] C.Binnie, Ph.D Thesis, Glasgow University, 1984.

[6.37] K.Kame, H.Kawasaki, T.Chigira, T.Naknisi, K.Kawabuch, M.Yoshim, "Extremely Low-noise MESFETs fabricated by MOCVD", Electron. Lett. 17 540 (1981).

CHAPTER 7 PERFORMANCE OF FETS

7.1 TESTING PROCEDURE

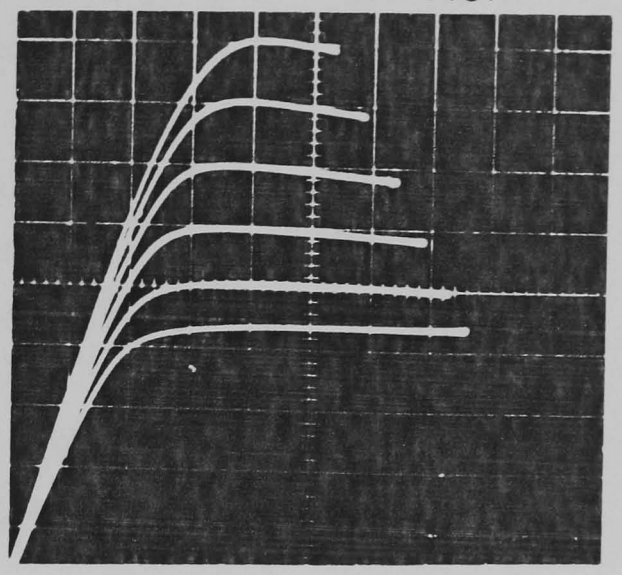
Transistors were prepared up to 40 at a time on 1.7 x 2mm dies as described in Chapter 6. The devices were examined in an optical microscope to check for defects such as a broken or missing gate, or a misaligned gate. A map of complete devices was drawn for each die. Usually more than half the devices would appear to be intact.

Devices were tested column by column with a Tektronix curve tracer connected to a simple prober with a simple manual stage. The recessed gate devices were tested during a visit to Plessey (Caswell).

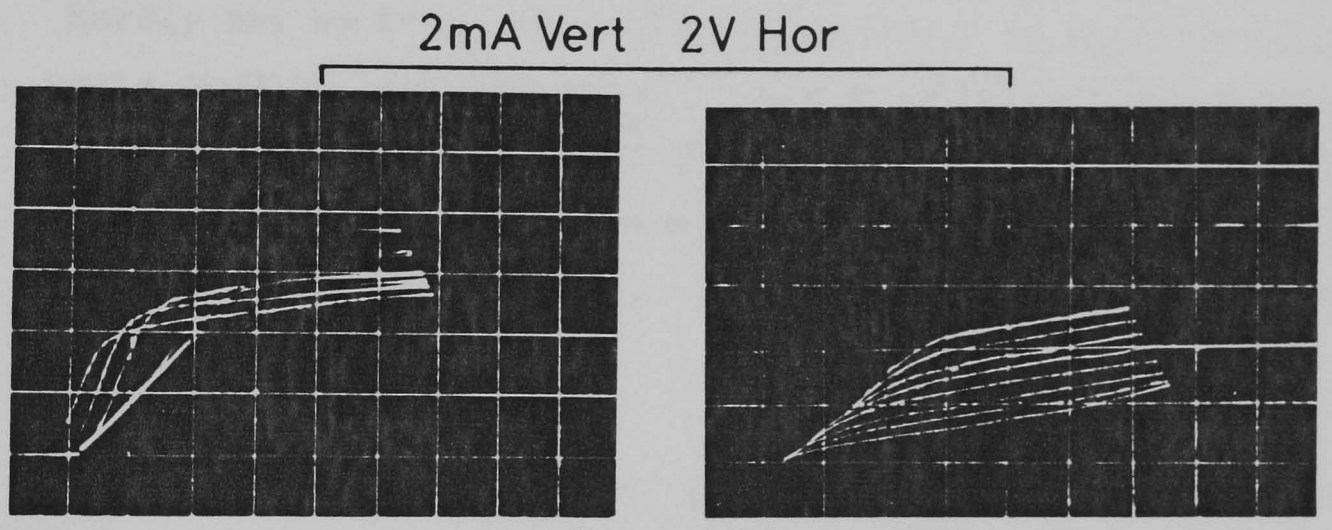
During testing of the first successfully fabricated die, the saturation current of each device was tested with the gate probe touching the gate pad, but not connected to the curve tracer. The gates of devices tested in this way evaporated. The problem did not occur if the gate was connected to the curve tracer from the start.

