https://theses.gla.ac.uk/

Theses Digitisation:

https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

The Fabrication and Evaluation of Silicon MOSFETs with 0.5 to 0.1 Micron Gate Lengths

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

by

Clive Malcolm Reeves

May 1987

(C) Clive Malcolm Reeves, 1987

ProQuest Number: 10997381

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10997381

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Summary

This Thesis describes the application of high resolution electron beam lithography and dry etching techniques to the fabrication of experimental silicon MOSFETs with sub 0.5 micron gate lengths.

HRN, a negative electron beam resist is investigated and demonstrated to have a resolution of 0.1 microns. Two alternative dry etching processes are reported for patterning 0.1 micron polysilicon gate electrodes using HRN masking. The first process uses chlorine plasma etching whilst the second process uses silicon tetra–chloride reactive ion etching.

MOSFET scaling theory is introduced and then used as a guide for designing the experimental sub 0.5 micron devices.

A full processing sequence is developed for fabricating the experimental devices. This process includes self–aligned ion implantation and rapid thermal annealing steps, to form the shallow source and drain drift regions. Devices with gate oxide thicknesses of 150 angstroms and channel doping levels in the range from $3x10^{16}$ to $1.2x10^{18}$ atoms/cm^3 have been implemented.

Electrical measurements are reported for the range of devices which have been fabricated. Results are included for MOSFETs with gate lengths of only 0.11 microns. At a channel doping level of $1.2x10^{18}$ atoms/cm^3 the 0.11 micron devices exhibit a transconductance of 70 mS/mm.

A slightly modified process is described which has been used to implement a second set of sub 0.5 micron devices together with 19 stage unloaded n–MOS ring oscillator circuits. Preliminary high speed measurements are reported for these circuits which have gate lengths of 0.23 microns. Minimum stage delays of 80 psec have been achieved with a corresponding power–delay product of 210 fJ.

Contents

Page

Chapter 3

Objectives and Experimental Approach

Chapter 4

Submicron Lithography and Etching

Chapter 4 cont.

Chapter 5

Selection of Critical Process Parameters

Page

Chapter 6

The MOSFET Processing Sequence

Chapter 7

B.T. Device Results

Chapter 7 cont.

Chapter 8

E.M.F. Device Results

Page

Chapter 1

Introduction

For many semiconductor devices, a significant improvement in performance can be achieved by reducing physical device dimensions. In the case of silicon MOSFETs, improved packing density, operating speed and power–delay product can be achieved[1.1]. The advance of silicon integrated circuit technology which has taken place during the past two decades has been, in part, due to a gradual reduction of device dimensions. During this period, integrated circuit technology has progressed from the Small Scale Integration (SSI) of approximately one hundred logic elements per chip to present day Very Large Scale Integration (VLSI)[1.2] with more than one hundred thousand logic elements per chip. Over this period the minimum feature size obtained by optical lithography has decreased from 10 microns to below 1 micron.

1.1 MOSFET Scaling

Various MOSFET scaling theories have been formulated in order to ensure optimum performance whilst reducing device dimensions. Dennard et al[1.3] proposed a set of MOSFET scaling rules, as shown in table (1–1), which predict a scaled device with similar electric field distributions as for a larger reference device. These field distributions have identical amplitudes as for the reference device but are scaled in co–ordinate space. When the linear dimensions are scaled by a factor 1/a, the channel doping level is increased by a factor a and operating biases are decreased by a factor 1/a. The table also shows the predicted relative performance for the scaled device. It can be seen that the maximum operating frequency increases by a factor a, the circuit packing density increases by a^2 and the device power consumption decreases by $1/a^2$. The power–delay product which is of special significance for digital applications decreases by $1/a^3$. Whilst this scaling theory is highly simplified, it does give a first order indication of the performance advantages which might be possible through device scaling.

1.2 Sub 0.5 Micron Lithography

The trends of continued reduction of device dimensions together with increased levels of circuit integration have required extensive developments in both wafer processing technology and optical lithography systems. Recently, optical wafer steppers have been used to demonstrate the fabrication of experimental high performance n– MOS VLSI circuits with 0.5 micron design rules[1.4,1.5]. Commercial research interest is at present directed towards the exploitation of VLSI technologies at these dimensions, for realising circuit densities of more than one million logic elements per chip.

The ultimate feature resolution which can be achieved using optical lithography is limited by the exposure wavelength[1.6]. It therefore seems likely that the latest optical wafer steppers which use deep U.V. excimer laser sources[1.7] to produce 0.5 micron patterns across large chip fields, are close to the ultimate limit for optical lithography.

Both electron beam lithography[1.8] and X– ray printing[1.9] can be used to define extremely high resolution patterns. In the case of electron beam lithography, a minimum linewidth of 100 angstroms has been reported[1.10], and for X– ray printing a value of 175 angstroms has been reported[1.11]. Whilst both techniques offer a resolution capability far beyond the limits for optical lithography, current generation exposure systems using these techniques are slow. Consequently, they are unsuitable for the patterning of complex VLSI wafers, where high wafer through– put is required. However, within the next five years, advances in electron beam and X– ray[1.12,1.13] exposure machine technology are likely to permit the commercial exploitation of VLSI technologies with 0.5 to 0.1 micron design rules. These machines will permit high wafer through– put together with 0.5 to 0.1 micron pattern fidelity and alignment. Meanwhile, considerable research into device fabrication methods and device optimisation is required at these dimensions.

1.3 MOSFETs with 0.5 to 0.1 Micron Gate Lengths

This thesis contains a preliminary study of MOSFET device fabrication for future VLSI applications which will utilise 0.5 to 0.1 micron design rules. The most critical aspect of device scaling relates to scaling of the active region under the gate. In the experiments described here, channel length scaling to between 0.5 and 0.1 microns is investigated. Full design rule scaling to these dimensions will require extensive research into isolation, contact window and interconnect scaling and is acknowledged to be too ambitious for this study.

A high resolution vector scan electron beam lithography system[1.8] is used for patterning the 0.5 to 0.1 micron gates. Whilst the vector scan system has a low wafer through–put, it is well suited to this research application where only a relatively small number of experimental devices are required.

1.4 Overview of the Thesis

The concept of MOSFET scaling is discussed in detail in Chapter 2, with special reference to sub 0.5 micron channel length scaling. Chapter 3 then describes the objectives of the thesis and introduces the proposed experimental MOSFET structure.

Chapter 4 describes the development of an electron beam lithographic and dry etching process which can be used to pattern the 0.1 micron polysilicon gate electrodes. Next, chapter 5 describes the processing steps which are of critical importance for MOSFET fabrication at these dimensions. A matrix of processing parameters is chosen for these steps in order to obtain a spread of device results. A more detailed description of the fabrication process is presented in chapter 6.

Having successfully fabricated the experimental devices, they were tested as described in Chapters 7 to 9. The d.c. characteristics are presented in Chapter 7 and various material and process parameters are calculated from the measurements.

The d.c. characteristics of a second set of devices are presented in chapter 8. These devices were fabricated using a slightly modified

process and a new mask set. The new mask set includes ring oscillator circuits for testing the high speed performance of the devices. Some preliminary high speed measurements obtained from these circuits are presented in chapter 9.

Finally, an overall conclusion to the thesis is presented in chapter 10 together with suggestions for future research.

References

1.1 G.E. Moore, 'VLSI: Some fundamental challenges,' IEEE Spectrum, vol 16, p530, 1979.

1.2 S.M.Sze (Ed.), 'VLSI technology,'McGraw–Hill, New York, 1984.

1.3 R.H. Dennard, F.H. Gaensslen, H. Yu, V.L. Rideout, E. Bassous and A.R. LeBlanc, 'Design of ion–implanted MOSFETs with very small physical dimensions,' IEEE J. Solid–State Circuits, vol. SC–9, p256, 1974.

1.4 W. Fichtner, E.A. Hofstatter, R.K. Watts, R.J. Bayruns, P.F. Bechtold, R.L. Johnson and D.M. Boulin, 'A submicron nMOS technology suitable for low power high speed circuits,'IEDM Tech. Dig., p264, Washington, 1985.

1.5 W. Fitchner, E.A. Hofstatter, R.K. Watts, R.J. Bayruns, P.F. Bechtold, R.L. Johnson and D.M. Boulin, 'High–speed low–power circuits fabricated using submicron NMOS technology,' IEEE Electron Dev. lett., vol EDL–6, p663, 1985.

1.6 B.T. Lin,'Optical methods for fine line lithograph,' in R.Newman (Ed),'Fine line Lithography,' North–Holland, Amsterdam, 1980.

1.7 V. Pol et al.,'Excimer laser–based lithography: a deep U.V. wafer stepper,' SPIE, Santa Clara, 1986.

1.8 G.R. Brewer (Ed),'Electron–beam technology in micro–electronic fabrication,' Academic, New York, 1980.

1.9 D. Maydon, 'X–ray lithography for microfabrication,' J. Vac. Sci. Technol., Vol 17, p1164, 1980.

1.10 H.G. Craighead, R.E. Howard, L.D. Jackel and P.M. Mankiewich, '10–nm linewidth electron beam lithography on GaAs,' Appl. Phys. Letts., Vol 42, p38, 1983.

1.11 D.C. Flanders, 'Replication of 175Å lines and spaces in polymethylmethacrylate using x–ray lithography,' Appl. Phys Lett., vol 36, p93, 1980.

1.12 J. Hersener and G. Lassing, 'A MOSFET, manufactured with synchrotron X–ray lithography,' Proc. Microcircuit Engineering 1986, p105, North–Holland 1986.

1.13 K. Suzuki, K. Okada, J. Matsui, N. Eudo, Y. Iida and M. Fukuma, 'Submicrometer CMOS fabrication using an X–ray stepper,' IEEE Electron Dev. Lett., vol EDL–6, p353, 1985.

Dimension	Scaling Factor
Linear Dimensions	$1/a$
Operating Voltages	$1/a$
Channel Doping	a
Max. Frequency	a
Packing Density	a^2
Power Consumption/Device	$1/a^2$
Power-delay Product	$1/a^3$

Table(1-1) MOSFET scaling rules which retain constant electric fields (CE scaling).

Chapter 2

The Scaled MOSFET

2.1 Introduction

Standard n−MOS VLSI technology requires both n−channel surface enhancement mode and depletion mode MOSFETs. The enhancement devices act as switching elements with the depletion devices as their active loads. In this chapter, the operation of conventional long channel MOSFETs is first explained and then a detailed discussion of MOSFET scaling is presented.

2.2 The Long Channel MOSFET

The n−channel surface enhancement mode MOSFET consists of a gate electrode insulated by a thin thermally grown gate oxide from the underlying p−type substrate. Highly doped n−type regions are formed in the substrate at both sides of the gate by self−aligned ion implantation. Source and drain contacts are then formed to these regions, resulting in the cross−sectional structure as shown in figure (2−1).

Various device dimensions will be defined during the course of this chapter, and referred to elsewere in the thesis. A number of these dimensions, including channel length (L) and width (W), gate oxide thickness (t_{ox}) and the source/drain junction depth (r_j) are shown in figure (2−2) where X, Y and Z axes are also defined.

Figure (2−3) shows an n−channel enhancement device with bias voltages applied. Once again, the notation is retained throughout the thesis and it can be seen that all voltages are defined with respect to the source. Under normal operating conditions, the gate and the drain are biased positively with respect to the source and the substrate (bulk) contact is biased negatively or at zero volts. i.e. V_g and V_d are positive and V_b is negative or zero.

MOSFETs can be biased into three different operating regions[2.1] which are termed **subthreshold**, **linear** and **saturation** and discussed separately below.

When the gate voltage (V_g) is below a certain threshold value, the p−type substrate has a high hole carrier concentration near to the substrate−to−oxide interface and depletion regions enclose the source and drain junctions, as shown in figure (2−4a). Under these conditions, the device is effectively turned off and only an extremely small drain current can flow. This region of operation is termed **subthreshold** and is discussed further in section 2.2.3. When the gate voltage (V_g) is increased, the electric field orthogonal to the oxide interface (E_x) increases which leads to bending of the substrate energy bands. Above a certain critical gate voltage, sufficient bending occurs to result in the formation of a thin layer of strong carrier inversion under the gate. This inversion layer opens a conducting channel between the source and the drain and the device is effectively turned on as shown in figure (2−4b). The device threshold voltage (V_{th}) is defined as the gate voltage (V_g) required to establish strong carrier inversion and can be modelled using equation (2.1). The substrate energy band diagram at threshold is shown in figure (2−5).

$$V_{th} = \Phi_{ms} + 2\Psi_B + \frac{N_{ss}.q}{C_{ox}} + \frac{(2\epsilon_s q N_A(2\Psi_B - V_b))^{1/2}}{C_{ox}} \qquad (2.1)$$

When:

Φ_{ms} = metal-semiconductor work function difference
Ψ_B = intrinsic-extrinsic Fermi level difference
ϵ_s = permittivity of silicon
q = electronic charge
N_A = channel doping level
N_{ss} = surface states/unit area
C_{ox} = gate oxide capacitance/unit area = ϵ_{ox}/t_{ox}
ϵ_{ox} = permittivity of SiO_2
t_{ox} = gate oxide thickness

Figure (2−6) shows a graph of threshold voltage (V_{th}), modelled by equation (2.1), as a function of channel doping level (N_A) for various oxide thicknesses (t_{ox}). A substrate voltage (V_b) of zero volts has been assumed together with a surface state density (N_{ss})

of zero. Work function difference values (Φ_{ms}) were chosen appropriately for a heavily doped n− type polysilicon gate electrode[2.2]. Curves are plotted for oxide thicknesses (t_{ox}) of 300, 150, 80, 40 angstroms which, as will be seen later, give an appropriate range for devices with 0.5 to 0.1 micron channel lengths.

N− channel depletion devices operate similarly to enhancement devices except that they have a negative threshold voltage (V_{th}) and so are normally turned on. The negative threshold voltage is achieved by shallow arsenic implantation during fabrication.

For both enhancement and depletion mode devices, when the gate voltage (V_g) exceeds the threshold voltage (V_{th}), the devices can be biased into linear and saturation operating regions, as discussed next.

2.2.1 Linear Region

Operation in the linear region occurs at low drain voltage (V_d) when the gate voltage (V_g) exceeds the threshold voltage (V_{th}). Under these conditions, the drain current (I_d) flows in the channel by a carrier drift mechanism, caused by the longitudinal electric field (E_y). Since this field varies linearly with applied drain voltage (V_d), the drain current (I_d) also varies linearly.

In the linear region, the drain current (I_d) is also proportional to ($V_g - V_{th}$) since this term controls the channel region carrier concentration via the gate oxide capacitance.

Equation (2.2) represents a simple model for the drain current due to carrier drift in the linear region for very low drain voltages (V_d) (ie $V_d \approx kT/q$). The equation can be considered as the product of a carrier charge density term, a carrier mobility term and a longitudinal electric field term.

$$I_d = W.C_{ox}(V_g - V_{th}).\mu_n.V_d/L \qquad (\text{for } V_d \approx kT/q) \qquad (2.2)$$

With terms defined overleaf.

When:

W = channel width

L = channel length

μ_n = electron mobility in the inversion layer

k = Boltzmann constant

T = absolute temperature

and other terms as defined previously.

2.2.2 Saturation Region

When the gate voltage (V_g) exceeds the threshold voltage (V_{th}), as the drain voltage (V_d) is increased, the linear relationship with drain current (I_d) is gradually lost and the drain current becomes saturated. This occurs when the longitudinal electric fields (E_y) are sufficiently large to reduce the degree of bending of the substrate energy bands near to the drain, such that no carrier inversion occurs in this region. The region is depleted of carriers and the device is said to be **pinched–off**. Channel current still flows, and once the carriers reach this region they are injected across to the drain. Beyond pinch–off, any additional drain voltage is developed across the depletion region and consequently the channel current remains saturated at the initial pinch–off value. Figure (2.4c) illustrates a device at the onset of pinch–off, and figure (2.4d) shows a device operating beyond pinch–off . The drain saturation current (I_{dsat}) is modelled by equation (2.3) from which it can be seen that the saturation current is proportional to $(V_g - V_{th})^2$ and is independent of drain voltage.

$$I_{dsat} = \frac{mW}{L}.\mu_n C_{ox}(V_g - V_{th})^2 \qquad \text{(const. mobility)} \qquad (2.3)$$

When:

m = function of channel doping profile (typ. m=1/2)

and all other terms are as previously defined.

It should be noted that the long channel MOSFET is designed such that the electric fields orthogonal to the gate oxide (E_x) are very much greater in magnitude than the longitudinal fields (E_y) in the channel region. This enables a simple **gradual channel** approximation to be used to formulate a full analytical equation which is shown below as equation (2.4). This equation models the device behaviour in the linear region and during the transition to the saturation region. It should be noted that the equation assumes no surface states (N_{ss}) and a back bias voltage (V_b) of zero volts. This equation which models long channel MOSFETs with considerable accuracy, appropriately reduces to equation (2.2) for small drain voltages (V_d) and can be reduced to equation (2.3) at the onset of saturation.

$$I_d = \mu_n(W/L)C_{ox}.\{(V_g-2\Psi_B-V_d/2)V_d- \frac{(8\epsilon_s qN_A)^{1/2}[(V_d+2\Psi_B)^{3/2}-(2\Psi_B)^{3/2}]}{3C_{ox}}\} \qquad (2.4)$$

With all terms as previously defined

The electrical I– V output characteristic curves for a typical n– channel enhancement mode MOSFET are shown in figure (2.7a) where drain current (I_d) is plotted as a function of drain voltage (V_d) for various gate voltages (V_g). Both linear and saturation regions are indicated.

The transconductance of a MOSFET is defined by equation (2.5) and an analytical expression for device transconductance, as shown by equation (2.6), can be obtained by substituting equation (2.3) into equation (2.5).

$$g_m = \frac{d\ I_{dsat}}{d\ V_g} \qquad \text{(for const } V_d\text{)} \qquad (2.5)$$

$$g_m = \frac{2mW\mu_n\epsilon_{ox}(V_g - V_{th})}{L} \qquad \text{(const. mobility)} \qquad (2.6)$$

The output conductance (g_d) of a MOSFET is defined by equation (2.7).

$$g_d = \frac{d\, I_{dsat}}{d\, V_d} \qquad \text{(for const. } V_g\text{)} \qquad (2.7)$$

For a MOSFET operating as an analogue small signal amplifier, the open circuit voltage gain (A_v) can be calculated using equation (2.8) from which it can be seen that increased voltage gain (A_v) is achieved through increased transconductance (g_m) and reduced output conductance (g_d).

$$Av = \frac{g_m}{g_d} \qquad (2.8)$$

2.2.3 Subthreshold Operation

As indicated earlier, when the gate voltage (V_g) is below the threshold voltage (V_{th}), the device is effectively turned off, with only a small drain current (I_d) which flows as a function of gate voltage. This current is due to a carrier diffusion mechanism across the source to drain gap. Figure (2.7b) shows a plot of subthreshold current for a typical MOSFET where drain current is plotted on a logarithmic scale as a function of gate voltage. The straight line region of the plot indicates an exponential relationship between subthreshold drain current (I_d) and gate voltage (V_g). The threshold voltage (V_{th}) is indicated by a deviation from this straight line. The subthreshold swing (S_w), is defined by equation (2.9) and gives a quantative measure of the abruptness with which the device turns on when the gate voltage (V_g) approaches threshold.

$$S_w = \log_e 10 . \frac{d(I_d)}{d(\ln(V_g))} \qquad \text{(for small } V_d\text{)} \qquad (2.9)$$

For a long channel MOSFET, the subthreshold swing (S_w) can be calculated analytically using equation (2.10), where C_D is the channel region depletion capacitance at threshold. This depletion region has a thickness (X_m) which can be modelled by equation (2.11). When C_D is much smaller than C_{ox}, a subthreshold swing of approximately 60 mV/decade is predicted at an operating temperature of 300K. However, in practice, this figure is usually degraded to about 70– 80mV/decade due to the presence of surface states.

$$S_w = \frac{kT}{q} . \log_e 10 . (1+C_D/C_{ox}) \qquad (2.10)$$

When:

C_D = channel region depletion capacitance = ϵ_s/X_m

X_m = channel region depletion thickness (at V_{th})

For $X_m = (2\epsilon_s(2\Psi_B - V_b)/qN_A)^{1/2}$ (2.11)

2.3 Short Channel Effects

The importance of MOSFET scaling for future silicon technologies has been identified in chapter 1, where it was also indicated that other device parameters must be scaled together with channel length (L) and width (W). These additional parameters include gate oxide thickness (t_{ox}), channel doping level (N_A), source/drain junction depths (r_j) and operating voltages (V_{th}, V_{DD} and V_b). However, if these parameters are not adequately scaled, the resulting device performance is degraded by a number of short channel effects.

Short channel effects occur when the longitudinal electric fields (E_y) are of comparable magnitude to the transverse fields (E_x).