7.2 PLANAR FETS WITH AL/AUPD GATES

The first devices to be made had all the gates exposed together (Section 6.5.3.1). The gate lengths varied from 1.1-1.8 microns. Most of the devices of the one die that was fabricated in this way were destroyed when they were tested with the gate floating (see above). The characteristic of one of the few surviving devices is shown in Figure 7.1 (Top). The device had a gate with dimensions 1.5 x 30 microns. The gate voltage was



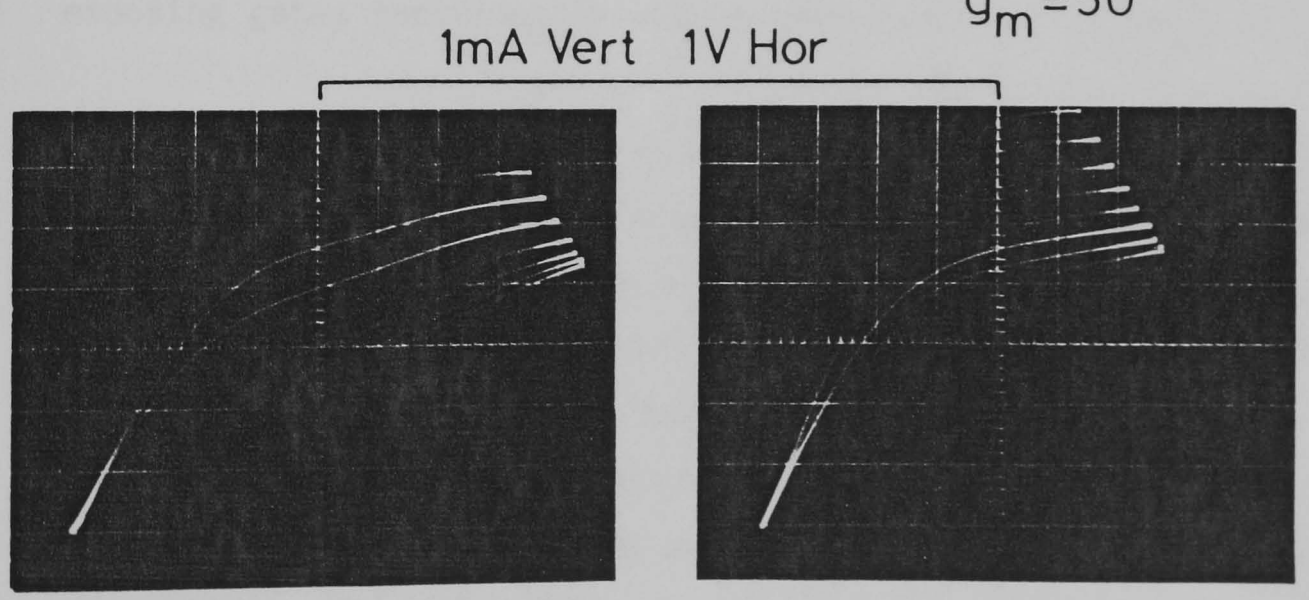
1·5 x 30 µm
$g_m = 80$



Before 'burn-in'