When this occurs, the field distributions become more two dimensional in character and the gradual channel approximation loses validity.

At the onset of short channel behaviour, device threshold voltage becomes reduced in comparison to the value for a long channel device and it also becomes a function of drain voltage. Device performance in the saturation region is also degraded, since the drain current does not properly saturate above pinch−off. Instead, the drain current continues to increase as a function of drain voltage, corresponding to an increased output conductance (g_d) which thereby reduces the open circuit gain (A_v). Whilst a detailed two dimensional analysis of the field distributions is required to explain these effects, a simplified understanding can be obtained by considering device depletion widths. Short channel effects occur when the gate is so short that the source and drain junction depletion regions begin to encroach significantly under the gate. The threshold voltage is therefore decreased since less gate charge is required to deplete and then invert the remaining region under the gate. The reduced threshold voltage for a short channel device $V_{th(S)}$, can be calculated approximately using equation (2.12) which incorporates a long channel threshold term (V_{th}), as defined earlier. This approximation assumes that the gate charge is required to deplete a trapezoidal region[2.3] under the gate, as shown in figure (2−8), where the depletion widths at the source (y_s) and the drain (y_d) can be modelled by equations (2.13i) and (2.13ii) respectively.

$$V_{th(S)} \approx V_{th} - \frac{qN_A X_m r_j}{2C_{ox}L} .\{[1+2y_s/r_j]^{1/2}-1\}.\{[1+2y_d/r_j]^{1/2}-1\} \qquad (2.12)$$

When:

V_{th} = long channel threshold voltage

r_j = source/drain junction depth

X_m = channel depletion thickness

y_s = source depletion width at surface

y_d = drain depletion width at surface

$$y_s = (2\epsilon_s[V_{bi}-2\Psi_B]/qN_A)^{1/2} \quad (2.13i)$$

$$y_d = (2\epsilon_s[V_{bi}-2\Psi_B+V_d]/qN_A)^{1/2} \quad (2.13ii)$$

V_{bi} = junction built-in voltage

As the drain voltage increases the drain depletion region widens further under the gate, thereby explaining the decrease in threshold voltage with drain voltage.

Whilst for long channel devices operating above pinch–off, the pinch–off depletion width does not significantly modulate the channel length as a function of drain voltage, in the case of short channel devices, this modulation becomes significant. As the pinch–off depletion region widens, the channel length is reduced leading to an increased channel conductance and drain current. The resulting electrical I–V output characteristic of a short channel device where channel length modulation is significant is shown in figure (2.9a).

The threshold voltage reduction resulting from short channel behaviour also leads to an increase in the subthreshold current with increased drain voltage as indicated in figure (2.9b). However, the subthreshold swing (S_w) remains constant.

As gate length is reduced further, severe short channel effects become apparent. Gate length is reduced to the point where under normal drain operating voltages the source and drain depletion regions merge and the device becomes **punched–through**. The drain voltage (V_d) at which punch–through occurs (V_{pch}) can be modelled by equation (2.14).

$$V_{pch} = \frac{qN_A(L-y_s)^2}{2\epsilon_s} \quad (2.14)$$

With y_s as previously defined.

The graph in figure (2.10) shows punch–through voltage (V_{pch}) as a function of channel doping level (N_A) for various channel lengths (L) in the range from 0.05 to 0.2 microns. It can be seen that for physically short channel lengths, extremely high channel doping levels

are required to avoid punch– through. For example a channel doping level of 5.5×10^{17} atoms/cm^3 is required to avoid punch– through for a 0.1 micron channel length at drain voltages up to 2 volts.

Once punched– through, a space charge limited current flows from the source to drain. This current, I_{pch} is a function of drain voltage and is modelled by equation (2.15), from which it can be seen that there is no dependence upon gate voltage. For a device operating in punch– through, the space charge current flows in parallel to any conventional channel current.

$$I_{pch} = \frac{9\epsilon_s \mu_n r_j W (V_d - V_{pch})^2}{8L^3} \qquad \text{(const. mobility)} \qquad (2.15)$$

A device operating in punch– through can not be turned off and therefore is unsuitable for digital applications. Furthermore, the high output conductance (g_d) resulting from punch– through, seriously degrades the open circuit voltage gain (A_v), also rendering the devices unsuitable for analogue applications.

In severe cases, devices are punched through at drain voltages of zero volts i.e. V_{pch} is equal to zero. This latter condition results in an I– V output characteristic as shown in figure (2.11).

2.4 Scaling strategies

Before considering scaling strategies, it should be noted that the channel length (L) of a self– aligned MOSFET is usually shorter than the physical gate length (L_g) due to lateral encroachment of the drift regions under the gate. This is caused by lateral straggle during implantation and lateral diffusion during any subsequent high temperature anneal steps. For scaling considerations, the channel length (L) term is of greater importance.

Short channel effects and the severe effect of punch– through can be avoided in physically small devices by suitably scaling all device parameters. Several important approaches to device scaling have been reported recently[2.4,2.5,2.6,2.7], although as yet they have

not been verified experimentally for devices with sub 0.5 micron channel lengths. Each approach is described below, with a view to channel length scaling into the 0.5 to 0.1 micron region.

2.4.1 Constant Electric Field (CE) Scaling.

In order to maintain electrical long channel behaviour in a device with a physically short channel length, Dennard et al.[2.4] proposed a simple scaling strategy as introduced briefly in chapter 1. All the device parameters for a large reference device are scaled according to the precise scaling rules of table (2−1). These rules have been formulated such that the scaled device will have identical electric field distributions to the reference device, but which are scaled in co−ordinate space. Assuming, that the reference device exhibits long channel behaviour, by applying this constant electric field scaling approach, termed **CE scaling**, the scaled device is predicted to exhibit similar long channel behaviour.

Once the scaling coefficient **1/a** has been chosen, it can be seen from table (2−1) that channel doping is increased by **a** in order to reduce depletion widths and that gate oxide thickness and source/drain junction depths (r_j) are reduced by **1/a**.

Several points should be noted regarding this approach to scaling. Firstly, since the positive supply voltage (V_{DD}) is scaled by **1/a**, the threshold voltage (V_{th}) should be similarly scaled. However, from the right hand term of equation (2.1), it can be seen that for correct scaling, a band bending term ($\Psi_{bending}$) which is equal to ($2\Psi_B - V_b$) should also be scaled by **1/a**. Unfortunately the Ψ_B term remains relatively unchanged during scaling and so compensation is required through an adjustment of the substrate voltage (V_b) which normally has a negative or zero value. However, the substrate voltage (V_b) can not be increased to a positive value as this would lead to leakage of the source/drain junctions. Therefore, an attempt to scale the band bending term ($\Psi_{bending}$) is possible only if the reference device has a negative substrate voltage and even then the full **1/a** scaling may not be possible. Inadequate scaling of this term ($\Psi_{bending}$) will result in a disproportionately high threshold voltage (V_{th}). The additional three terms in equation (2.1) Φ_{ms}, $2\Psi_B$ and

$N_{ss}q/C_{ox}$ lead to a further complication in threshold voltage scaling.

Another problem with CE scaling occurs because the subthreshold current does not scale correctly since it is due to a carrier diffusion as opposed to a carrier drift mechanism. This can be seen by equation (2.7) since the ratio of C_D to C_{ox} remains the same for both reference and scaled devices, resulting in a constant subthreshold swing (S_w). For correct subthreshold scaling, this value should also be scaled by a factor 1/a. Further reference to equation (2.7) suggests that the desired scaling of the subthreshold swing (S_w) can be achieved by reducing the absolute temperature (T) at which the devices operate. Accordingly, low temperature operation by cooling devices to 77K has been proposed and recently investigated[2.8,2.9,2.10,2.11,2.12]. This approach has a further benefit in that an improved electron mobility results at low temperature. However, if low temperature is not possible, the unscaled subthreshold performance must be incorporated into the circuit design compromise at the expense of other aspects of performance e.g. circuit switching speed.

It should also be noted that by increasing the channel doping level, as required for CE scaling, the electron mobility is reduced, in turn slightly degrading the performance of the scaled devices. When the channel doping is increased from $1x10^{16}$ to $1x10^{17}$ atoms/cm3, the electron mobility decreases from approximately 1300 to $800cm^2/V-s$ [2.13].

The CE scaling approach rapidly reaches scaling limits when applied to channel length scaling at dimensions below 0.5 microns. From circuit design considerations for n−MOS digital circuits, it is desirable to use a threshold voltage (V_{th}) for the enhancement devices such that $V_{th}= V_{DD}/4$ where V_{DD} is the positive supply voltage. From consideration of device noise margin and processing tolerances a minimum threshold voltage of 0.2 volts is considered acceptable[2.5]. This corresponds to a positive supply voltage of 0.8 volts.

The CE scaling approach is used here to demonstrate how the 1.3 micron channel length L.A.T.V. process[2.14,2.15] developed at I.B.M. in the mid 1970s can be used to predict a device with a 0.4 micron channel length. The device parameters for the L.A.T.V. device and the scaled 0.4 micron device are shown in table (2−2). The gate oxide thickness is reduced from 250 angstroms to 80

angstroms and the souce/drain junction depths are reduced from 0.35 microns to 0.12 microns. The peak channel doping is increased from $4x10^{16}$ to $1.2x10^{17}$ atoms/cm^3 and operating voltages are suitably reduced to the limit indicated above. A zero volts substrate bias is chosen to replace the L.A.T.V. value of -1 volts in an attempt to scale band bending terms, although a ratio of only 2.25 is achieved. The scaled 0.4 micron channel length device will be referred to as example **CE1**.

Whilst other reference devices could have been used, the L.A.T.V. process gives a good indication of the limit to which this scaling approach can be applied since the reference device has itself a relatively small channel length together with a relatively large supply voltage.

The problem of operating voltage limits can be overcome if an alternative approach to scaling is chosen whereby the electric field strengths are allowed to increase whilst preserving the same field patterns.

2.4.2 Constant Electric Field Pattern (CP) Scaling

Baccarani et al.[2.5] reported a scaling strategy whereby the operating voltages are scaled less strongly with channel length. This approach permits further scaling than the approach by Dennard et al[2.4] because the threshold voltage limit of 0.2 volts no longer restricts scaling.

Device parameters are scaled according to the scaling rules shown in table(2−3). For this approach, the channel length is scaled by a factor **1/a** and the operating voltages are independently scaled by **1/b**. The electric field strengths are therefore scaled by **a/b** and it can be seen that when **a** is equal to **b**, the CE scaling case results.

This scaling approach, termed **CP scaling**, ensures that the scaled device has similar field patterns as for the reference device but allows the field intensities to increase. Since the scaled device retains the field pattern of a long channel device, it should also exhibit long channel behaviour.

Having removed the operating voltage constraint which limited the CE scaling approach, a further degree of channel length scaling is

possible until two further limits are reached. Firstly, scaling of the source/drain junction depths becomes a problem since the drift regions are required to be both extremely shallow and of very high conductivity. Whilst junction depths of approximately 0.05 microns can be achieved by low energy arsenic implantation, these regions have very high sheet resistivities, even for high dose implants. In practice a junction depth of 0.07 microns can be achieved with an acceptably low sheet resistivity of approximately 100 ohms/square[2.16].

A second limit to scaling occurs when the gate oxide is so thin that carrier tunnelling though the oxide becomes significant. Tunnelling effects become rapidly significant for oxide thicknesses of less than 50 angstroms, with an exponential dependence upon oxide thickness. The tunnelling current is proportional to a tunnelling probability term (P_T) as defined by equation (2.16) for a rectangular barrier[2.17] with units as indicated. The term α_T is a constant which approaches unity as the electron effective mass in the insulator approaches the free electron mass. Assuming an α_T value of 1 and a silicon dioxide barrier height of 3.1 volts, the resulting tunnelling probability increases by a factor of $4x10^7$ each time the gate oxide thickness is reduced by 10 angstroms.

$$P_T \approx \exp(\alpha_T O_T^{1/2} t_{ox}) \qquad (2.16)$$

When:

P_T = tunnelling probability
O_T = rectangular barrier height (in volts)
α_T = constant related to electron effective mass
t_{ox} = oxide thickness (in angstroms)

Once again, the 1.3 micron L.A.T.V. process is used as a reference technology for predicting several scaled device examples as shown in table (2−4). Example CP1 is a scaled 0.4 micron device, similar to the example CE1 except that the positive suppy voltage (V_{DD}) has been increased to 2 volts. The scaling factors a and b are equal to 3.25 and 1.25 respectively. This requires a substrate voltage of −0.6 volts and a channel doping level of $3x10^{17}$

atoms/cm^3, with all other parameters as for CE1.

Example CP2 predicts a 0.3 micron device with an operating voltage of 1 volt. For this example a is equal to 4.3 and b is equal to 2.5. A channel doping level of $3x10^{17}$ atoms/cm^3 is required together with a 60 angstrom gate oxide. A 0.08 micron source/drain junction depth is required together with a zero volts substrate bias. This junction depth is close to the processing limit identified above and therefore represents the practical limit for CP scaling.

Finally, example CP3 predicts a 0.2 micron device with an operating voltage of 1 volt. For this example a and b are equal to 6.5 and 2.5 respectively, requiring a channel doping level of $6x10^{17}$ atoms/cm^3 and a 40 angstrom gate oxide. Also, a 0.05 micron junction depth is required together with a substrate bias of zero volts. Whilst the junction depth required is below the practical limit identified above, this example has been included for the purpose of extending CP scaling to the point where oxide tunnelling effects are significant.

It should be noted that whilst CP scaling can be used to predict a desirable long channel behaviour by retaining long channel field patterns, several second order effects may occur due to the increase in electric field intensities. These second order high field effects are discussed further in section 2.5.

2.4.3 Generalised Long Channel (GL) Scaling

Brews et al.[2.6] observed that a simple empirical equation can be used to predict whether a given device will exhibit long channel or short channel behaviour. This equation uses various device parameters to calculate the minimum channel length (L_{min}) for which long channel behaviour is predicted, and is shown below as equation (2.17). Alternatively, any device with a smaller channel length is predicted to display short channel behaviour. The terms X_s and X_d represent the source and drain depletion thicknesses and are defined by equations (2.18i) and (2.18ii) respectively.

$$L_{min} = 0.4[r_j t_{ox}(X_s+X_d)^2]^{1/3} \qquad (2.17)$$

When:

L_{min}	= minimum L for long channel behaviour	(in um)
r_j	= source/drain junction depth	(in um)
T_{ox}	= gate oxide thickness	(in Å)
X_s	= source depletion thickness	(in um)
X_d	= drain depletion thickness	(in um)

For $X_s = (2\epsilon_s[V_{bi}-V_b]/qN_A)^{1/2}$ (2.18i)

and $X_d = (2\epsilon_s[V_d+V_{bi}-V_b]/qN_A)^{1/2}$ (2.18ii)

Figure (2−12) shows curves of L_{min}, calculated using equation (2.17), as a function of channel doping level (N_A) for various applied drain voltages (V_d). An oxide thickness of 150 angstroms has been used together with a junction depth (r_j) of 0.1 microns. The curves correspond to applied drain voltages (V_d) of 1, 2, 3, 4 and 5 volts.

Generalised long channel scaling permits some device parameters to be scaled strongly whilst others are scaled more gradually in order to comply with technology and device limits. However, as for CP scaling, no account is taken of the strength of the electric fields within the device and so care must be taken to avoid the second order effects of increased field strengths as discussed later in section 2.5. Whilst Brews et al. have demonstrated the accuracy of their empirical relationship for channel lengths of one micron and greater, it is used here, speculatively, to predict devices with sub half micron channel lengths.

Since the empirical equation takes no account of threshold voltage (V_{th}), this must be considered independently. A suitable approach, termed **GL scaling**, is described below for three example devices.

For each of the three devices, a junction depth (r_j) of 0.1 microns was chosen together with a substrate voltage (V_b) of zero volts and a positive supply voltage (V_{DD}) of 2 volts. The digital circuit design consideration that $V_{th}= V_{DD}/4$ was then applied to obtain a threshold voltage of approximately 0.5 volts. Next, gate

oxide thicknesses of 150, 80, and 40 angstroms were chosen for the three devices which are referred to as examples **GL1**, **GL2** and **GL3** respectively. Using equation (2.1) or the graphical representation in figure (2−6), it was found that the required threshold voltage is predicted for channel doping levels of approximately $1.2x10^{17}$, $4x10^{17}$ and $1.2x10^{18}$ atoms/cm^3 respectively. By setting the drain voltage (V_d) in equation (2.18ii) equal to the positive supply voltage (V_{DD}) of 2 volts, the minimum channel length was then determined for each device. The construction in figure (2−12) shows this derivation of L_{min} for the GL1 example. Accordingly, L_{min} values of 0.40, 0.23 and 0.12 microns were obtained for the GL1, GL2 and GL3 device examples. Allowing for a small margin of error, nominal channel lengths (L) of 0.45, 0.25 and 0.15 microns are predicted. The parameters for each of the devices are summarised in table (2−5).

2.4.4 Other Approaches to Scaling

Whilst emphasis has been placed up till this point upon scaling to maintain long channel behaviour, for certain applications it may be preferable to scale using other criteria. As indicated earlier, subthreshold swing (S_w) does not scale with other parameters, resulting in a degraded turn on characteristic which may not be acceptable for certain digital applications. In such cases, scaling to preserve constant subthreshold current can be used, as reported by Sokel[2.7]. Optimum circuit performance may then be achieved with devices which exhibit short channel behaviour.

2.5 High Electric Field Effects

When the electric fields within a MOSFET are permitted to increase beyond normal levels, a number of second order effects begin to modify the behaviour of the device[2.18]. These effects can be considered separately as those caused by the longitudinal field components (E_y) and those by the transverse field components (E_x).

2.5.1 Longitudinal Field (E_y) Effects

High longitudinal fields lead to velocity saturation of the channel carriers and in severe cases hot carrier phenomena result.

2.5.1.1 Velocity Saturation

As the longitudinal field (E_y) increases, the channel carriers eventually reach saturation velocity. It should be noted that carrier velocity saturation and drain current saturation are different phenomena. As discussed earlier, current saturation is caused by device pinch–off and under conditions of constant mobility leads to a quadratic relationship between saturation current and $(V_g - V_{th})$. However, under conditions of carrier velocity saturation, the drain saturation current and transconductance are reduced. The saturation current (I_{dsat}) then varies linearly with $(V_g - V_{th})$ as modelled by equation (2.19) and shown in figure (2–13). Consequently, the device transconductance becomes independent of gate voltage and is modelled by equation (2.20).

$$I_{dsat} = WC_{ox}(V_g - V_{th})v_s \qquad \text{(velocity sat.)} \qquad (2.19)$$

$$g_m = WC_{ox}v_s \qquad \text{(velocity sat.)} \qquad (2.20)$$

When v_s = carrier saturation velocity.

For devices operating in punch–through, under conditions of velocity saturation the space charge limited punch–through current is modelled by equation (2.21) rather than equation (2.15) which only applies for the constant mobility case.

$$I_{pch} = \frac{2\epsilon_s v_s r_j W(V_d - V_{pch})}{L^2} \qquad \text{(velocity sat.)} \qquad (2.21)$$

2.5.1.2 Hot Electron Phenomena

When the longitudinal fields become extremely large in the drain pinch–off region, electrons are able to gain sufficient energy to cause impact ionisation. The resulting electrons contribute to an increased drain current whilst the holes travel down to the substrate contact leading to a significant substrate current. This substrate current in conjunction with the substrate resistance leads to a lowering of the source junction barrier height which further increases the channel current thereby leading to more impact ionisation[2.19].

Another significant hot carrier effect results from electrons gaining sufficient energy to surmount the energy barrier at the silicon to gate oxide interface which is approximately equal to 3.1 eV. These electrons either become trapped in the oxide or pass through thereby contributing to the gate current. The trapped electrons increase the oxide fixed charge and therefore lead to long term threshold voltage instability[2.20].

2.5.2 Transverse Field (E_x) Effects

High transverse electric fields lead to a degradation of both carrier mobility and saturation velocity. In addition, high fields in the oxide may lead to catastrophic oxide breakdown.

2.5.2.1 Mobility and Saturation Velocity Degradation

Both mobility and saturation velocity are degraded within the thin inversion layer of a MOSFET due to the high transverse fields which are required to establish carrier inversion. For the scaled device, high channel doping levels result in even higher transverse fields and the degradation effects become more significant. For example, when the average transverse field increases from 1×10^5 to

$5x10^5$ V/cm, the inversion layer electron mobility decreases from approximately 900 to 670 $cm^2/V-s$. Over the same field range the saturation velocity decreases from approximately $9x10^6$ to $5x10^6$cm/s [2.21].

2.5.2.2 Gate Oxide Breakdown

An additional problem arises from the increased electric fields in the gate oxide which are required to establish inversion for higher channel doping levels. Generally, oxide films thicker than 1000 angstroms can support field strengths of up to $9x10^6$ V/cm, with catastrophic dielectric breakdown resulting above this limit. However, it has been shown experimentally that the maximum field strength increases for thicknesses below 1000 angstroms. For example, an 80 angstrom oxide can support a maximum field of approximately $2x10^7$ V/cm [2.22,2.23,2.24]. This is convenient for sub 0.5 micron device scaling where oxide thicknesses in the range 150 to 40 angstroms are proposed.

The effects of mobility and saturation velocity degradation lead only to a reduced enhancement of device performance with scaling. However the effects of hot carrier oxide charging and oxide oxide breakdown are more serious, leading respectively to threshold voltage instability and device failure.

2.6 Summary

A simple description of MOSFET operation has been given. The short channel effects which result from inadequate scaling have been described. Several alternative scaling strategies have been introduced which attempt to preserve long channel behaviour for physically small devices. Finally, a qualitative description has been given of the high field effects which may degrade the performance of scaled MOSFETs.

References

2.1 S.M. Sze, 'The Physics of Semiconductor Devices', Wiley, New York, 1981, 2nd ed.

2.2 W.M. Werner, 'The work function difference of the MOS–system with aluminium field plates and polycrystalline silicon field plates,' Solid State Electron., vol 17, p769, 1974.

2.3 L.D. Yau, 'A simple theory to predict the threshold voltage of short channel IGFETs,' Solid State Electron., vol. 17, p1059, 1974.

2.4 R.H. Dennard, F.H. Gaensslen, H. Yu, V.L. Rideout, E. Bassous and A.R. LeBlanc, 'Design of ion–implanted MOSFETs with very small physical dimensions,' IEEE J. Solid–State Circuits, vol. SC–9, p256, 1974.

2.5 G. Baccarani, M.R. Wordeman, R. H. Dennard, 'Generalized scaling theory and its application to a 1/4 micrometer MOSFET design,' IEEE Trans. Electron Devices, vol. ED–31, p452, 1984.

2.6 J.R. Brews, W. Fichtner, E.H. Nicollian, S.M. Sze, 'Generalized guide for MOSFET miniaturization,' IEEE Electron Device Lett., vol EDL–1, p2, 1980.

2.7 R. Sokel, 'Transistor scaling with constant subthreshold leakage,' IEEE Electron Device Lett., vol EDL–4, p85, 1983.

2.8 A. Kamgar, 'Miniaturization of Si MOSFETs at 77K,' IEEE Trans Electron Devices, vol ED–29, p1226, 1982.

2.9 G. Gildenblat, L. Colonna–Romano, D. Lau and D.E. Nelson, 'Investigation of cryogenic CMOS performance,' IEDM Tech. Dig., p268, Washington, 1985.

2.10 J.C.S. Woo and J.D. Plummer, 'Short–channel effects in MOSFETs at liquid–nitrogen temperature,' IEEE Trans. Electron Devices, vol ED–33, p1012, 1986.

2.11 J.Y.–C. Sun, Y. Taur, R.H. Dennard, and S.P. Klepner, ' Submicrometer–channel CMOS for low–temperature operation,' IEEE Trans. Electron Devices, vol ED–34, p19, 1987.

2.12 J.T. Watt, B.J. Fishbein and J.D. Plummer, 'Low temperature NMOS with caesium–implanted load devices,' IEEE Trans. Electron Devices, vol ED–34, p28, 1987.

2.13 C. Jacoboni, C. Canali, G. Ottaviani and A.A. Quaranta, 'A review of some charge transport properties of silicon,' Solid State Electron., vol 20, p77, 1977.

2.14 H.N. Yu, A. Riesman, C.M Osburn and D.L. Critchlow, '1–um MOSFET VLSI technology: Part I – An overview,' IEEE Trans. Electron Devices, vol ED–26, p318,1979.

2.15 R.H. Dennard, F.H. Gaensslen, E.J. Walker and P.W. Cook, '1–um MOSFET VLSI technology: Part II – Device designs and characteristics for high–performance logic applications,' IEEE J. Solid–St. Circuits, vol SC–14, p247, 1979.

2.16 C.M. Osburn and M.Y. Tsai, S. Roberts, C.J. Lucchese and C.Y. Ting, 'High conductivity diffusions and gate regions using self aligned silicide technology,' presented at Electrochem. Soc. 1st Int. Symp. on VLSI (Detroit,MI),1982.

2.17 V. Kumar and W.E. Dahlke, 'Characteristics of Cr–SiO_2–nSi tunnel diodes,' Solid State Electron., vol 20, p 143, 1977.

2.18 E. Sangiori, E.A. Hofstatt, R.K. Smith, P.F. Bechtold and W. Fitchner, 'Scaling issues related to high field phenomena in submicrometer MOSFETs,' IEEE Electron Dev. Lett., vol EDL–7, p115, 1986.

2.19 T. Kamata, K. Tanabashi and K. Kobayashi, 'Substrate current due to impact ionization in MOS–FET,' Jpn. J. Appl. Phys., vol 15, p1127, 1976.

2.20 E. Sun, J. Moll, J. Berger and B. Alders,'Breakdown mechanism in short−channel MOS transistors,' IEEE Tech. Dig. Electron Device Meet., Washington D.C., p478, 1978.

2.21 R.W. Coen and R.S. Muller, 'Velocity of surface carriers in inversion layers on silicon,' Solid State Electron., Vol 23, p35, 1980.

2.22 C.M. Osburn and D.W. Ormond, 'Dielectric breakdown in silicon dioxide films on silicon,' Solid State Sci. Technol., vol 119, p581, 1972.

2.23 I.−C. Chen, S.E. Holland and C. Hu, 'Electrical breakdown in thin gate and tunneling oxides,' IEEE Trans. Electron Devices, vol ED−32, p413, 1985.

2.24 Y. Hokari, T. Baba and N. Kawamura, 'Reliability of 6−10 nm thermal SiO_2 showing intrinsic dielectric integrity,' IEEE Trans. Electron Devices, vol ED−32, p2485, 1985.

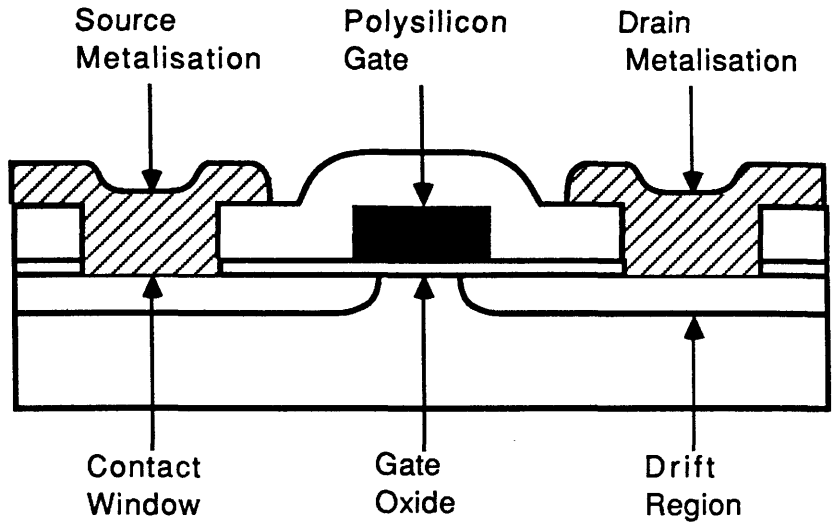


Figure (2-1) MOSFET in cross-section.

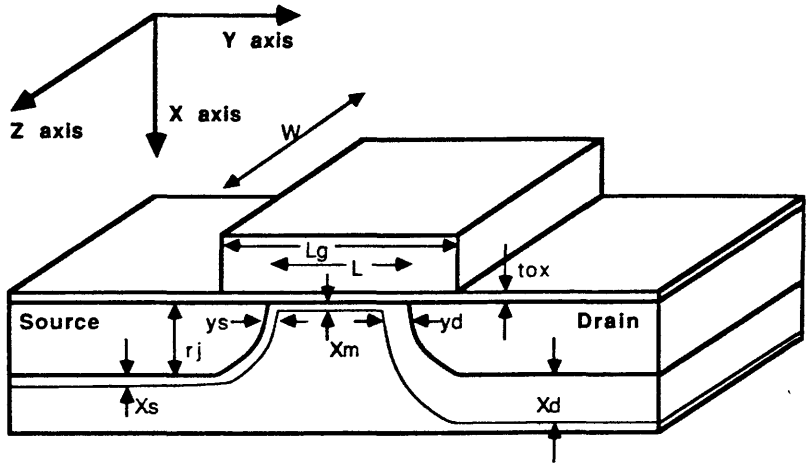
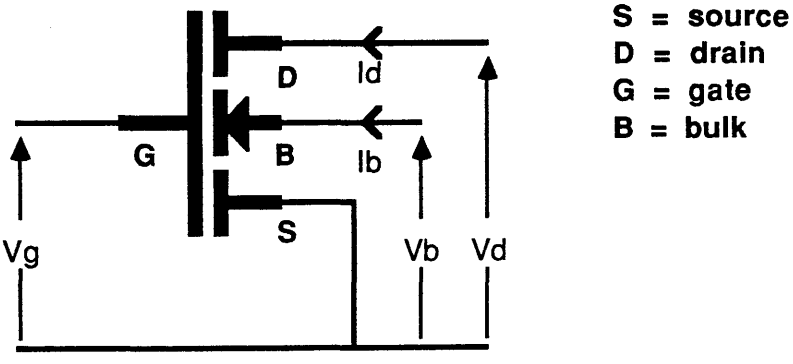
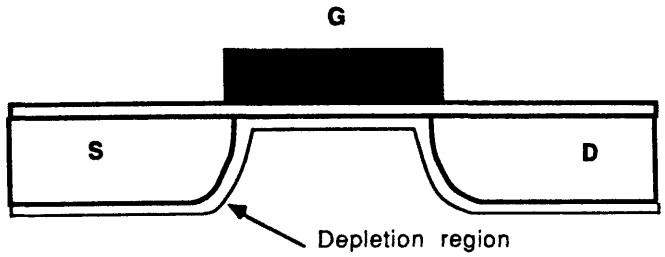


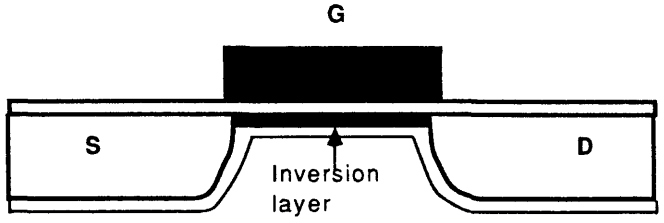
Figure (2-2) MOSFET with labelled device dimensions.



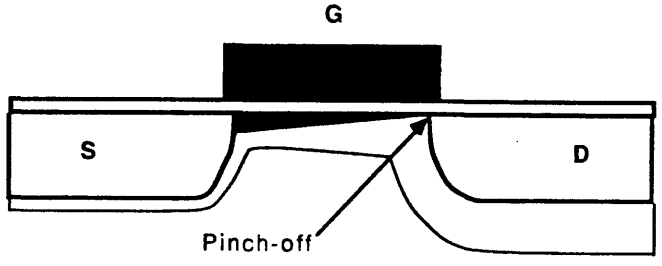
Figure(2-3) N-type enhancement mode MOSFET showing the applied voltage notation which is used throughout the thesis.



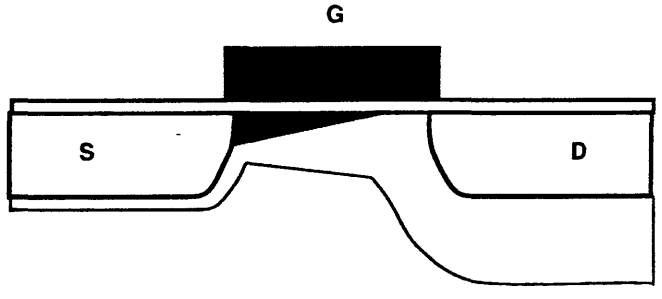
(a) Below threshold.



(b) Above threshold.



(c) At pinch-off.



(d) Beyond pinch-off.

Figure (2-4) A set of diagrams showing a MOSFET under various operating conditions.

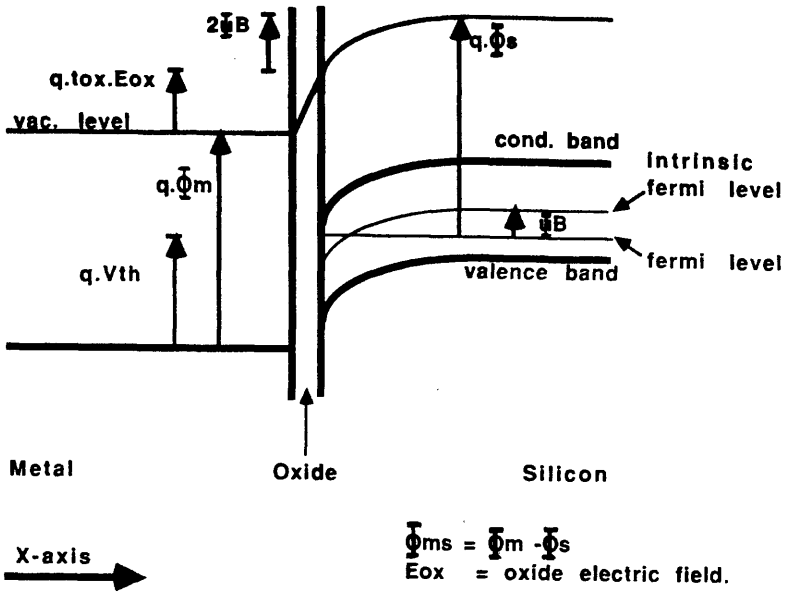


Figure (2-5) Energy band diagram of a MOS structure at the carrier inversion threshold.

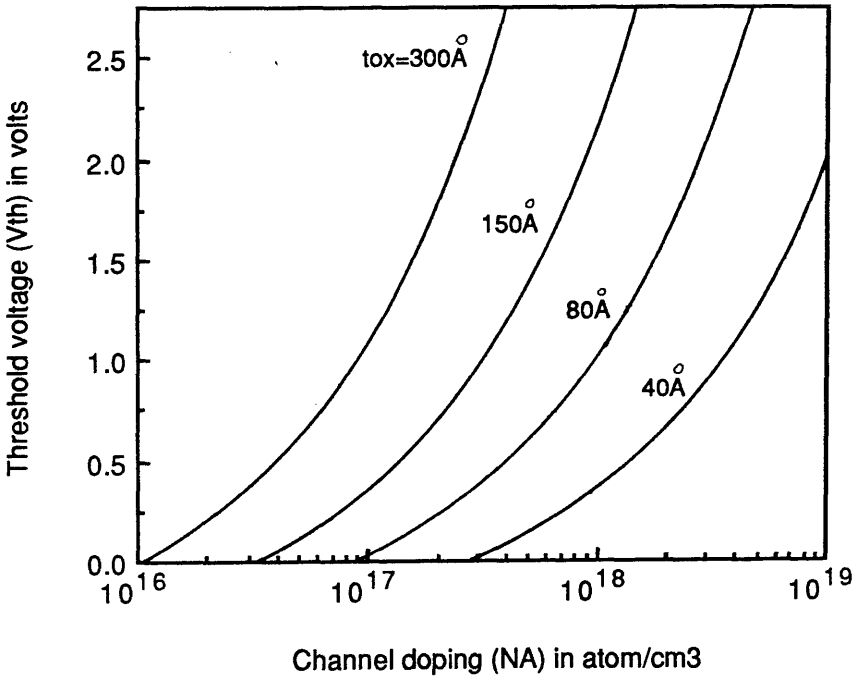
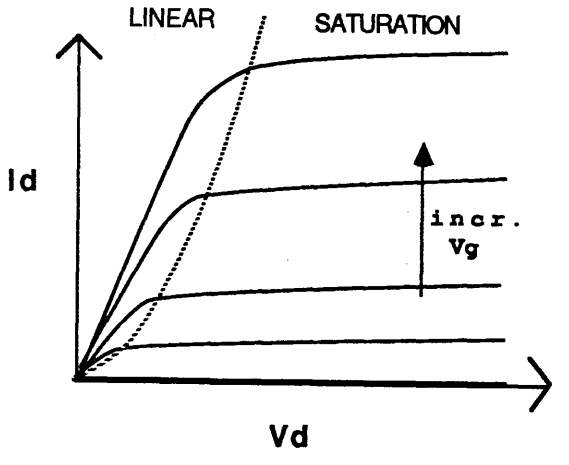
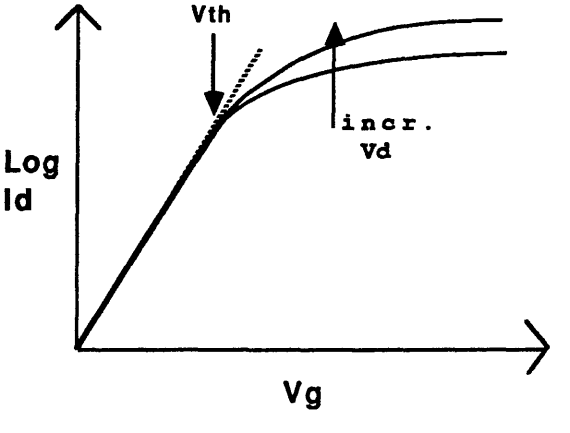


Figure (2-6) Graph of threshold voltage (V_{th}) versus channel doping (N_A) for a range of oxide thicknesses (t_{ox}) from 300 to 40Å.



(a) I-V output characteristic.



(b) Subthreshold characteristic.

Figure (2-7) Typical electrical characteristics of a MOSFET.

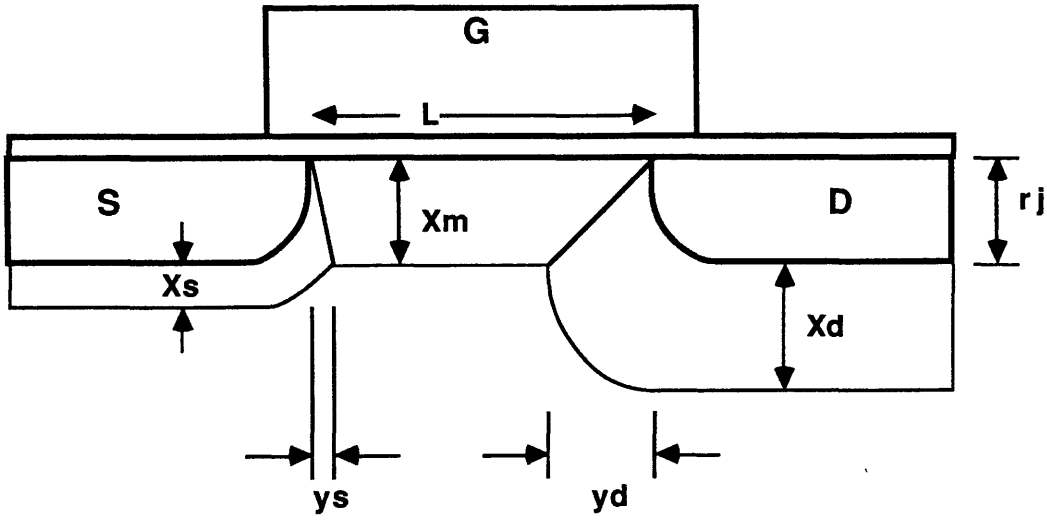
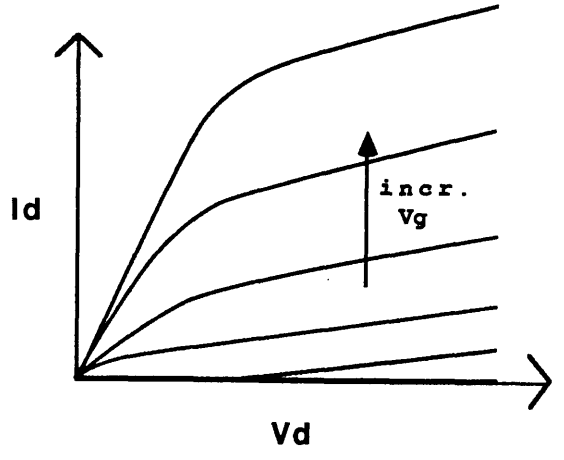
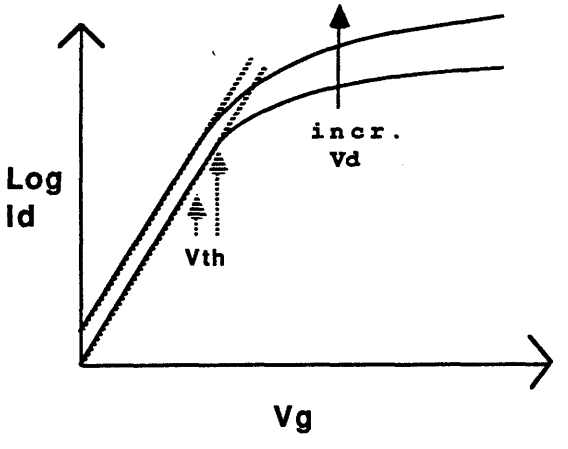


Figure (2-8) Illustration of trapezoidal substrate charge sharing model for short channel MOSFETs.