0·7 x 20µm

After 'burn-in'
$g_m = 50$



0·1 x 10µm
$g_m = 100$

0·075 x 10µm
$g_m = 60$

Planar Al-Au/Pd gates

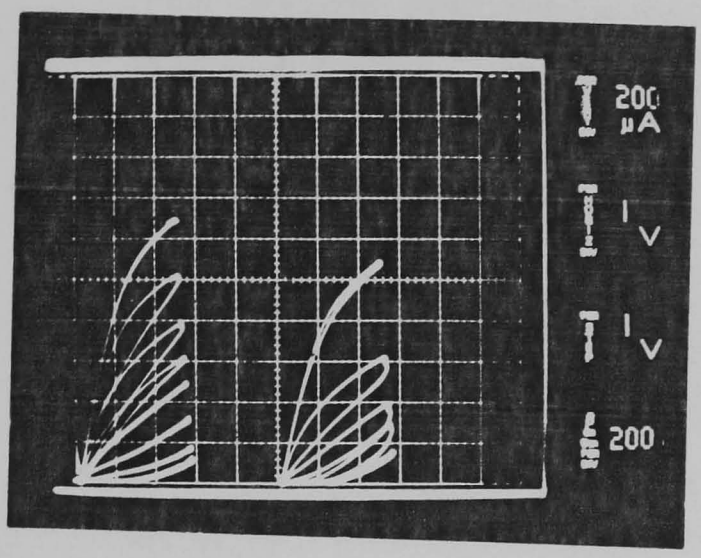
Fig. 7.1 Characteristics of planar FETs with Al-Au/Pd gates.
Gate dimensions are given as Length x Width
Transconductances, g_m, are given in mS/mm of gate width

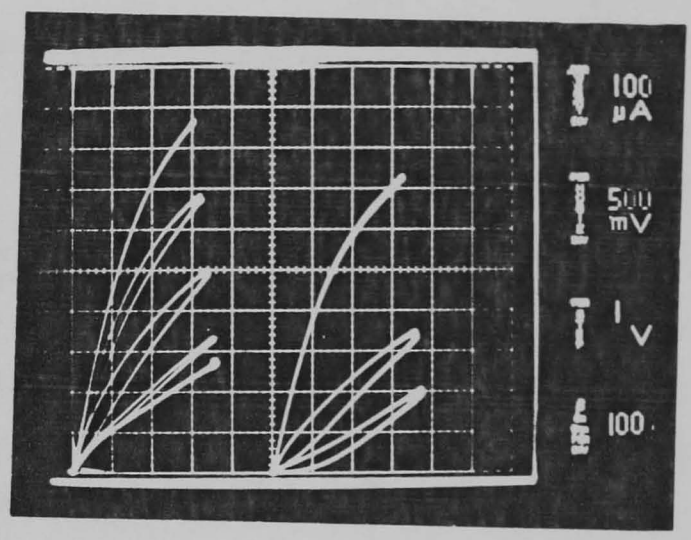
stepped by 0.8V. The device saturates well, but could not be pinched off, as was to be expected (Section 6.3). The transconductance was 80mS/mm of gate length. At low gate voltages the device exhibits negative resistivity instead of flat saturation. Wada et al. [7.1] explains this kind of behaviour in terms of velocity overshoot. Negative resistivity is seen in the characteristics of some commercial devices [7.2]. The device has a relatively high saturation voltage and this is probably due to the the parasitic resistance of the 10 micron drain source gap, as most of the applied voltage would be dropped outside the gate region where the saturating mechanims occur.

It was found with these devices that the gate excercised hardly any control when a device was first tested, but if the gate voltage was increased to its maximum (9.6V), and then brought back down to 4V, the characteristic was much improved. This appeared to indicate the presence of an interfacial layer under the gate that was being heat-treated by the current flowing during breakdown.

The results for this first group of transistors were encouraging, and it was decided to make smaller devices by exposing gates individually with a small spot size (6.5.3.2).