(a) I-V output characteristic



(b) Subthreshold characteristic

Figure (2-9) Typical electrical characteristics for a MOSFET displaying short channel effects.

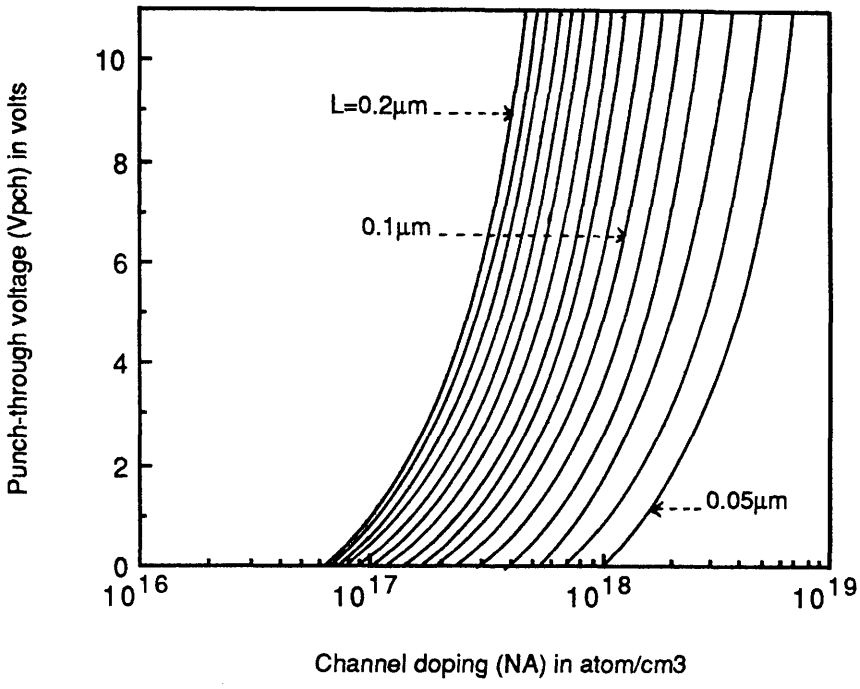


Figure (2-10) Graph showing the punch-through voltage (V_{pch}) as a function of channel doping (N_A), for channel lengths in the range 0.2 to 0.05 um (from equation (2.14)).

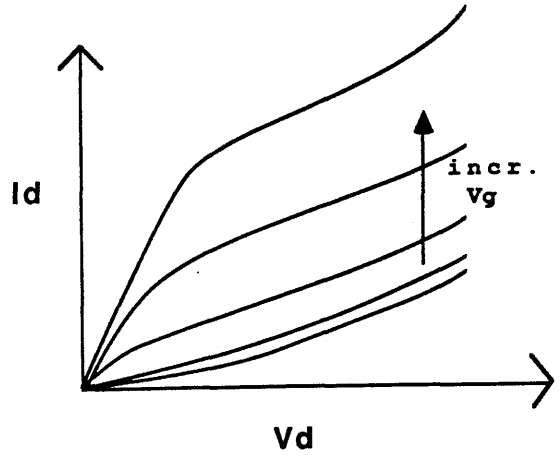


Figure (2-11) Typical I-V output characteristic for a punched-through MOSFET.

Dimension	Scaling Factor
L	1/a
t_{ox}	1/a
r_j	1/a
N_A	a
V_{DD}	1/a
V_{th}	1/a
$\Psi_{(bending)}$	1/a

Table (2-1) Showing MOSFET scaling rules for constant electric field scaling (CE scaling).

Dimension	LATV	CE1	a
L	1.3um	0.4um	3.25
t_{ox}	250Å	80Å	3.125
r_j	0.35um	0.12um	3
N_A	$4x10^{16}atm/cm^3$	$1.2x10^{17}atm/cm^3$	3
V_{DD}	2.5V	0.8V	3.125
V_{th}	0.6V	0.2 V	3
V_b	-1V	0V	
$\Psi_{(bending)}$	1.8V	0.8V	2.25

Table (2-2) Showing the application of constant electric field (CE) scaling to predict a MOSFET with a 0.4um channel length (CE1) using a reference LATV device.

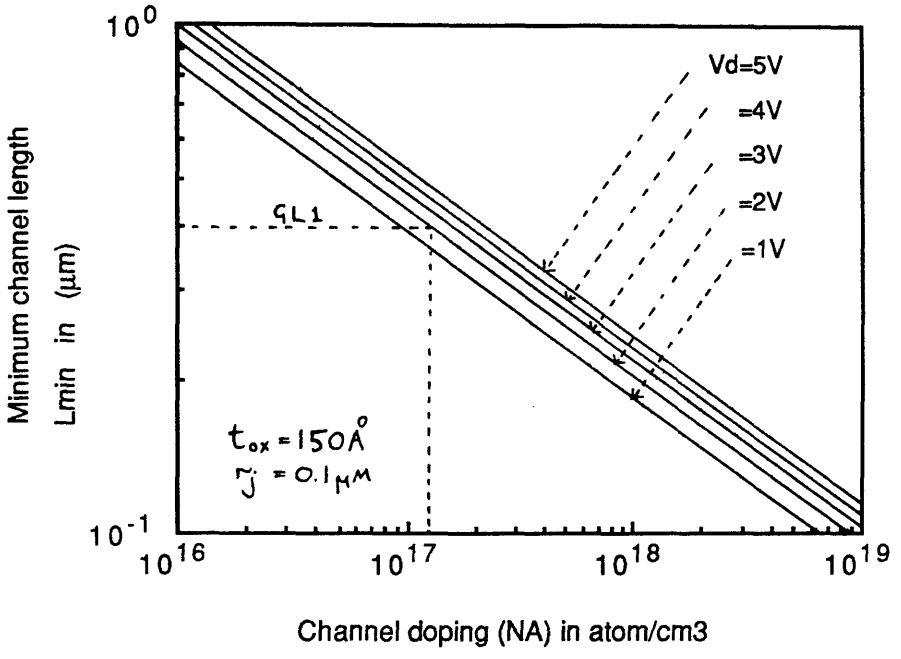


Figure (2-12) Graph showing L_{min} as a function of channel doping (N_A) for various applied drain voltages (V_d) (from equation (2.17)).

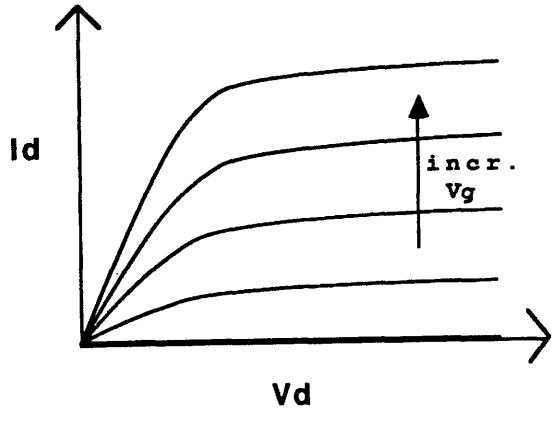


Figure (2-13) Typical I-V output characteristic for a MOSFET under conditions of carrier velocity saturation.

Dimension	Scaling Factor
L	1/a
t_{ox}	1/a
r_j	1/a
E Fields	a/b
N_A	a^2/b
V_{DD}	1/b
V_{th}	1/b
Ψ(bending)	1/b

Table (2-3) Showing MOSFET scaling rules for constant field pattern scaling (CP scaling).

Dimension		LATV	CP1	CP2	CP3
L	(um)	1.3	0.4	0.3	0.2
a		1	3.25	4.3	6.5
b		1	1.25	2.5	2.5
t_{ox}	(Å)	250	80	60	40
r_j	(um)	0.35	0.12	0.08	0.06
N_A	(atm/cm^3)	$4x10^{16}$	$3x10^{17}$	$3x10^{17}$	$6x10^{17}$
V_{DD}	(V)	2.5	2	1	1
V_{th}	(V)	0.6	0.5	0.25	0.25
V_b	(V)	-1	-0.6	0	0
Ψ(bending)	(V)	1.8	1.4	0.8	0.8

Table (2-4) Showing three predicted devices obtained using constant field pattern (CP) scaling and a LATV reference device. The examples, CP1 CP2 and CP3 are scaled MOSFETs with 0.4, 0.3 and 0.2 um channel lengths.

Parameter	GS1	GS2	GS3
L (um)	0.45	0.25	0.15
t_{ox} (Å)	150	80	40
r_j (um)	0.1	0.1	0.1
N_A (atm/cm^3)	$1.2x10^{17}$	$4x10^{17}$	$1.2x10^{18}$
L_{min} (um) for V_{DD}=2V & V_b=0V	0.40	0.23	0.12
V_{th} (V)	≈0.5	≈0.5	≈0.5

Table(2-5) Table shows three predicted MOSFETs obtained using generalised long channel (GL) scaling, for $V_{th} = V_{DD}/4 = 0.5V$.

Chapter 3

Objectives and Experimental Approach

3.1 Introduction

This chapter describes the objectives of the study and a suitable experimental approach. A modified 3 micron n−MOS process, which incorporates a high resolution electron beam lithography step has been chosen to investigate MOSFET scaling.

3.2 Scaling for Future VLSI Applications

As indicated earlier, it is likely that electron beam lithography and X−ray printing machines will soon permit the exploitation of VLSI circuits with 0.5 to 0.1 micron design rules. In the case of n−MOS technology, considerable research into the optimisation of device performance and processing technology is required at these dimensions.

To date, the scaling theories of chapter 2 have been used to implement MOS technologies with design rules down to the 1 micron level and in some cases the 0.5 micron level[3.1]. However, no serious attempt has been made to confirm, by experiment, the scaling predictions for technologies at smaller dimensions.

Full design rule scaling to the 0.5 to 0.1 micron level would require scaling of channel length (L) and width (W), together with scaling of isolation widths, contact window dimensions and interconnect metalisation linewidths. Such an approach will require an extremely extensive research programme and therefore was considered far too ambitious for this study. Instead, a limited approach to scaling has been chosen whereby channel region scaling is confined to the X−axis and the Y−axis directions. For this approach, channel length (L) and the oxide thickness (t_{ox}) are reduced and channel doping level (N_A) is increased. However, other features such as channel width (W), contact window dimensions and interconnect linewidths remain unchanged, permitting extensive use of conventional optical lithography

and standard n−NOS processing techniques. Consequently, high resolution electron beam lithography and dry etching techniques were required only at the critical 0.5 to 0.1 micron gate patterning step.

This limited approach to scaling can be used to investigate the scaling predictions of chapter 2, by varying the appropriate device parameters. Predictions of long channel and short channel behaviour can then be investigated for very short physical channel lengths.

To date, several research groups worldwide[3.2,3.3,3.4,3.5,3.6,3.7] have fabricated experimental MOSFETs with channel lengths in this range, using similar limited scaling approaches. However, in most cases, little attention has been paid to scaling theory and consequently the electrical characteristics of the devices have been poor.

Also, experimental MOSFETs have been fabricated recently with extremely narrow channel widths using electron beam lithography[3.8].

3.3 Objectives of the Study

The objective of this study has been to fabricate MOSFETs with gate lengths in the range 0.5 to 0.1 microns. Since short channel effects are known to degrade device performance, the experimental devices have been designed, where possible, in accordance with MOSFET scaling theory. The resulting devices were tested and evaluated with the following aims:

(1) To determine whether short channel effects can be suppressed in physically small MOSFETs through the use of scaling theory.

(2) To investigate the relative improvement in MOSFET performance which might be possible through scaling to these dimensions.

(3) To investigate quantitatively the second order effects which are known to result from device scaling. These effects include mobility and saturation velocity degradation and hot electron phenomena as discussed qualitatively in chapter 2.

(4) To investigate the practical limit for MOSFET scaling and the factors which determine this limit. Directions for future research can

be identified, where the limits are set by current processing technology.

3.4 The Experimental Device Structure

Figure (3–1) shows a standard polysilicon gate MOSFET using 3 micron design rules. The source and drain drift regions are formed by self–aligned ion implantation using the gate as a mask. Device isolation is achieved using the LOCal Oxidation of Silicon (LOCOS) technique[3.9] whereby a thick field oxide is grown selectively. Silicon nitride is used to protect the active areas during oxidation.

Figure (3–2) shows the experimental device structure as chosen for this study. The structure of the device and consequently the processing techniques have been modified in several important respects.

The use of high resolution electron beam lithography and dry etching techniques to pattern 0.5 to 0.1 micron polysilicon gate electrodes imposes a constraint upon the device structure. Thin line patterning is often unreliable where the lines are exposed across large vertical step discontinuities. Such steps normally occur wherever the LOCOS field oxide meets a device active region. However, whilst a reduced field oxide thickness is required to improve the patterning of thin lines, it would normally lead to increased parasitic capacitance under the bond pads. Therefore, a two step LOCOS process was chosen to give separate control of field oxide thickness and bond pad oxide thickness. This permits a relatively small field oxide step to be achieved whilst retaining a low parasitic capacitance under the bond pads.

The process must also be modified to incorporate extremely shallow source and drain drift regions using a suitable high dose low energy arsenic implant. Any subsequent anneal step required to remove implantation damage and to activate this implant must not cause any significant dopant redistribution since this would lead to an increased drift region junction depth (r_j). In addition, it would lead to increased lateral encroachment under the gate and thereby reduce the effective channel length (L). A Rapid Thermal Anneal (RTA)[3.10] step was chosen to minimise this redistribution.

The shallow source and drain drift regions present an additional problem. When aluminium contacts are sintered into extremely shallow drift regions, aluminium diffuses through the junction and forms a junction short circuit. This phenomenom, termed junction spiking[3.11], would be particularly severe for 0.1 micron junction depths required here. The device structure was therefore modified to include an increased junction depth in the locality of contact windows. This could be achieved either by direct implantation through the contact windows or by implantation using a separate lithographic mask level.

The implementation of reduced gate oxide thickness (t_{ox}) and increased channel doping level (N_A) required no special structural modifications.

The development of suitable high resolution electron beam lithographic and dry etching techniques for patterning polysilicon linewidths down to 0.1 microns is described in chapter 4.

3.5 Summary

The objectives of the thesis have been identified together with a suitable experimental approach. The chapter concludes with a description of the experimental MOSFETs with sub 0.5 micron gate lengths.

References

3.1 M.R. Wordeman, A.M. Schweighart, R.H. Dennard, G. Sai–Halasz and W.W. Molzen, 'A fully scaled submicrometer nMOS technology using direct–write e–beam lithography,' IEEE Trans. Electron Devices, vol ED–32, p2214, 1985.

3.2 C.E. Binnie, 'The fabrication of small geometery MOSFETs using electron beam lithography,' Ph.D. Thesis, University of Glasgow, Feb 1985.

3.3 P. Parrens and F. Buiguez, 'Submicron device fabrication using e.beam masking and R.I.E.,' Proc. of Microcircuit Engineering 82, Grenoble, 1982.

3.4 I.P. Thomas, A.E. Glaccum, P. Hendy, B. Martin, C.J. Pollard, 'Sub-micron Si MOSFETs for analogue and digital applications,' presented at IEE colloquium on 'Device physics, technologies and modelling for the submicron era,' (p2/1), London, 1983.

3.5 W. Fichtner, R.K. Watts, D.B. Fraser, R.L. Johnson and S.M. Sze, '0.15um Channel-length MOSFETs, fabricated using e-beam lithography,' IEEE Electron. Device Lett., Vol EDL-3, p412, 1982.

3.6 R.E. Howard, L.D. Jackel, R.G. Swartz, P. Grabbe, V.D. Archer, R.W. Epworth, E.L. Hu, D.M. Tennant and A.M. Voshchenkov, 'Buried channel MOSFETs with gate lengths from 2.5um to 700A,' IEEE Electron Device Lett., Vol EDL-3, p332, 1982.

3.7 H.-S. Lee and L.C. Puzio, 'The electrical properties of subquarter-micrometer gate-length MOSFETs,' IEEE Electron Device Lett., vol EDL-7, 1986.

3.8 R.E. Howard, W.J. Skocpol, L.D. Jackel, P.M. Mankiewich, L.A. Fetter, D.M. Tennant, 'Single switch events in nanometer-scale Si MOSFETs,' IEEE Trans. Electron Devices, vol ED-32, p1669, 1985.

3.9 J.A. Appels, E. Kooi, M.M. Paffen, J.J.H. Schlorje and W.H.C.G. Verkuylen, 'Local oxidation of silicon and its application in semiconductor technology,' Philips Res. Rep., vol 25, p118, 1970.

3.10 A. Gat, 'Heat pulse annealing of arsenic-implanted silicon with a CW arc lamp,' IEEE Electron Device Lett., vol EDL-2, p85, 1981.

3.11 S.M. Sze (Ed), 'VLSI Technology,' p367, McGraw-Hill, Singapore, 1983.

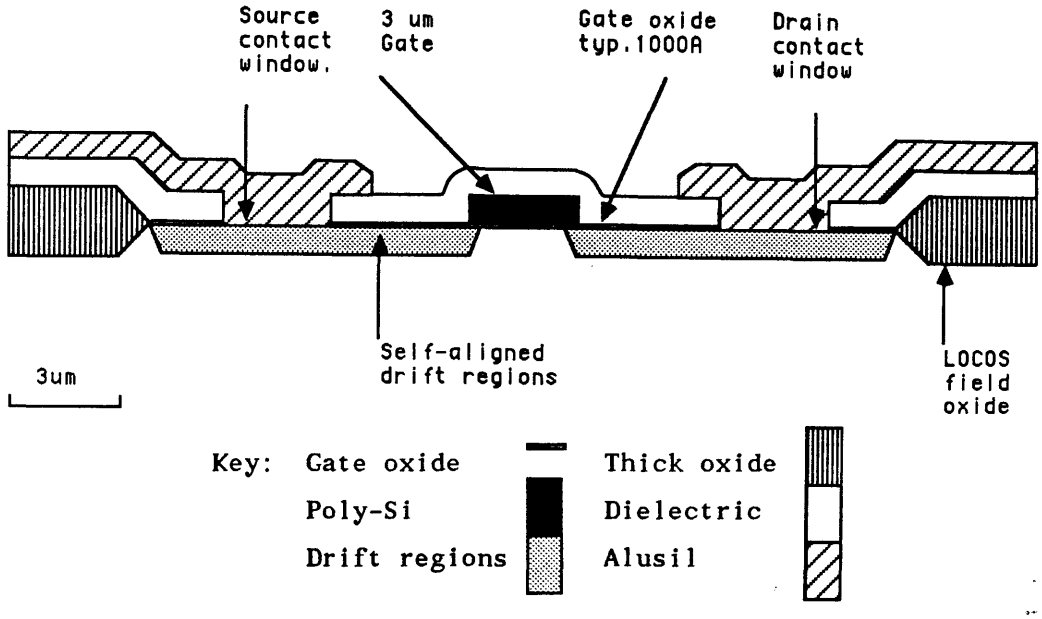


Figure (3-1) A standard polysilicon gate MOSFET with 3 micron design rules.

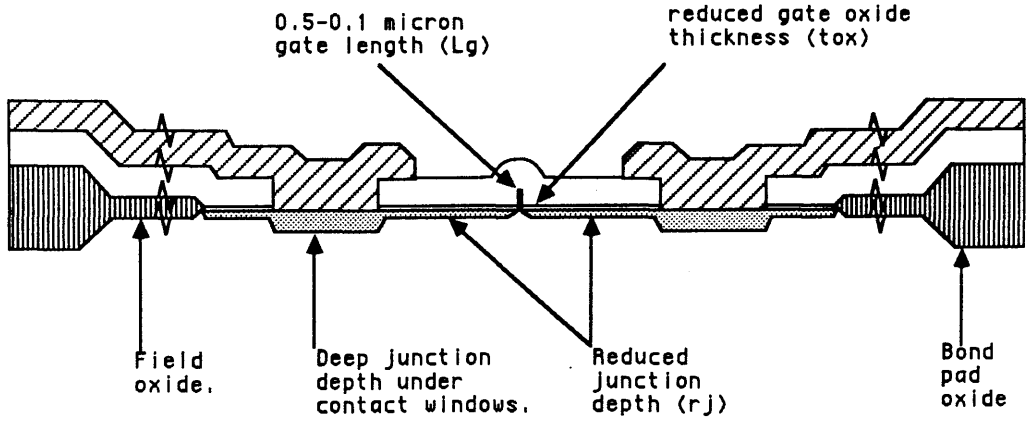


Figure (3-2) The experimental MOSFET with a 0.5 to 0.1 micron gate length together with required structural changes. (shading key as above).

Chapter 4

Submicron Lithography and Etching

4.1 Introduction

This chapter describes the use of negative electron beam resist and dry etching techniques for patterning 0.1 micron features. The characterisation of the Philips resist, HRN, is reported together with its application as a dry etch mask. Two procedures are then described for dry etching 0.1 micron polysilicon gate electrodes.

4.2 The Glasgow Electron Beam Lithography System

All submicron electron beam lithography required for the study was undertaken at Glasgow University using a Philips PSEM 500 Scanning Electron Microscope which has been modified into a very high resolution Electron Beam Lithography (E.B.L.) system. The system is described briefly below and further details are reported elsewhere[4.1,4.2].

The modifications enable the beam deflection coils to be driven under external software control. This is achieved using an Ithaca microprocessor to drive a scan generator which contains two 16 bit digital to analogue converters. The converters are used to 12 bit accuracy to obtain a 4096 by 4096 pixel exposure field.

Complex exposure patterns, which are reduced to a series of rectangles, are then scanned sequentially by the electron beam. The exposure dose for each rectangle is independently controlled by varying the beam dwell time at each pixel position. Prepared pattern files consisting of corner co-ordinate and exposure data are stored on floppy disk and accessed by the Ithaca microprocessor as required. Since only the coil scanning and beam blanking circuits are modified, all other controls, including frame size, spot size and beam energy can be adjusted as for the standard machine. Consequently, a minimum exposure field of 1.5 x 1.2 microns can be selected, together with a minimum beam spot size of 80 angstroms at an

energy of 50keV. The PSEM based E.B.L. system is shown in figure (4–1).

Exposure pattern files are developed on a GEC4180 computer using the DESIGN computer aided design software package[4.1]. This design aid was developed by Dr. W.S. Mackie at Glasgow University for rapid pattern data preparation and in its latest form it incorporates pattern repeat and multi–level mask design features. Following data preparation using DESIGN, the pattern files are then transferred to the Ithaca microprocessor system and stored on floppy disk for later use.

4.3 Electron Beam Resists

Electron beam resists are available which operate in either a positive or a negative sense[4.3]. In the case of positive resists, the exposed regions are dissolved during development, whilst for negative resists the unexposed regions are dissolved.

Both types of electron beam resist can be used to define conducting electrode structures with sub micron dimensions. For positive resists, a technique of lift–off is usually employed whilst for negative resists, a dry etching technique is preferred.

4.3.1 Positive Resists : The Lift–Off Technique

Positive organic resists such as poly–methyl–methacrylate (PMMA), are composed of long molecular chains with high molecular weight. During exposure to an electron beam, extensive chain sission takes place resulting in fragmented molecules with low molecular weight. A developer is then chosen which dissolves the molecules at a rate dependent upon molecular weight. The exposed regions, with low molecular weight, dissolve rapidly whilst the unexposed high molecular weight regions dissolve relatively slowly. After development, the resist image has windows opened where the exposure has taken place. By careful control of exposure dose, a resist profile can be formed with an under–cut, permitting metalisation by lift–off[4.3]. For this technique, metal is evaporated from a point source and is

deposited both onto the resist and through the resist windows onto the substrate. The undercut profile ensures that the metal layer is discontinuous at the window edges. The resist is dissolved by immersion in a resist solvent, causing the overlying metalisation to float off. Only the metal which was deposited directly onto the substrate remains. The complete process is shown in figure (4.2a).

The under–cut profile can be enhanced by using a two layer resist technique[4.4] with the lower layer having a higher exposure sensitivity. During exposure, a wider feature becomes exposed in the this lower layer, which improves the under–cut. The technique has been used to pattern linewidths down to 100 angstroms on thin membranes.

4.3.2 Negative Resists : Dry Etching Technique

Negative organic resists consist of molecular chains with low molecular weight. During electron beam exposure, extensive cross–linking occurs between the molecules and an increased average molecular weight results. Once again, a developer is chosen which dissolves the molecules at a rate dependent upon molecular weight. The unexposed regions of low molecular weight dissolve rapidly whilst the highly cross–linked exposed regions dissolve relatively slowly. Only the exposed regions remain after development.

An electrode pattern can be formed by etching a pre–deposited layer using a mask of patterned resist. For submicron patterns, this requires the use of a negative electron beam resist. The etching is required to be highly anisotropic to prevent lateral etching under the mask and so dry, rather than wet, etching is employed. The process is completed when the residual resist has been stripped (see figure (4.2b).

4.3.3 Positive versus Negative Resists

Positive electron beam resists generally have higher exposure contrast than negative resists and do not suffer from swelling during development as commonly occurs with negative resists. Consequently,

positive resists usually achieve a higher ultimate resolution[4.3]. The compatibility with lift– off and superior resolution have resulted in a general preference for positive electron beam resists.

Both lift– off and dry etching techniques are well established for use with conventional optical lithography. However, for modern VLSI processing, it has emerged that higher yields can be obtained using dry etching[4.5]. Consequently, most VLSI processes totally avoid the use of lift– off.

This move to dry etching, has also led to the development of new negative electron beam resists with improved performance, such as the Philips HRN resist[4.6]. This resist, which is based on a styrene co– polymer, has been characterised at Glasgow for 0.1 micron lithography,as described below.

4.4 Characterisation of HRN Resist

4.4.1 Spin Characteristics

HRN200 (20% solids by weight), as supplied by the Philips Redhill Research Laboratories, was diluted in Microposit resist thinner to obtain HRN150 (15% solids). The characteristic thickness versus spin speed curves were determined for both dilutions. Films were spun for 60 seconds at a range of speeds from 1000 to 8000 r.p.m., baked at 120°C for 20 minutes to drive off the casting solvents and then measured using a Talystep stylus probe. The results are shown in figure (4– 3) and indicate that a thickness range from approximately 1.1 to 0.25 microns can be obtained at these dilutions.

4.4.2 Standard HRN Processing

Resist preparation was standardised for this study, with films being spun for 60 seconds at 4000 r.p.m. using HRN150. The films were baked at 120°C for 20 minutes, to obtain a standard thickness of 0.36 microns.

HRN can be developed using Methyl– Iso– Butyl– Ketone (MIBK) and rinsed in Iso– Propyl– Alcohol (IPA). After electron beam

exposure, HRN films were developed for 15 seconds in MIBK at 22.8°C (+ 0.2°C), then rinsed for 15 seconds in IPA at the same temperature and finally blow dried using compressed nitrogen. Unfortunately, this simple process was found to leave scum marks on the substrate which still persisted when development and rinse times were increased. However, the scumming was removed by repeating the development, rinse and dry cycle. Therefore a standard process was chosen which incorporated this double development cycle. Each cycle was performed using identical conditions as described above.

4.4.3 Sensitivity and Contrast

For negative resists, the final image thickness after development is a function of the exposure dose. An increased exposure dose leads to more cross–linking and fewer molecules are dissolved. Figure (4–4) shows an idealised resist thickness versus exposure dose characteristic curve for a negative resist. The film thickness is normalised to its initial value and a logarithmic scale is used for exposure dose. The figure indicates two important features of negative resists. Firstly, it shows that no image results for exposure below a certain critical dose, termed the gel dose (D_g). Secondly, it shows that above this dose, a logarithmic relationship exists between image thickness and exposure dose. The gamma value (γ) of a resist, which gives a measure of the exposure contrast, is defined by equation (4.1). D_m is defined as the projection of a straght line which intersects the gel dose point and the 50% thickness point ($D_{0.5}$) as shown in figure (4–4).

$$\gamma = \mathrm{Log}_{10}[D_m/D_g]^{-1} \qquad (4.1)$$

Gel dose and gamma value are both useful parameters for comparing the performance of two different negative resists. However such comparisons are only possible when identical substrate materials and exposure energy are used.

A simple experiment was undertaken to determine the gel dose

and gamma value for HRN when exposed on a silicon substrate, using a 50keV electron beam. HRN films were spun using standard conditions to obtain a resist thickness of 0.36 microns. The films were then exposed with patterns consisting of 120 x 30 micron rectangles with various exposure doses. An E.B.L. exposure frame size of 400 x 300 micron, a spot size of 0.125 microns and a beam energy of 50keV were chosen. The beam current was measured by aligning the beam into Faraday cup4.1 and was then used by the Ithaca microprocessor to set correct beam dwell times for the desired exposure. After exposure and development, the pattern thicknesses were measured by Talystep. The results are shown in figure (4−5), and it can be seen that under these conditions, HRN has a gel dose of 145 uC/cm^2 and a gamma value of 3 (for $D_{0.5}$= 210uC/cm^2 and D_m= 310uC/cm^2). Whilst negative electron beam resists are usually associated with low contrast, the gamma value (γ) for HRN is close to that of PMMA positive resists, suggesting that a comparable resolution might be possible for HRN. For comparison, Evacite PMMA has a gamma of 4.6, under similar exposure conditions[4.2].

4.4.4 HRN Resolution

The use of HRN for patterning linewidths down to 0.1 microns was investigated. An exposure frame size of 100 x 80 microns was chosen resulting in pixel dimensions of approximately 0.025 and 0.02 microns. At this frame size, a 0.1 micron line corresponds to 5 pixels, when parallel to the longer axis. A beam spot size of 160 angstroms at an energy of 50keV was chosen for the exposures.

HRN150 resist was spun on silicon substrates and baked using the standard conditions to obtain a thickness of 0.36 microns. Patterns of isolated lines were then exposed for widths from 3 to 21 pixels and exposure doses from 350 to 1200 uC/cm^2. The lines were isolated from each other to minimise inter−proximity effects whereby back−scattered electrons from one exposure pattern lead to an increased exposure dose at all adjacent patterns[4.1,4.2,4.3]. After exposure, the patterns were developed using the standard development conditions described earlier and then viewed using a scanning electron microscope (SEM). Linewidth measurements were recorded and are

shown in figure (4−6) in the form of a plot of linewidth verses log exposure dose.

From figure (4−6), it can be seen that 0.1 micron linewidths are readily achieved for 5 pixel lines exposed in 0.36 micron thick HRN resist films. The finest linewidth measured was 0.07 microns for a 3 pixel line with an exposure dose of 800 uC/cm^2. For exposure by five pixels, a 0.1 micron linewidth required an exposure dose of 600uC/cm^2. It was observed that a 20% change in exposure dose caused only a 10% change in resist linewidth for the 0.1 micron condition described above, indicating extremely good linewidth control with exposure dose.

4.4.5 Standard Exposure Parameters

A set of standard exposure parameters, as shown in table (4−1), were selected from the results of figure (4−6) for patterning various nominal linewidths. Further investigation showed that 0.1 micron lines could be reliably reproduced to within $\pm$0.01 microns using standard exposure parameters together with standard processing conditions.

4.4.6 Resist Profiles

For application as a dry etch mask, the resist height and cross−sectional profile are of considerable importance. Consequently, a technique was developed for viewing the HRN patterns in cross−section.

HRN films were spun onto silicon substrates and prebaked using standard conditions. The samples were scribed and cleaved into 5mm squares. Each sample was further scribed with two 2mm co−linear scribe marks, separated by a gap of approximately 1mm. Prior to exposure in the E.B.L. system, each sample was rotated to align the longer axis of the exposure field perpendicular to the two scribe marks. Next, a pattern of submicron lines was exposed at ten positions, in an arrangement of two inter−leaved rows of five and located across the scribe gap, as indicated in figure (4−7). Computer

goniometer control was used to ensure correct movement to each consecutive exposure site. The patterns were developed and cleaved along the scribe axis. Then, each cleaved sample was coated with a thin film of gold–palladium (typ. 100 angstroms) and mounted in a SEM to image the cross–sectional resist profile. The gold palladium film was required to avoid charging effects during this SEM analysis.

Figure (4–8) shows an electron micrograph of a typical HRN line, obtained by this technique. The line was exposed using the standard exposure parameters and development conditions for a nominal linewidth of 0.1 microns and it can be seen that the resulting linewidth of is very close to this value. The height is approximately 70 percent of the original film thickness of 0.36 microns. It can also be seen that the overall resist profile is highly rectangular with nearly vertical sidewalls, both of which are desirable for dry etch masking. Larger linewidths, also exposed under standard conditions, displayed similar vertical sidewalls. All linewidths were in excellent agreement with the nominal values of table (4–1).

Using the standard processing conditions, but with the exception of reduced resist spin speed, different resist profile heights were investigated. It was found that a maximum height of 0.6 microns could be obtained for 0.1 micron linewidths, using standard exposure conditions. When aspect ratios of greater than 6 to 1 were attempted, the resist collapsed sideways.

4.4.7 Narrow line, grating and Mesh Patterns

HRN was investigated further by patterning narrow lines, grating and mesh patterns. However, these patterns were not applied to device fabrication.

A reduced and therefore non–standard film thickness of 0.11 microns was used for improved resolution. A minimum linewidth of 0.06 microns was then obtained for a 3 pixel line at an exposure dose of 800uC/cm^2, a spot size of 160 angstroms and a frame size of 100 x 80 microns. Figure (4–9) shows an electron micrograph of a 0.075 micron wide HRN line, patterned using 3 pixels and an exposure dose of 1100uC/cm^2.

Submicron pitch gratings were patterned into the 0.11 micron

films. The minimum grating pitch gives a further indication of the resolution of a resist, since the effect of electron back–scattering from the substrate becomes more significant than for isolated lines[4.3]. Figure (4–10) shows an electron micrograph of the finest pitch grating obtained at a frame size of 100 x 80 microns. Single pixel lines were exposed at a pitch of seven pixels using an exposure dose of 1600uC/cm^2. The resulting grating consists of lines with a width of approximately 0.09 microns and a pitch of 0.17 microns.

Having demonstrated isolated lines and grating patterns, figure (4–11) shows an electron micrograph of a two dimensional mesh, again obtained using a HRN film thickness of 0.11 microns and a frame size of 100 x 80 microns. Lines were patterned in each direction using a width of one pixel, a pitch of 7 pixels and an exposure dose of 1400uC/cm^2. The resulting mesh has a pitch of 0.17 and 0.2 microns in each direction.

The best resolution for any organic negative resist reported to date is 0.023 microns[4.7] for patterning on thin substrates. Whilst HRN does not achieve this resolution, an excellent performance down to 0.1 microns has been demonstrated.

4.5 Dry Etching of Polysilicon

A dry etch process, suitable for patterning 0.1 micron polysilicon gates over thin gate oxides must satisfy several requirements. Firstly, the process must offer a high degree of etch selectivity between the polysilicon and the thin underlying gate oxide layer. This is necessary as the oxide layer must not be etched through. Secondly, the resist mask erosion rate must be sufficiently low to ensure that an effective mask remains throughout the etching process. Thirdly, etching is required to be highly anisotropic to produce gate electrodes with vertical or near vertical sidewalls. This is important because the gate will be used as a self–aligned implant mask. Finally, the etching process should minimise any crystallographic damage to the active region of the MOS structure[4.8].

It should be noted that wet etching is generally unsuitable for submicron patterning because it does not offer the required anisotropy and resolution.

4.5.1 Dry Etching Machines

Several types of dry etching machine are currently used for microfabrication. The three most important types, ion beam etching (IBE), planar plasma etching (PE) and reactive ion etching (RIE)[4.9], all share several basic features. A low energy plasma is generated by either d.c. or r.f. discharge across a chamber containing a gas at low pressure. The gas is ionised and the ions are then accelerated by an electric field towards the substrate. When the ions impinge onto the substrate they attack the surface atoms by a partly chemical and partly physical mechanism, resulting in a gradual erosion. Dry etch machines usually optimise either the physical or chemical surface etching mechanisms, or combine both. In general, physical etching mechanisms are more directional and chemical mechanisms give higher etch selectivities. The three most important types of dry etch machine in current use, are described next.

4.5.1.1 Ion Beam Etching

Ion beam etching (IBE) essentially uses a physical erosion mechanism. A noble gas plasma, typically of argon, is generated in a gun assembly by d.c. discharge. The ions are then extracted from the plasma and accelerated through an acceleration grid to a specific energy. The ions pass from the gun into the sample chamber in the form of a beam which uniformly impinges onto the substrate. Chamber pressures below 10−5 mTorr are usually employed with ion energies in the range of 500 to 2000eV.

Since the etch process has a physical erosion mechanism, most materials have a similar etch rate and consequently etch selectivity is poor[4.10]. Although the beam is highly directional, anisotropy is often degraded by re−deposition, facetting and trenching effects.

Ion beam etching is generally unsuitable for MOS device fabrication due to the high crystallographic damage levels which occur.

4.5.1.2 Planar Plasma Etching

A planar plasma etching (PE) machine consists of a chamber containing two parallel electrode plates. One electrode is grounded and the other electrode is driven by an r.f. generator, connected through an impedance matching circuit. The sample is mounted on the grounded electrode and the chamber is then evacuated. An etch gas is then introduced into the chamber and the r.f. generator is used to establish a plasma. The etch gas is chosen such that it dissociates to produce ions which will react with the substrate material to be etched. Chamber pressures in the range 0.1 to 1 torr are usually chosen. The r.f voltage applied across the electrodes has two functions, it both creates the plasma and then accelerates the resulting ions and electrons. The plasma potential changes polarity with respect to the electrodes, twice per r.f. cycle. Consequently, the ions and electrons are alternately accelerated towards each electrode. Substrate etching occurs only during the part of the cycle when the reactive ions are directed towards the grounded electrode. Since the total ion and the total electron charge reaching the electrode during each cycle are equal, insulated substrates can be etched without charging effects. Also, because the ions have a lower mobility than the electrons, an equilibrium is reached whereby a d.c. voltage component is developed across a dark space region which separates the main plasma from each electrode as seen in figure (4−12). This d.c. voltage component effectively increases the proportion of the cycle time for which the ion current flows and results in the ions having a higher mean energy.

Plasma etching processes are complex and in many cases, the exact mechanisms of the surface reactions are not fully understood. However, in all cases, chemically reactive ions impinge onto the substrate and react to form volatile products. For some etch gases and etch conditions, the directionality of the ions leads to highly anisotropic substrate etching. Due to the highly chemical nature of the etching, very high etch selectivities can be obtained between different materials[4.11].

Plasma etch rates are often improved by reducing the area of the grounded electrode with respect to the powered electrode. This leads to a higher current density and dark space d.c. voltage at the grounded electrode than for the powered electrode. The increased

voltage component increases the mean energy of the ions and so the physical erosion mechanism is enhanced. A capacitor is required in series with the powered electrode to support a d.c. voltage component which arises from this asymmetry between dark space voltages.

4.5.1.3 Reactive Ion Etching

A reactive ion etching (RIE) machine[4.12] is essentially similar to a plasma etching (PE) machine except that the substrate is mounted onto the powered electrode which is again capacitively coupled. The whole chamber casing is then grounded and acts as a large electrode. This results in an extremely asymmetrical configuration and consequently a very high dark space d.c. voltage component occurs between the plasma and the powered electrode. The resulting mean ion energy is increased with respect to planar plasma etching. Also, a lower operating pressure, typically between 0.1 and 0.01 torr, results in fewer ion collisions and therefore increases the directionality of the ions. Both effects lead to an enhanced physical erosion mechanism.

Figure (4−13) shows a typical reactive ion etching system. The d.c. bias voltage across the r.f. generator coupling capacitor can be monitored and is approximately equal to the d.c. voltage component across the dark space at the powered electrode.

4.5.1.4 Plasma Etching versus Reactive Ion Etching

The distinction between planar plasma etching and reactive ion etching rests with the configuration of the machine used and not necessarily the etching mechanisms, since for both techniques, the etching is partly chemical and partly physical. In general, planar plasma etching relies more strongly on a chemical mechanism whilst for reactive ion etching, the physical mechanism is enhanced. Consequently, planar plasma etching usually results in higher etch selectivity, whilst reactive ion etching results in improved anisotropy. It should be noted that reactive ion etching is likely to result in slightly increased levels of crystallographic damage due to the higher ion energies involved.

4.5.2 Polysilicon Etching

Both plasma and reactive ion etching techniques have been investigated for this study. Initially an Electrotech 317 planar plasma etching system at B.T.R.L. was used with pure chlorine. Later work was then carried out on a Plasma Technology RIE 80 reactive ion etching system at Glasgow, using silicon tetra–chloride.

4.5.2.1 Submicron Plasma Etching with Chlorine

The Electrotech system at B.T.R.L. had been previously characterised for a 3 micron polysilicon gate process and sub 0.5 micron etching had not been attempted previously.

A batch of silicon wafers for polysilicon etching experiments was first prepared at B.T.R.L. Oxide layers were grown to a thickness of 150 and 300 angstroms, on which 0.3 micron polysilicon was deposited. The polysilicon was then implanted with a phosporous dose of $6x10^{15}$ atoms/cm^2 at 40keV and annealed in a nitrogen ambient at 1000°C for 10 minutes. The implantation and anneal steps, which are required to lower the gate resistivity for actual devices wafers, were included to ensure etching consistency between test wafers and later device wafers.