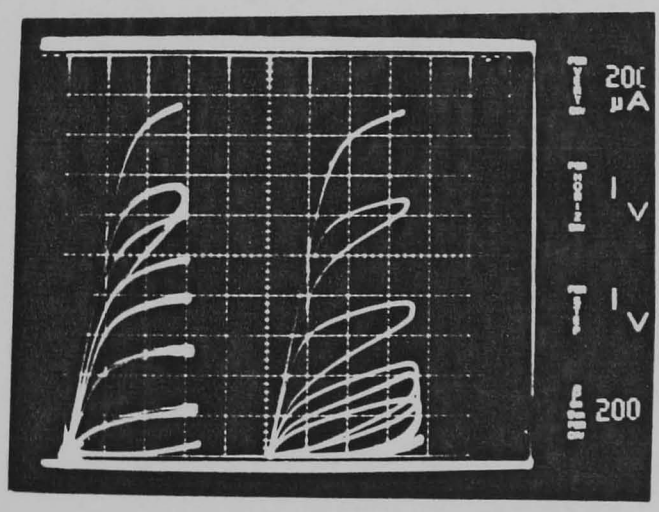
The middle pair of characteristics of Figure 7.1 show the characteristics of a 0.7 micron gate device fabricated by this method. The characteristics before and after burn-in are shown. The improvement after burn-in is obvious, although the transconductance of only 30 mS/mm for this particular device is rather poor, and was somewhat lower than the average of about 40 mS/mm for all the devices on the same die whose chactaristics were taken. The saturation voltage for the devices on this die was 6V compared with 4V for the previous devices, and this was probably because the ohmic contacts were alloyed at 330°C rather than 400°C, which has lead to increased contact resistance. The high contact resistance has lead to a low value for the

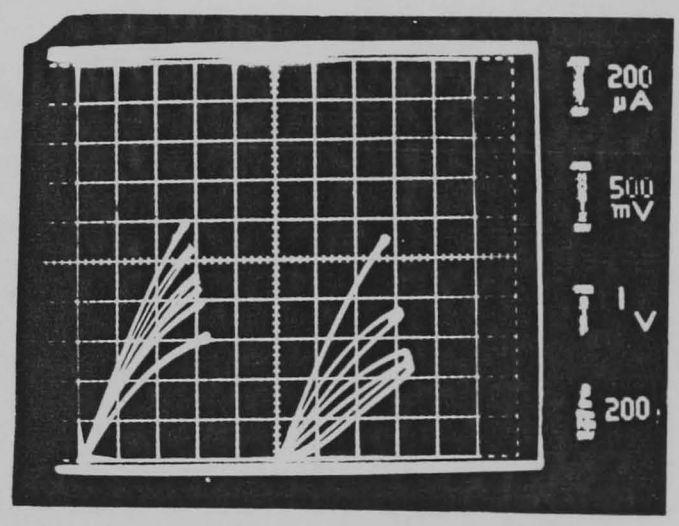




Before 'burn-in' 0·2 x 5µm After 'burn-in'

g_m = 60 100 40 80





0·35 x 5µm 0·14 x 5 µm

g_m = 50 90 30 80

Recessed Al-Au/Pd gates

Fig. 7.2 Characteristics of recessed-gate FETswith Al-Au/Pd gates Characteristics were recorded with illumination (left) and without (right.
Transconductances, g_m, are given in mS/mm of gate width

saturatation current, and hence a low transconductance. All further devices employed ohmic contacts alloyed at 400°C.

The bottom pair of characteristics shown on Figure 7.1 are from devices fabricated on a different die. These were the two smallest devices fabricated during the course of this work, with measured gate lengths of 0.1, and 0.075 microns respectively. The devices nearly saturate with no applied gate voltage, but the saturation is less good as the gate voltage increases. The transconductances are better than obtained on the previous die, being 85 and 75 mS/mm respectively, which is similar to that obtained for the 1.5 micron device. The 0.075 micron device is believed to be the smallest MESFET ever made.

None of the devices with gates smaller than 1 micron showed negative resistivity; this may have been because of current leakage through the buffer layer.