The wafers were patterned at Glasgow with sub 0.5 micron HRN lines. Standard electron beam exposure and development conditions were used to obtain resist linewidths down to 0.1 microns and once again, the patterns were arranged across scribe gaps to permit profile evaluation.

The HRN patterned wafers were then etched using their standard etching conditions which resulted in highly anisotropic etching and good selectivity over oxide. A chlorine chamber pressure of 300 mtorr and an r.f. current of 2 amps (r.m.s), at a frequency of 380kHz, were used.

End–point detection of the polysilicon etching was achieved visually. During the final moments of etching, optical interference fringes could be seen shrinking into the centre of the wafer due to

slight preferential etching towards the wafer edge. The disappearance of these fringes was used to indicate 100% etching (at the wafer centre). A controlled amount of over–etching was required to compensate for any variation in the thickness of the polysilicon or its native oxide (10–30 angstroms) which also required etching. In addition, any non–uniformity arising from the etching machine required compensation. In accordance with standard device fabrication at B.T.R.L., a 15% over–etch time (at wafer centre) was chosen[4.13].

The patterned test wafers were etched with 15 and 30% over–etch times and for no over–etch, as determined at the wafer centre. The residual HRN masking was later stripped using a non critical oxygen plamsa etching step. The etch rates of polysilicon and HRN were determined from Talystep measurements of samples, before and after resist stripping. In addition, the oxide erosion was determined by measurement of residual oxide thickness for various over–etch times. Polysilicon and HRN etch rates were measured as 1100 Å/min and 650 Å/min respectively, giving a relative etch selectivity of 1.7 to 1. The etch rate of silicon dioxide was determined to be approximately 40 Å/min, resulting in selectivity to polysilicon of 25 to 1. For the 30 percent over–etch case using an 150 angstrom oxide, the resulting oxide thickness was measured to be close to 100 angstroms, as measured by ellipsometer.

Next, the wafers were cleaved to examine the etched profiles. Figure (4–14) shows an electron micrograph for a line which received a 15% over–etch. The line shows an over–cut profile. However, during the course of the study, a variation from over–cut to under–cut was observed for nominally identical 15% over–etch conditions. This variation is indicated in figure (4–15), but the exact cause was not established. An un–monitored variation in the plasma etching parameters is likely. However, it should be noted that consistent profiles were obtained for wafers etched sequentially on the same day.

The investigation has demonstrated the use of chlorine plasma etching for patterning polysilicon gate lines with a width of approximately 0.1 microns. This technique was then applied to experimental device fabrication.

Figure (4–16) shows a chlorine plasma etched 0.15 micron polysilicon line, which is connected to a 3 micron track. The

patterning of mixed high and low resolution patterns is required to make electrical connection to the submicron gate electrodes.

4.5.2.2 Submicron Reactive Ion Etching in Silicon Tetra– Chloride

The Plasma Technology system was investigated for polysilicon etching for two reason. Firstly to see if improved submicron profile control could be obtained with reactive ion etching, and secondly to enable polysilicon etching to be undertaken independently at Glasgow. The Plasma Technology machine had not been previously characterised for polysilicon etching, and so some preliminary etch rate measurements for polysilicon, silicon dioxide and HRN were undertaken.

Test wafers were prepared at the Edinburgh Microfabrication Facility with layer thicknesses as for the previous etching experiments. A number of wafers were patterned with large HRN features and then reactive ion etched using silicon tetra– chloride[4.14]. A 13.56 Mhz r.f. power supply was used and both power and chamber pressure were varied to optimise etching.

A silicon tetra– chloride flow rate of 55– 60 cc/sec was used to obtain a chamber pressure of 10 mtorr. Power levels of 100, 50 and 25 watts were used resulting in d.c. bias values of 400, 280 and 200 volts respectively. Etch rates were then measured and the results are shown in table (4– 2) and figure (4– 17). It can be seen that whilst the polysilicon etch rate decreases with reduced power, the selectivity over HRN improves and a selectivity of 2 to 1 is obtained at 25 watts. This enhanced selectivity can be explained through the decreased d.c. bias at lower power which leads to reduced physical etching.

The effect of chamber pressure was investigated at a constant power level of 25 watts. The polysilicon test wafers were patterned with HRN features, and then etched at pressures of 10 and 50 mTorr. Wafers with thick silicon dioxide layers were also etched to ascertain the oxide etch rate. The results are shown in the lower part of table (4– 2) and in figure (4– 18). It can be seen that polysilicon selectivity with respect to both HRN and oxide, increases with pressure, which once again can be explained by the reduced d.c. bias.

At 50 mtorr, the bias is only 70 volts, resulting in a polysilicon etch selectivity of 5 to 1 with respect to HRN and 12 to 1 with respect to silicon dioxide.

Using the profiling technique, 0.1 micron lines were etched into polysilicon using the 10mtorr and 50 mtorr etch conditions at 25 watts. It emerged that these conditions produced a tapered profile and a significant under–cut profile respectively. Whilst both conditions are capable of 0.1 micron resolution, neither is ideal for MOSFET fabrication. The low pressure condition produces a suitable profile but the selectivity with respect to oxide is low at only 7 to 1. The high pressure gives an improved selectivity of 12 to 1 but has an extremely slow polysilicon etch rate and results in an undesirable under–cut profile.

A process was chosen which exploits the etching properties at both operating pressures. The wafers were etched at a chamber pressure of 10 mtorr until the etch end–point was observed via the thin film interference fringes on the wafer. Under these conditions, the polysilicon etch rate is 200 A/min and a slightly tapered profile results. Once the end–point had been reached, the plasma was turned off and the chamber pressure was increased to 50 mtorr for the desired over–etch step.

This two step etch process was used to etch 0.1 micron lines into 0.3 micron thick polysilicon layers, stopping on 150 and 300 angstron oxide layers. Etch times of zero, 5, 10 and 20 minutes were chosen for the second etching step, which corresponds approximately to zero, 10, 20 and 40 percent over–etching at the original conditions (calculated from relative etch rate data). Samples were prepared with 0.1 micron lines and etched using the two etch steps. The residual resist was removed and the samples were cleaved and examined in profile. Figure (4–19) shows electron micrographs of the resulting polysilicon features. As the over–etch time increases from zero to 5 and then 10 minutes, the tapered profile becomes steeper. For 20 minutes, an under–cut is apparent. Even for samples with the thinner gate oxides (150 angstroms) and the longest over–etch times (20 minutes), the oxide was not etched through.

Good consistency was observed between all samples etched under the same conditions, indicating that a high degree of sidewall profile control was possible.

For device fabrication, a 5 minute over–etch time was chosen in order to retain the taper whilst giving an effective over–etch of approximately 10 percent. HRN resist was removed after etching using an oxygen plasma. Any residual resist which remained was then stripped by immersion in fuming nitric acid for 5 to 10 minutes.

4.5.2.3 Chlorine versus Silicon Tetra–Chloride Etching

Both the chlorine plasma etching process and the silicon tetra–chloride reactive ion etching process have been demonstrated to be suitable for 0.1 micron patterning. The chlorine process offers a higher poly–silicon etch rate together with a higher selectivity over oxide of 25 to 1. However, the process requires stringent safety precautions due to the highly toxic nature of chlorine. In contrast, the silicon tetra–chloride process gives improved profile control via the two step etching process, although etch rates and selectivities are inferior. These obvervations are in agreement with the earlier discussion on plasma verses reactive ion etching. The damage levels resulting from each etching process have not been determined and will require further investigation.

4.6 Summary

The negative electron beam resist, HRN, has been used to pattern linewidths to 0.1 microns and below. The resist image has then been used with two different dry etching processes to pattern 0.1 micron polysilicon gate electrodes. Firstly the standard B.T.R.L. chorine plasma etching process has been demonstrated and secondly a new reactive ion etching process, developed at Glasgow, has been demonstrated which gives improved profile control for 0.1 micron features. In both cases, 0.1 micron gate patterns have been etched over 150 angstrom gate oxides.

References

4.1 W.S. Mackie, 'Techniques and applications of very high resolution electon beam lithography,' Ph.D. Thesis, University of Glasgow, April 1984.

4.2 C.E Binnie, 'The fabrication of small geometry MOSFETs using electron beam lithography', Ph.D. Thesis, University of Glasgow, Feb 1985.

4.3 G.R. Brewer (Ed),'Electron–beam technology in micro– electronic fabrication,' Academic, New York, 1980.

4.4 S.P. Beaumont, T.Tamamura, C.D.W. Wilkinson, 'A two layer resist system for efficient lift–off in very high resolution electron beam lithography,'Proc. Microcircuit 80, R.P. Kramer (Ed.), Delft University Press, 1981.

4.5 S.M. Sze, 'VLSI technology,' McGraw–Hill, Singapore, 1983.

4.6 Data sheet from P.W. Whipps, Philips Redhill Laboratories, Surrey. 'Negative–acting electron beam resist: type HRN and HSN, 1980.

4.7 T. Tammamura, K. Sukegawa and S. Sugawara, 'Resolution limit of negative electron beam resist exposed on a thin–film substrate,' J. Electrochem. Soc., vol 129, p1831, 1982.

4.8 L.M. Ephrath and D.J. DiMaria, 'Review of RIE induced radiation damage in silicon dioxide,' Solid State Technol., vol 24, p182, 1981.

4.9 A.R. Reinberg, 'Dry processing for fabrication of VLSI devices,' in N.G. Einspruch (Ed) VLSI Electronsics, Academic, New York, 1981, vol 2.

4.10 C.M. Melliar–Smith and C.J. Mogab, 'Plasma–assisted etching techniques for pattern delineation,' in J.L. Vossen and W. Kern. (Eds), 'Thin film processes,' Academic, New York, 1979.

4.11 B. Chapman, 'Glow Discharge Processes,' Wiley, New York, 1980.

4.12 L.M. Ephrath, 'Reactive ion etching for VLSI,' IEEE Trans. Electron Devices, vol ED–28, p1315, 1981.

4.13 Many helpful discussions with I.P. Thomas at B.T.R.L. are acknowledged.

4.14 M. Sato and H. Nakamura, 'Reactive ion etching of aluminium using $SiCl_4$,' J. Vac. Sci. and Technol, vol 20(2), p186, 1982.

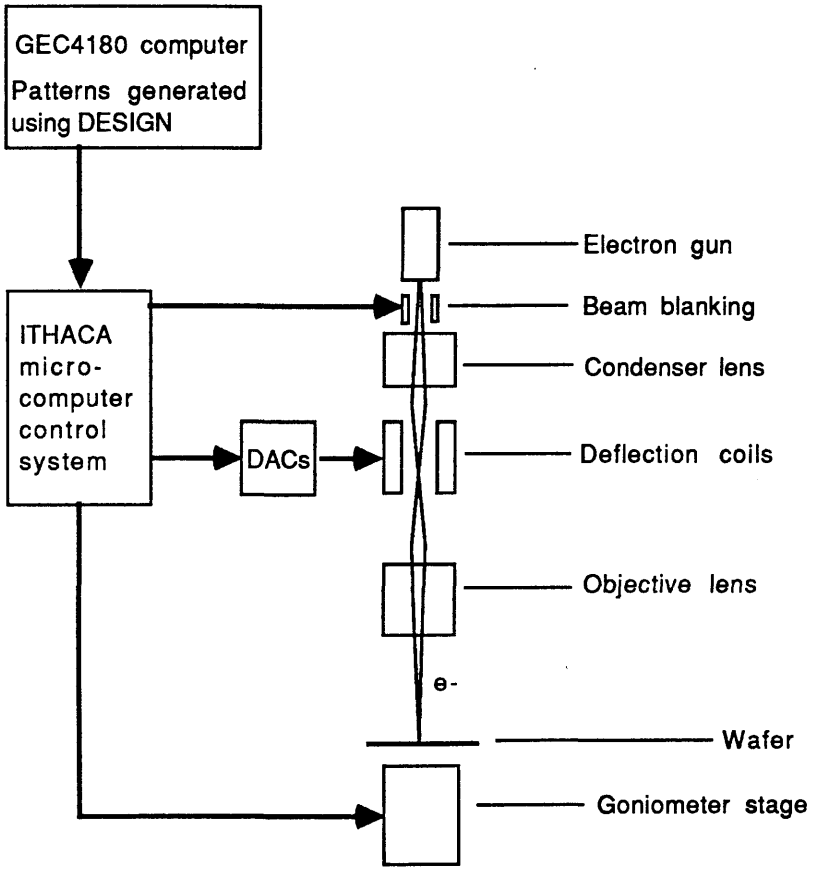


Figure (4-1) The PSEM based High Resolution E.B.L. System at Glasgow.

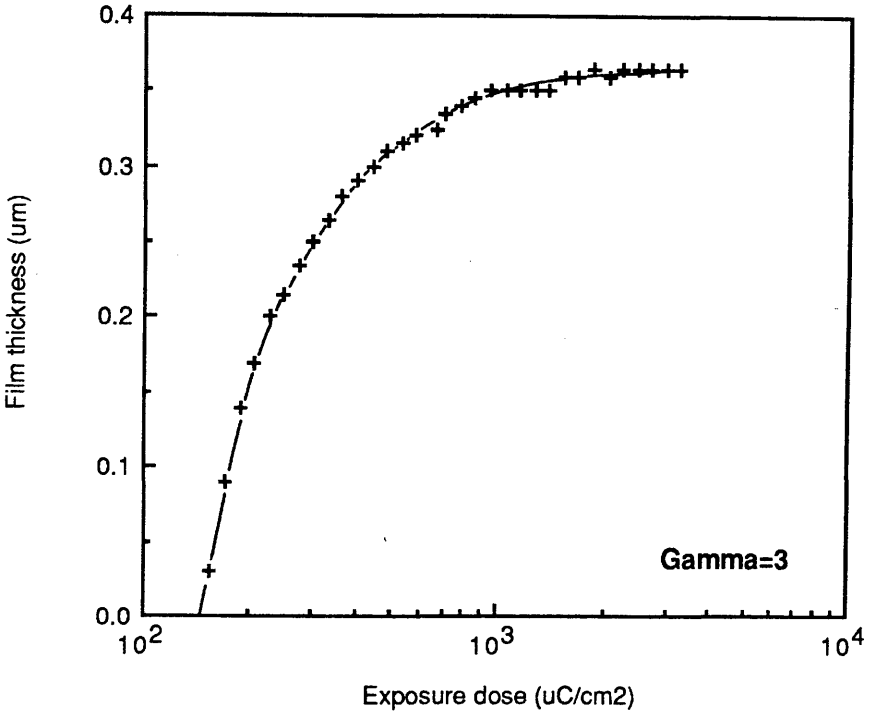


Figure (4-5) Resist thickness versus exposure dose characteristic curve for HRN patterns onto a silicon substrate at an exposure energy of 50keV.

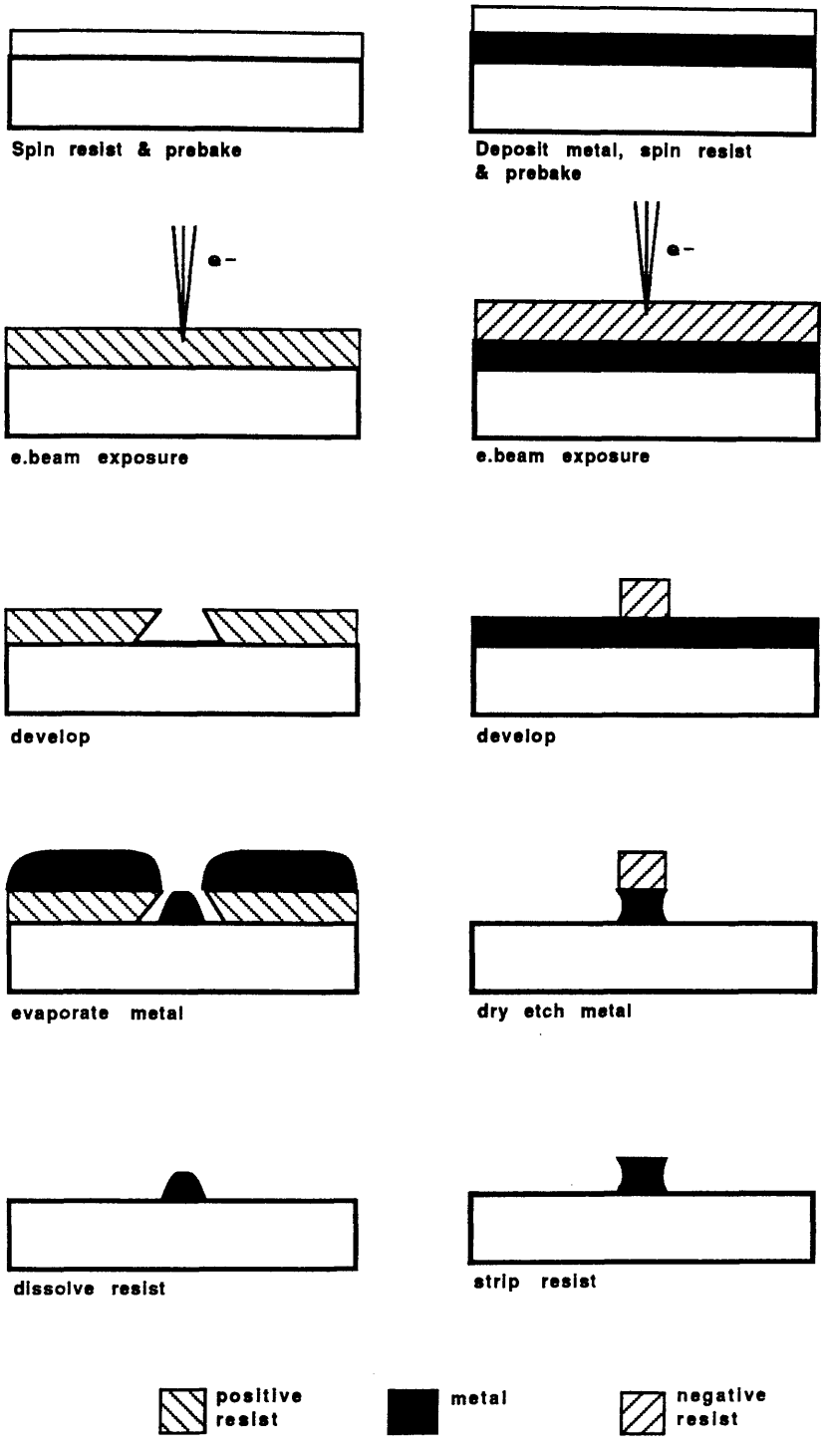


Figure (4-2) The use of positive and negative electron beam resists for submicron lithography. (a) Shows positive resist and the lift-off technique and (b) shows negative resist and dry etching.

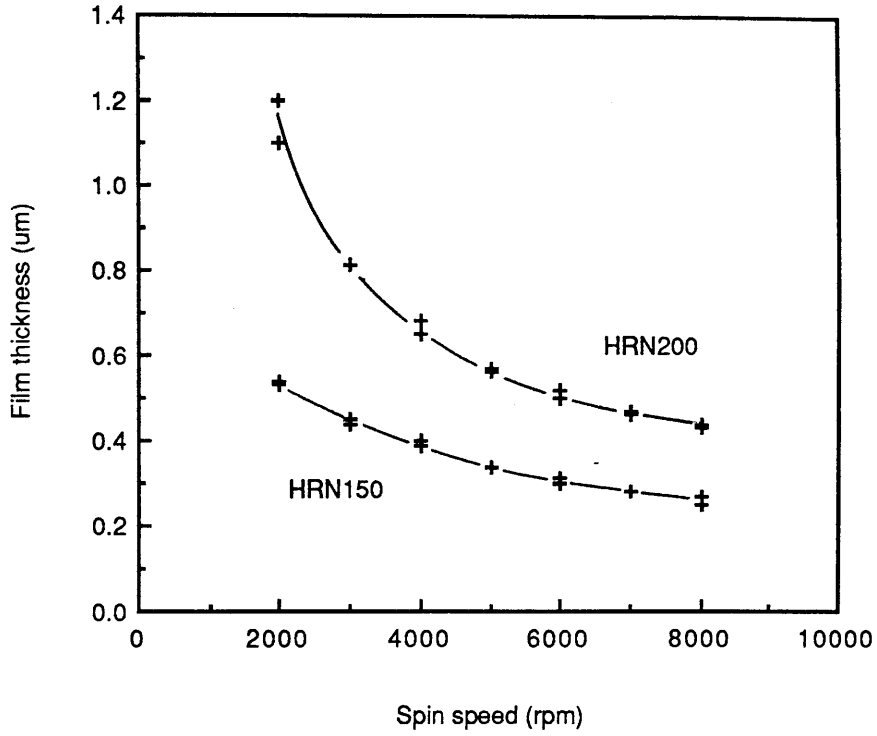


Figure (4-3) Spin speed versus thickness curves for HRN resist at two dilutions (HRN200=20% solids and HRN150=15% solids).