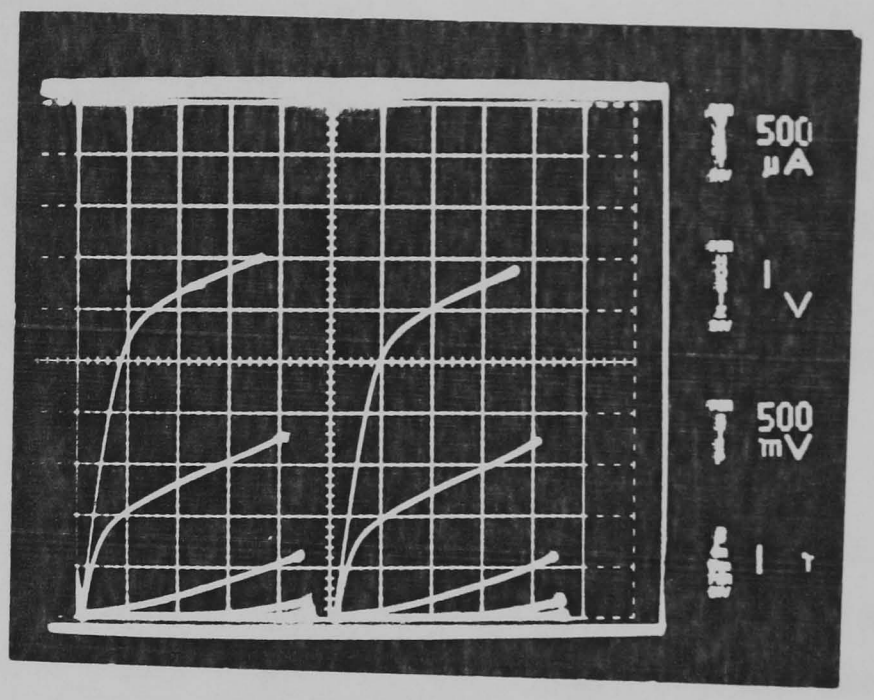
7.3 DEVICES WITH RECESSED GATES

The planar devices did not pinch off because the active layer was too thick, so, in order to fabricate devices that would pinch off, recessed gates were used on two dies (6.5.3.3).

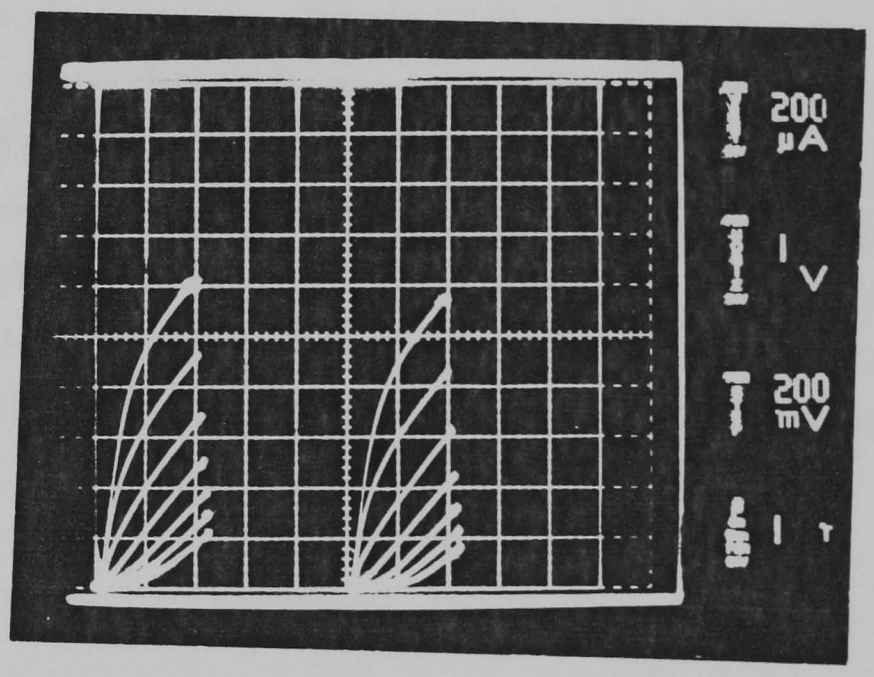
7.3.1 Al/AuPd Gates

The characteristics of some of the devices fabricated with recessed Al/AuPd gates are shown in Figure 7.2. The most striking feature, when compared to the characteristics of the planar devices of Figure 7.1, is their 'loopiness'. This is probably caused by interfacial states under the gate, formed because the gate metal was deposited under rather poor vaccuum conditions. The upper pair of characteristics shows the change that occurred after burn-in of the gate of one of the devices.