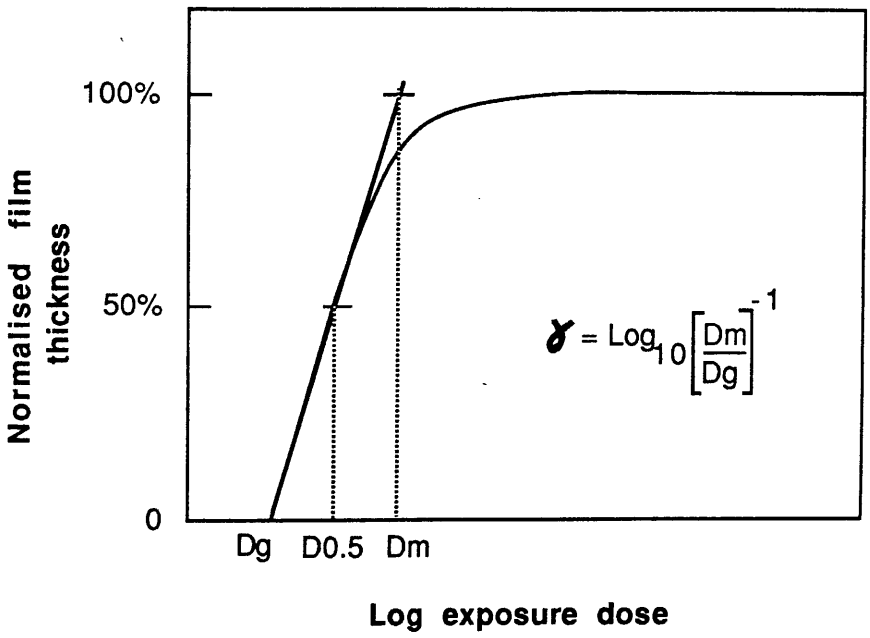


Figure (4-4) Normalised resist thickness versus exposure dose characteristic for a negative electron beam resist.

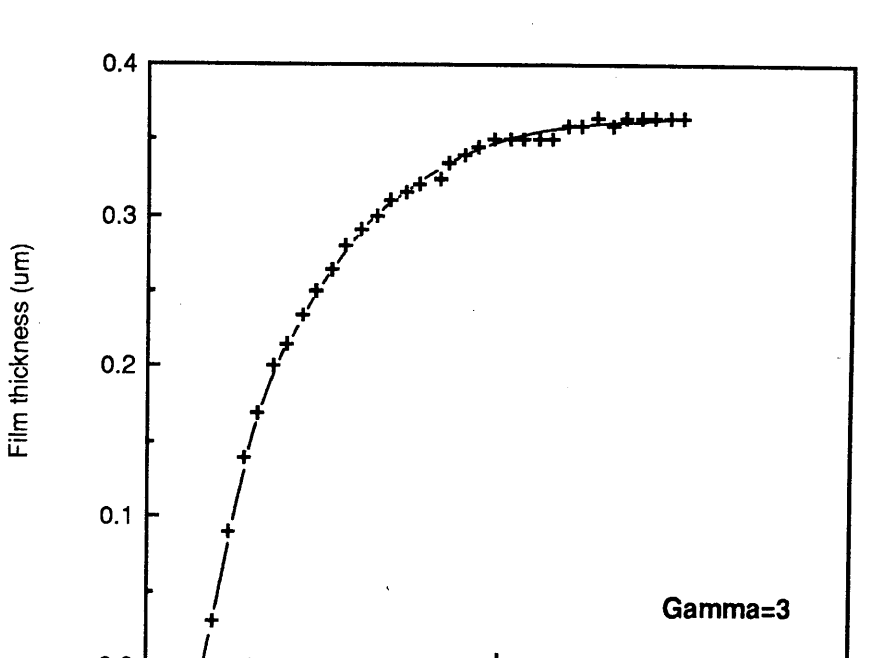


Figure (4-5) Resist thickness versus exposure dose characteristic curve for HRN patterns onto a silicon substrate at an exposure energy of 50keV.

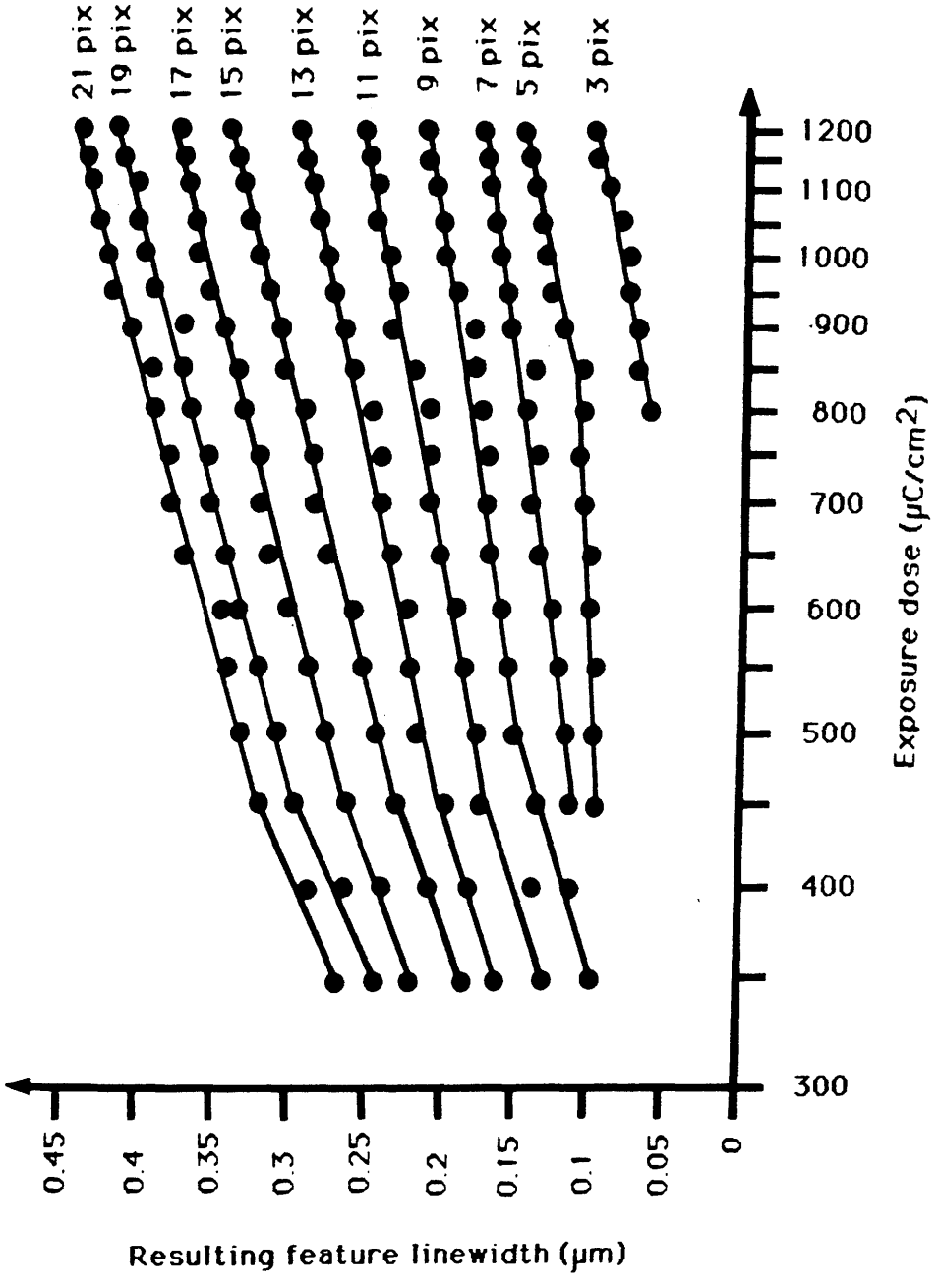


Figure (4-6) Linewidth versus exposure dose curve for HRN lines patterned onto silicon substrates at an exposure energy of 50keV. Curves are shown for various pixel numbers exposured using a 100 x 80 micron frame size.

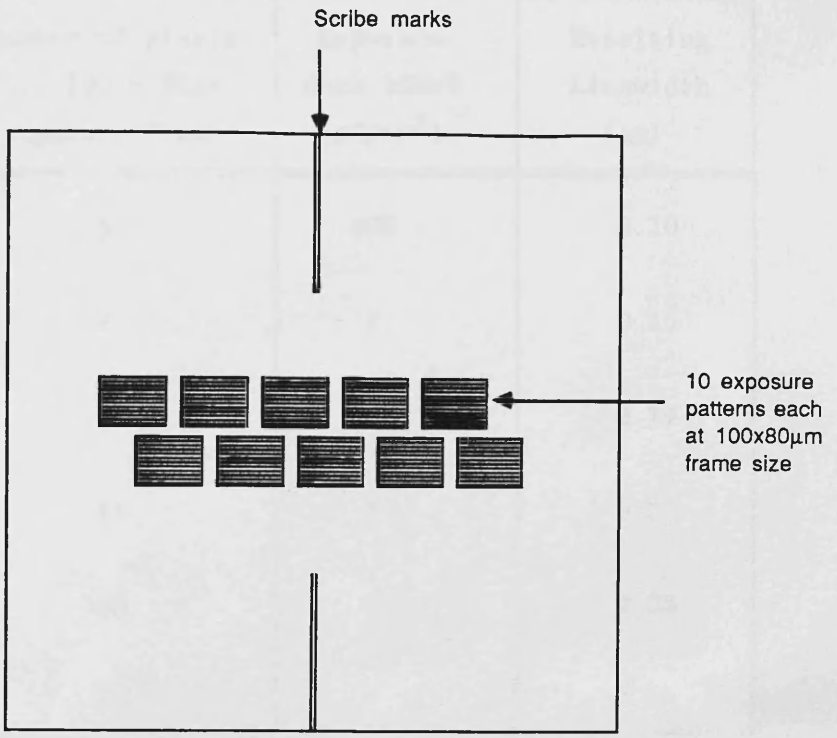


Figure (4-7) An arrangement of ten patterns located across a scribe gap for subsequent cleaving and SEM evaluation.

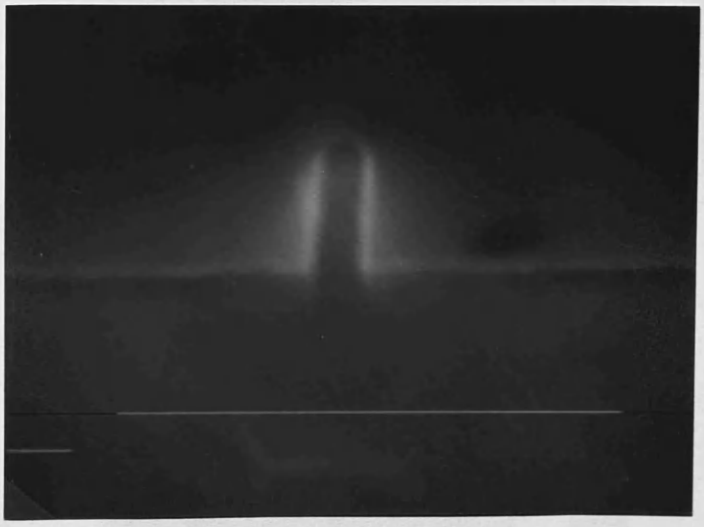


Figure (4-8) Electron micrograph showing a 0.1 μm HRN line viewed in cross-section (marker=1um).

Number of pixels for 100 x 80μm exposure frame	Exposure dose 50keV ($\mu C/cm^2$)	Resulting Linewidth (μm)
5	600	0.10
8	"	0.15
10	"	0.19
15	"	0.27
20	"	0.35
25	"	0.44

Table (4-1) Table showing standard exposure parameters for patterning HRN linewidths down to 0.1 μm using a 100 x 80 micron exposure frame.

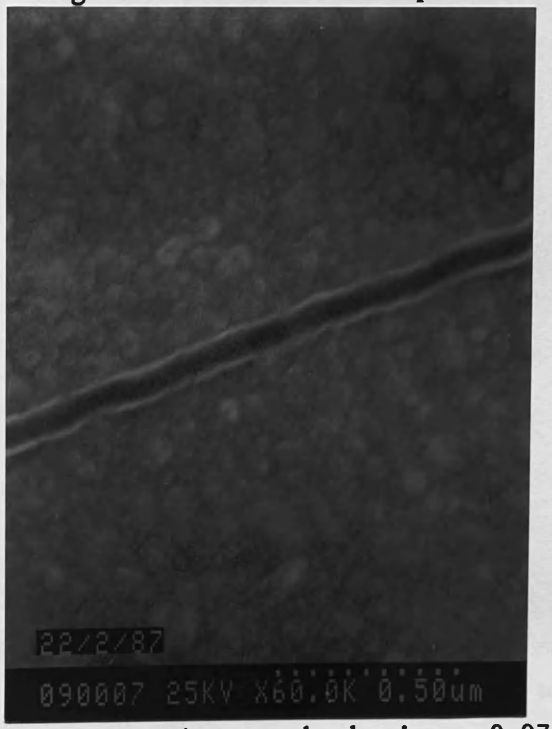


Figure (4-9) Electron micrograph showing a 0.075 μm line patterned into HRN resist.

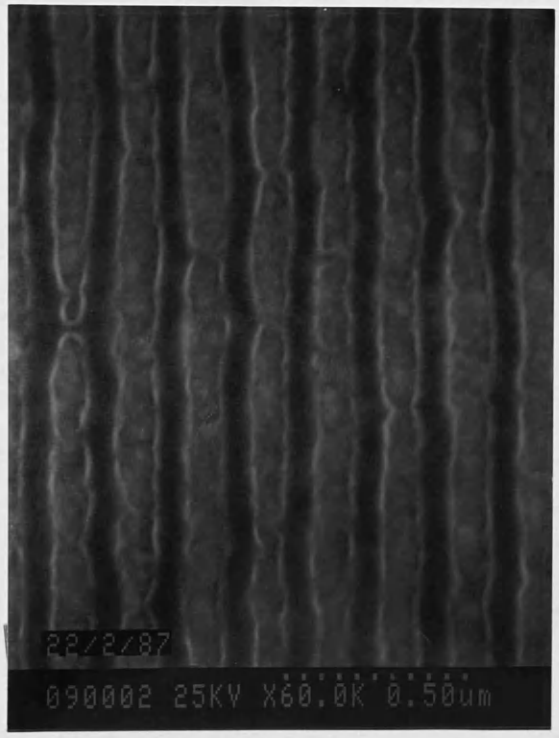


Figure (4-10) Electron micrograph showing a grating, patterned into HRN resist. The grating pitch is 0.17 microns.

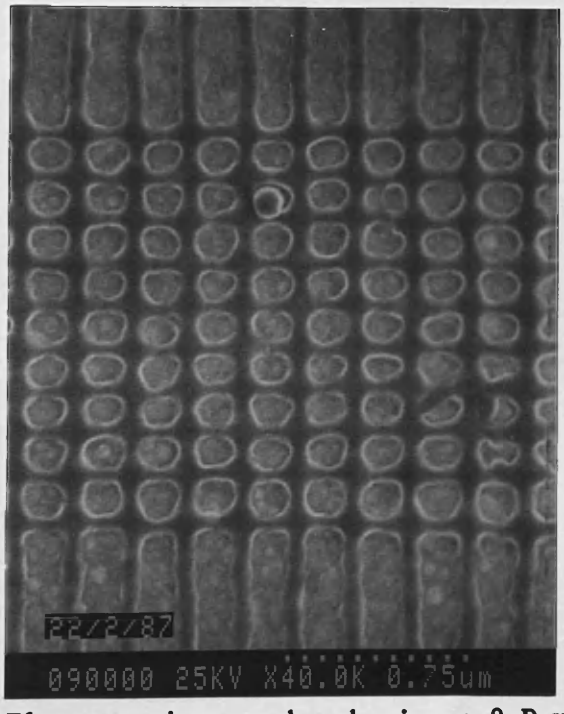


Figure (4-11) Electron micrograph showing a 2-D mesh, patterned into HRN resist. The pitch is 0.17 and 0.2 microns in each respective direct.

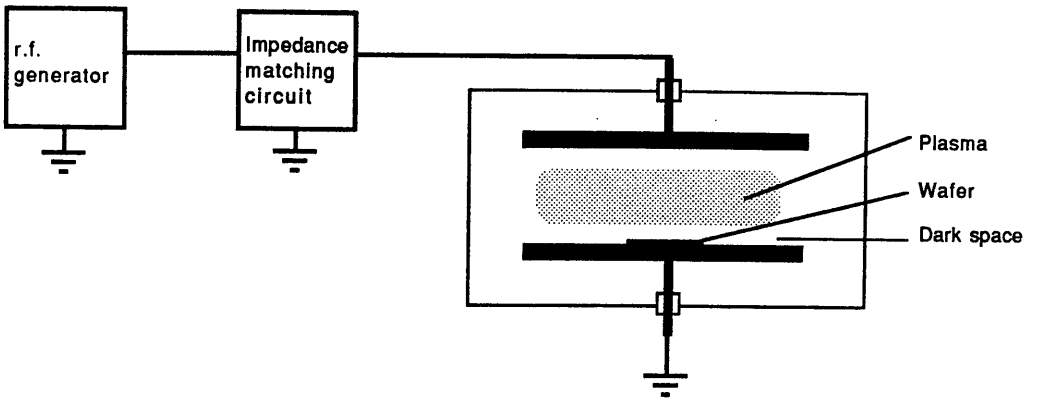


Figure (4-12) Planar plasma etching system.

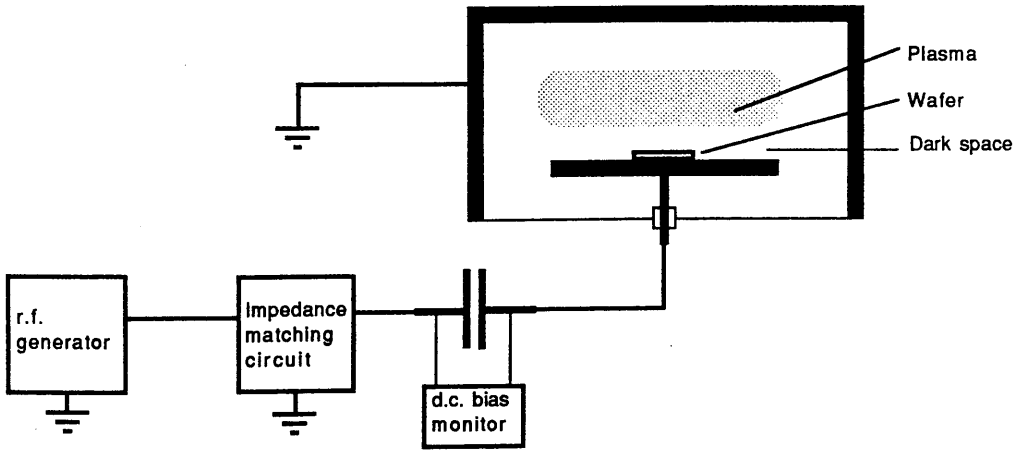


Figure (4-13) Reactive ion etching system.

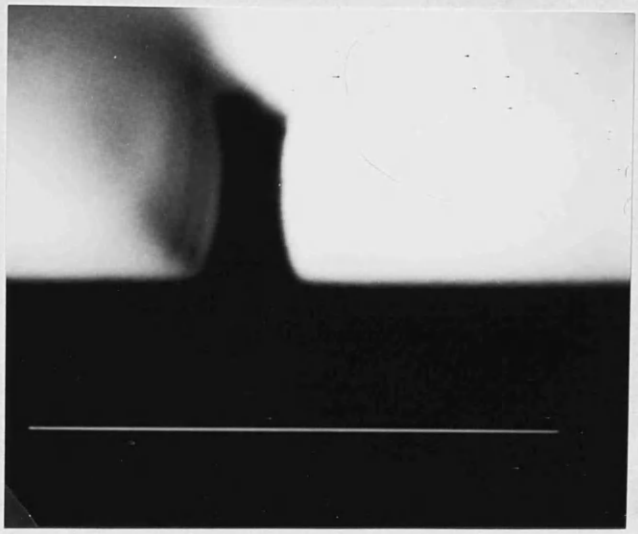


Figure (4-14) Electron micrograph showing a submicron polysilicon line in cross-section. The line was plasma etched in chlorine with a 15% over-etch time (marker=1um).