0·35 x 20 µm
g_m = 175



0·16 x 10 µm
g_m = 140



0·13 x 5 µm
g_m = 60

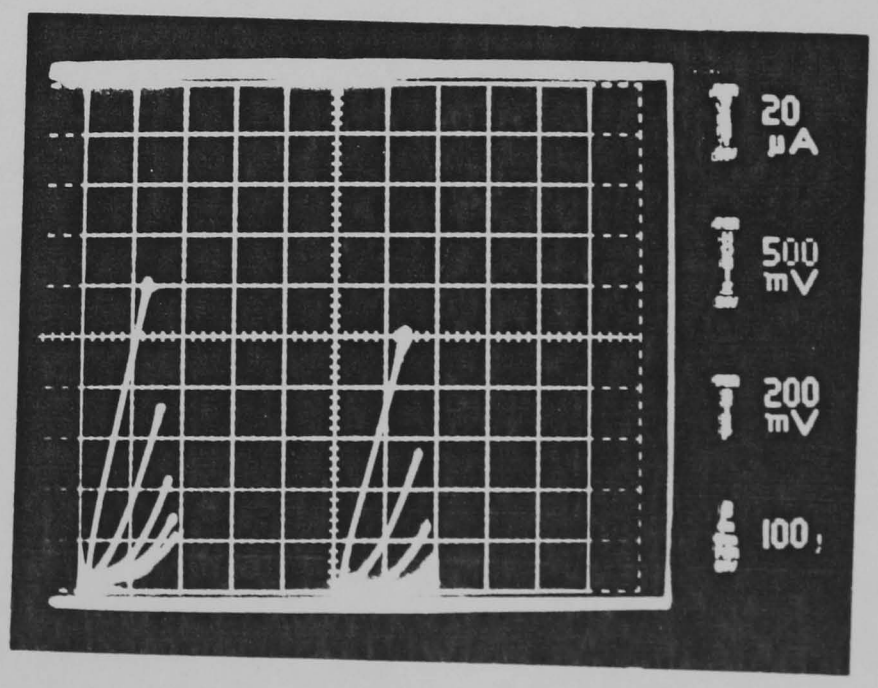


Fig. 7.3 Characteristics of recessed-gate FETs with Ti/Al gates
Characteristics were recorded with illumination (left) and without (right)
Transconductances, g_m, are given in mS/mm of gate width

The saturation current decreased as did the transconductance. The saturation current would be expected to decrease since the gate metal would be in more intimate contact with the GaAs surface after burn-in; why the transconductance also decreases is not understood. Examination of the diode characteristics showed that the voltage required for 10 microamps to flow (an arbitrary definition of the 'turn-on' voltage of the diode) dropped from 4V to 1.5V after burn-in. Device characteristics became less 'loopy' after burn-in.

The lower pair of characteristics shown on Figure 7.2 show that while devices with 0.35 micron gates pinched off well at about 7V, and showed some indication of saturation, the narrower devices could not be pinched off, and did not saturate. A feature of all the devices tested was that when the characteristics were measured in light and in dark, it was found that the transconductance was greater and the saturation current (taken as that at the top of the knee of the characteristic) was slightly smaller in the dark. The transconductances obtained for these recessed devices in the dark are similar to those obtained for planar devices in light.

7.3.2 Ti/Al Gates

These devices had much better diode characteristics than any of the Al/AuPd devices. Turn-on of the gate diode was at about 0.4V, but the characteristic did not contain a flat portion below turn-on as the Al/AuPd devices had done, the characteristic being a smooth curve.

Some transistor characteristics are shown in Figure 7.3. The 0.35 micron devices showed reasonable saturation, and had saturation voltages of about 1V, compared with 4V for the planar devices. The reduction in saturation voltage occurred because the recess caused a greater proportion of the applied voltage to

be dropped across the depletion region under the gate than occurred in planar devices. The transconductances of 0.35 and 0.16 micron devices were much better than any devices fabricated with Al/AuPd gates, probably because titanium forms conducting oxides and insulating interfacial layers do not occur. The transconductance of the 0.13 micron device appears to be poor in comparison, but if the device had been taken up to a similar current as the others, the transconductance might have improved. The smaller devices did not saturate and this was probably again because of the graded interface between the active and buffer layers of the VPE wafer. The interface has a greater effect in recessed devices than in planar ones because the conducting channel has been thinned and interface leakage forms a larger proportion of the total current flowing under the gate.

The characteristics of the 0.35 micron device are quite similar to those of SAINT-FETs [7.3], which also do not show complete saturation and also have transconductances of about 175 mS/mm. Interestingly also, the characteristics of the smaller devices from this work are rather similar to those of a SAINT-FET with a deep implant (thicker active layer). The inference that might be drawn from this is that if FETs were fabricated on thinner active layers, or with deeper recesses, the characteristics of devices with gates smaller than 0.2 microns might improve markedly.

The vertical spread of an implantation gives rise to a graded interface to the underlying semi-insulating GaAs, similar to that between the buffer and the active layer on the wafer used here. Thus the decrease in transconductance close to threshold seen on both sets of devices might be attributable to the same cause, viz. leakage through the graded interface below the active layer.

7.4 CONLUSIONS

Devices were fabricated with gate lengths varying from 1.5 down to 0.075 microns using planar and recessed technologies. The longer planar devices showed some evidence of overshoot effects, while in the shorter planar devices, conduction at the graded interface to the buffer layer appeared to mask the effect. Recessed devices had lower saturation voltages than planar devices, but did not saturate as fully. Experiments with MBE- or MOCVD-grown wafers with abrupt interfaces (0.5-1nm) should show overshoot effects for small devices, and also give better saturation for recessed devices.

In the context of the project as a whole, fabrication of mesa-isolated solid substrate transistors was the last step that was achieved from the strategy outlined in Section 1.8. Failure to achieve a removable proton implantation mask and lack of time prevented transistors from being made on thin substrates of GaAs.

Thin substrates of GaAs may well exhibit interesting properties because of the built-in potential of exposed surfaces of GaAs due to surface states. The potential associated with these states amounts to about -0.7 volts usually, and on bulk substrate, has the effect of depleting the GaAs of electrons to a depth corresponding to a reverse bias diode potential of 0.7 volts. For a substrate doped with 10^{18} n-type carriers cm^{-1} this depth is about 25nm. In the case of a thin membrane, there will be a potential on both the exposed surfaces, and it seems likely that a potential well will form, and the electrons will be confined to a thin layer in the centre of the membrane. This raises the questions of whether a Schottky barrier FET will work on such a substrate, what the operating characteristics of such a device will be if it does work, and whether novel devices which utilise the physics of the membrane to advantage can be fabricated.

The techniques developed in this work concerning fine line lithography on thin and solid substrates, high accuracy alignment, and membrane fabrication will allow exploration of the limits of conventional devices such as the GaAs MESFET, and perhaps the development of entirely new devices which take advantage of the physics of the very small.

References

[7.1] T.Wada and J.Frey, "Physical Basis of Short-Channel MESFET Operation", IEEE Trans. Electron Dev. ED-26 476 (1979).

[7.2] L.O.Chua and Y.W.Sing, "Non-Linear Lumped Circuit Model of GaAs MESFET", IEEE Trans. Electron Dev. ED-30 825 (1983).

[7.3] N.Kato, K.Yamasaki, K.Asai, and K.Ohwada, "Electron Beam Lithography in n^+ Self Aligned GaAs MESFET Fabrication", IEEE Trans. Electron Dev. ED-30 663 (1983).