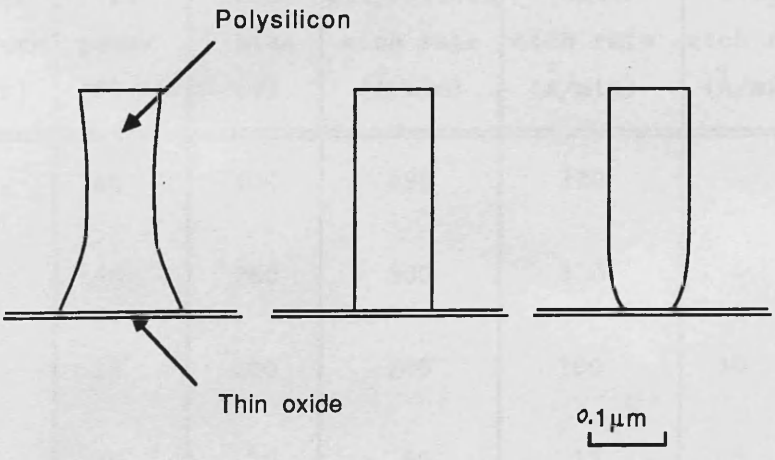


Figure (4-15) Diagram showing variation in polysilicon etched profiles obtained using chlorine plasma etching.

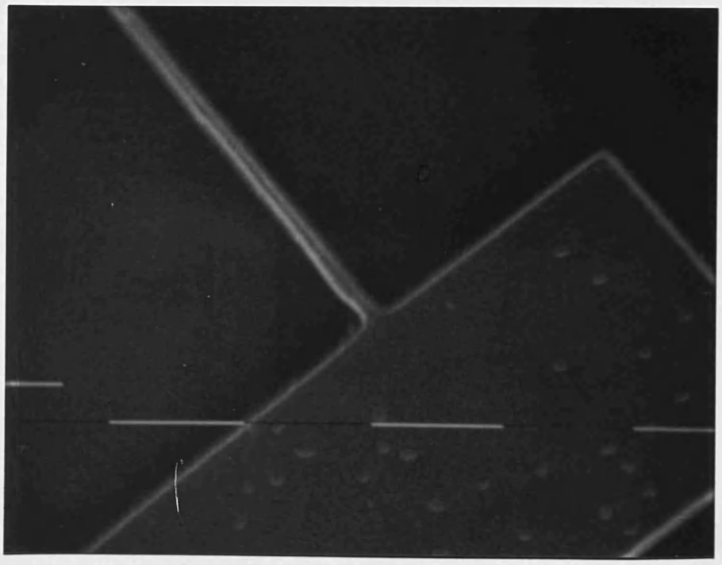


Figure (4-16) Electron micrograph showing a 0.15 μm polysilicon line connecting with a 3 μm track. The pattern was etched using chlorine (markers=1um).

Chamber pressure (mtorr)	r.f. power (W)	d.c. bias (V)	polysilicon etch rate (Å/min)	HRN etch rate (Å/min)	SiO_2 etch rate (Å/min)
10	100	400	490	380	-
10	50	280	300	330	-
10	25	200	200	100	30
50	25	70	60	12	5

Table (4-2) Table showing measured etch rates for $SiCl_4$ RIE as a function of r.f. power and chamber pressure. (d.c. bias voltages also shown).

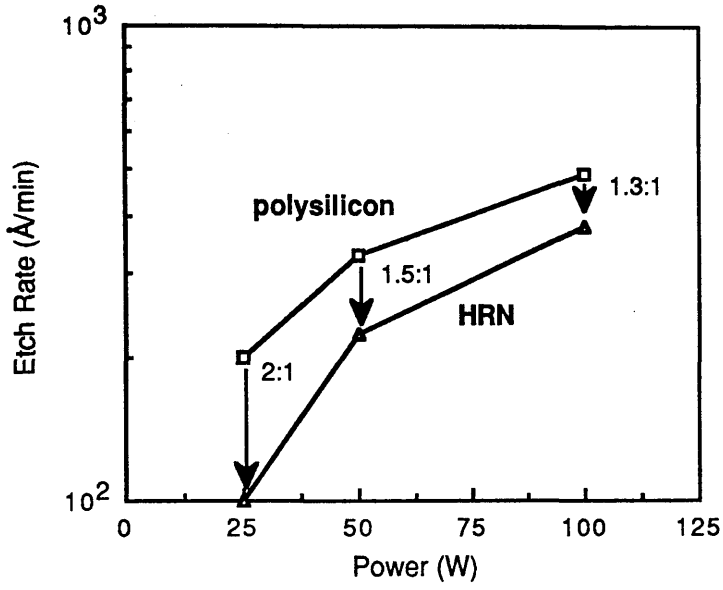


Figure (4-17) Graph showing etch rates for $SiCl_4$ RIE as a function of r.f. power level. Etch rates are indicated for polysilicon and HRN.

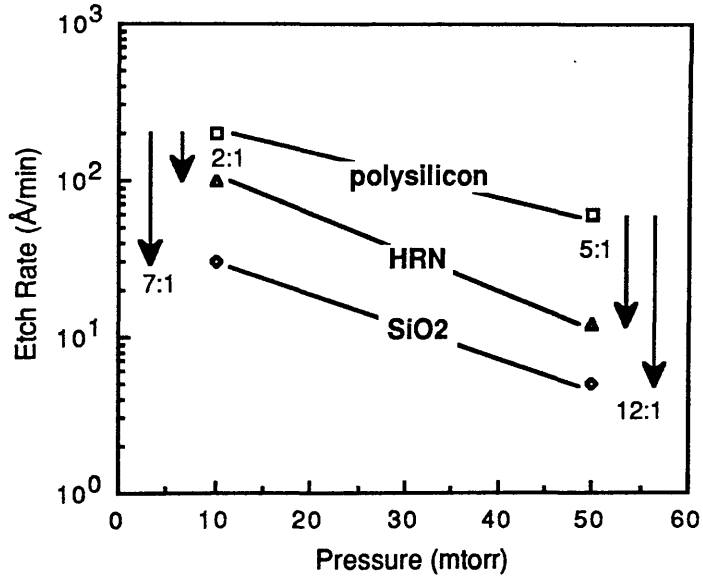
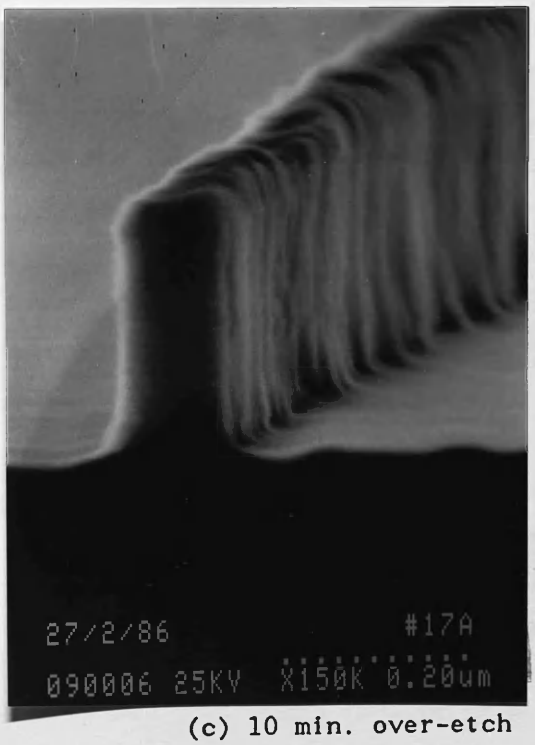


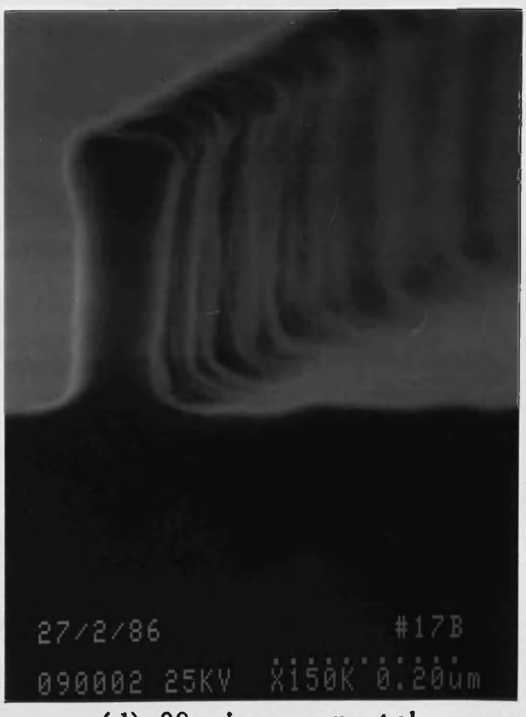
Figure (4-18) Graph showing etch rates for $SiCl_4$ RIE as a function of chamber pressure. Etch rates are indicated for polysilicon, HRN and silicon dioxide.

(a) No over-etch

(b) 5 min. over-etch.



(c) 10 min. over-etch



(d) 20 min. over-etch.

Figure (4-19) Set of four electron micrographs showing etched polysilicon profiles for $SiCl_4$ RIE Over-etch times of zero, 5, 10 and 20 min. at 50mtorr were used. In each case the HRN was 0.1 microns.

Chapter 5

Selection of Critical Process Parameters

5.1 Introduction

This chapter describes the critical n−MOS processing steps which are required for the limited scaling approach described in chapter 3. A matrix of processing parameters was chosen for these steps and suitable implant parameters were selected using the results of SUPREM II process simulations.

5.2 Critical Device Structure

The full processing sequence required to fabricate the devices is described in chapter 6. However, only the steps associated with the formation of the gate, channel and drift regions are of critical importance for device scaling. Other necessary, but non critical, steps are associated with device isolation, contact formation, interconnect metalisation and electron beam alignment. The critical steps are listed below:

(1) Gate oxidation.
(2) Channel implantation.
(3) Polysilicon deposition and implantation.
(4) High temperature anneal.
(5) Gate patterning.
(6) Source/drain implantation.
(7) Rapid Thermal Anneal

The significance of most of these steps has been described earlier. Several require further explanation.

The high temperature anneal step, which follows the polysilicon deposition and implantation is required both to anneal the gate oxide

and to activate all earlier implants. It may lead to significant dopant re–distribution in the channel region which could affect device operation.

The etched gate profile is especially important since the gate is used as a self–aligned implant mask during the formation of the critical source/drain drift regions.

Finally, the rapid thermal anneal step (R.T.A.) is required to activate source/drain implants. This step should minimise any dopant redistribution, either downwards, or laterally under the gate.

5.3 Process Matrix

A period of approximately 18 months was required to fully process one batch of wafers. Consequently, a processing matrix was chosen to investigate a range of critical processing parameters using a single wafer batch. The risk that one poorly chosen or poorly executed process step might have ruined the whole investigation was acknowledged.

A processing matrix was chosen with various gate oxide thicknesses (t_{ox}) and channel doping levels (N_A) and where possible, parameters were selected in accordance with scaling considerations.

Table (5–1) shows the scaled device examples from chapter 2, which are all predicted to exhibit long channel behaviour. CE1 is a 0.4 micron scaled from a LATV device using constant electric field (CE) scaling. CP1, CP2 and CP3 are 0.4, 0.3 and 0.2 micron devices again scaled from a LATV device but using constant field pattern (CP) scaling. GS1, GS2 and GS3 are 0.45, 0.25 and 0.15 micron devices predicted by the generalised long channel appoach (GL). The devices require oxide thicknesses from 150 to 40 angstroms.

Prior to this study, oxide thicknesses to a minimum of 300 angstroms had been grown and characterised at B.R.T.L. In response to the need for thinner oxides, a series of oxide growth experiments were undertaken at B.T.R.L. to grow thicknesses down to 150 angstroms and more recently to 100 angstroms. Gate oxide thicknesses (t_{ox}) of 300 and 150 angstroms were chosen for the process matrix. It is acknowledged that these oxide thicknesses fall

short of the scaling requirements.

Device processing with extremely thin oxides presents several problems. Firstly, the oxide must be grown uniformly, avoiding any pin–holes. Secondly, the oxide must be grown under conditions which minimise the interface states, the fixed oxide charge and the mobile ionic charge. Finally, the gate etching step must be able to stop at the thin oxide without etching through it. Together, these factors make device fabrication with thin oxides an extremely complex issue.

Six channel doping levels (N_A) in the range $3x10^{16}$ to $1.2x10^{18}$ atoms/cm^3 were chosen in accordance with the scaling examples and with reference to SUPREM simulations (see section 5.4.2). The resulting process matrix, which shows both oxide thickness and channel doping level is shown in figure (5–1). The scaling examples are also shown.

A constant source/drain junction depth (r_j) of approximately 0.1–0.12 microns was chosen, as a compromise between scaling requirements and parasitic source/drain resistance effects.

Due to the extended time scales involved in device fabrication, it was considered too risky for the whole investigation to depend upon the successful processing of a single wafer batch. Consequently, a second wafer batch was processed at the Edinburgh Microfabrication Facility. These later devices were fabricated using a new mask set, designed at Glasgow, and a slightly modified fabrication process. The new mask design included the implementation of ring oscillator circuits and therefore significantly extended the scope of the investigation. The critical processing steps are described next for both fabrication runs.

5.4 Early Devices (B.T.R.L.)

The earlier devices were fabricated with a device structure as shown in figure (5–2). The sub 0.5 micron gates were etched using the chlorine planar plasma etching process which produced a desirable rectangular gate profile as described further in chapter 7. The gate oxidation step is described first.

5.4.1 Gate Oxide Growth

Gate oxidation was performed in a dry furnace at 785°C. The low oxidation temperature was chosen to ensure long oxidation times and therefore good uniformity and reproducibility. Formation of thin oxides at this temperature has been previously reported by Irene[5.1]. Whilst rapid thermal processing techniques have been shown recently to produce extremely high quality thin oxides[5.2], suitable facilities were not available for this project.

Experiments undertaken at B.T.R.L. established that nominal oxide thicknesses of 300, 150 and 100 angstroms could be obtained for oxidation times 990, 300 and 120 minutes at 785°C. The thin oxides were measured by ellipsometer and determined to be 315, 150 and 105 angstroms. There was no significant variation in measured oxide thickness for thin oxides grown in B.T.R.L. and E.M.F. furnaces under nominally the same oxidation conditions.

In accordance with the process matrix, device wafers were processed with nominal oxide thicknesses of 300 and 150 angstroms, using the above conditions. (Further wafers with 100 angstom oxides are currently under fabrication through continued collaboration with B.T.R.L.[5.3]).

5.4.2 Channel Implantation and Anneal

The SUPREM II simulation software package[5.4] was used to aid selection of boron channel implant parameters. For the earlier wafers, a high temperature anneal step of 10 minutes at 1000°C was chosen. Boron re-distributions were modelled both before and after high temperature annealing.

A double boron implant was chosen to obtain each of the six channel doping levels required. A shallow implant at 30keV was chosen for channel region scaling and a deeper implant at 80keV was chosen for deep punch-through suppression[5.5]. Table (5-3) shows the six pairs of boron implant doses which were chosen. In each case, a simple code is used for identifying each pair of implant parameters (CH1 to CH6).

Figure (5-3) shows SUPREM simulatation results for the six

implant conditions, when implanted through a 150 angstrom gate oxide. Each profile shows two distinct peaks corresponding to the two implant energies used. These peaks are at depths of approximately 0.1 and 0.25 microns. It should be noted that the in each case, the doping level at the oxide interface is only about 15% of the peak level.

Figure (5−4) shows the same simulated implants, but following the 1000°C anneal step. The double peaks have been reduced by diffusion. There is still a decrease towards the interface, although in this case the minimum values are approximately 50% of the peak values. The nominal doping values indicated in the process matrix are shown in the figure as dashed lines.

The optimisation of the channel doping profile is complex and will be discussed later. In general, a low surface concentration leads to a higher carrier mobility but increased short channel effects. High doping at a deeper level is required to prevent punch−through.

5.4.3 Source/drain Implant and R.T.A.

The source/drain implant and R.T.A. anneal were chosen as a compromise between two effects. From a consideration of scaling theory, a shallow junction depth (r_j) in the range 0.12 to 0.06 microns is required. However for the experimental devices, the source/drain contact window separation was not scaled, leading to very high parasitic resistance of the source drift region (R_S) and consequently to a reduced extrinsic transconductance ($g_{m(ext)}$) as shown by equation (5.1). As the junction depth is reduced, this series resistance effect becomes more significant.

$$g_{m(ext)} = \frac{g_{m(int)}}{1 + g_{m(int)} \cdot R_S} \qquad (5.1)$$

When:

$g_{m(ext)}$ = extrinsic g_m

$g_{m(int)}$ = intrinsic g_m

R_S = parasitic series source resistance

The R.T.A. requirement was satisfied using a Heatpulse anneal step. Wafers were annealed using an 8 second cycle at 950°C.

SUPREM simulation was used to model the implantation and Heatpulse anneal steps. For an arsenic implant of $2.5x10^{15}$ atoms/cm^2 at 70keV through a 150 angstrom oxide the simulation results indicated a junction depth of 0.12 microns and a sheet resistivity of 94.5 ohms/sq after annealing. A junction depth of 0.073 microns and a sheet resistivity of 130 ohms/sq were predicted for the same dose, but at an energy of 40keV. Figures (5–5a) and (5–5b) show the two respective total doping profiles. In each case, the junctions are formed where the arsenic and the earlier boron implant doping levels become equal. This can be seen in the two figures where a CH3 channel implant has been assumed. Peak arsenic concentrations of $2x10^{20}$ atoms/cm^3 are predicted.

The earlier devices were implanted with $2.5x10^{15}$ atoms/cm^2 at 70keV in accordance with the lower sheet resistivity predicted above. A small number of devices were fabricated using the 40keV implant parameters.

5.5 Later Devices (E.M.F.)

The later devices were fabricated using slightly different processing conditions. Firstly the high temperature anneal step was increased to 20 minutes at 1050°C resulting in further diffusion of the channel dopant and an increased doping at the oxide interface. Secondly, the silicon tetra–chloride dry etching process was introduced to obtain a tapered gate profile (see section 4.5.2.2). The taper results in a slightly graded implant masking effect at the gate edges which is likely to lead to a reduced effective junction depth (r_j) near the channel.

A polysilicon oxidation step was included to grow a 200 angstrom sidewall oxide onto the gate structure. This refinement was included to reduce the gate to source overlap capacitance and was performed prior to the self–aligned implant. The structure of the later device is shown in figure (5–6).

5.5.1 Process Matrix for later devices

A considerably reduced process matrix was used for the later devices, with only three different channel doping levels of $1.2x10^{17}$, $3x10^{17}$ and $6x10^{17}$ atoms/cm^3. Oxide thicknesses of 150 and 300 angstroms were again used, as indicated in table (5– 4).

5.5.2 Channel Implantation and Annealing

Figure (5– 7) shows the six simulation profiles for boron channel implants after the 20 minute 1050^{o}C anneal step. Considerable diffusion is predicted under these conditions, resulting in a uniform doping concentration to a depth of 0.2 microns. For each profile, the interface concentration is now approximately 90 percent of the peak value. Once again, the desired channel doping levels are indicated with dashed lines. It should be noted that only the CH3, CH4 and CH5 implant parameters were used for device fabrication.

5.5.3 Source/Drain Implant

No Heatpulse anneal facilities were available for these wafers, and so an unconventional rapid furnace anneal was used. This technique was developed at the E.M.F. to simulate a R.T.A. step and was performed by loading a single wafer onto a quartz carrier and inserting it into a furnace tube at 1000^{o}C for several seconds[5.6]. A quartz carrier of low thermal mass was used, to minimise the wafer heat– up time.

SUPREM was used to simulate the dopant re– distribution occuring for various anneal times. Each simulation included CH3 channel implants, together with an arsenic implant of $2.5x10^{15}$ atoms/cm^2 at 70keV. A 300 angstom gate oxide was used for the simulation since it corresponds approximately to the thickness of a 150 angstrom gate oxide after the sidewall oxidation step. The predicted junction depths and sheet resistivities are shown in table (5– 5) for a range of anneal times. It can be seen that no significant

re−distribution is predicted for anneal times of less than 60 seconds. Consequently a time of 20 seconds was chosen for the device wafers. Allowing for a short warm−up period, this corresponded to 30 seconds in the furnace. A junction depth of 0.1 microns and a sheet resistivity of 90.2 ohm/sq were predicted.

5.6 Device Coding

A simple device processing code has been chosen to aid rapid identification of the processing conditions. The code is defined in table (5−6) and distinguishes the process route, channel doping level (N_A) and gate oxide thickness (t_{ox}). The code is used extensively for the device results of chapters 7 to 9.

5.7 Summary

Process parameters have been chosen for gate oxidation, channel and source/drain implants, together with the required anneal steps.

Two device designs have been described. The later devices are predicted to have a more uniform channel doping profile, a reduced effective source/drain junction depth. Process simulation results have been presented for the critical implantation steps.

References

5.1 E.A. Irene, 'Silicon oxidation studies, some aspects of the initial oxidation regime,' J. Elecrochem Soc., vol 125, p1708, 1978.

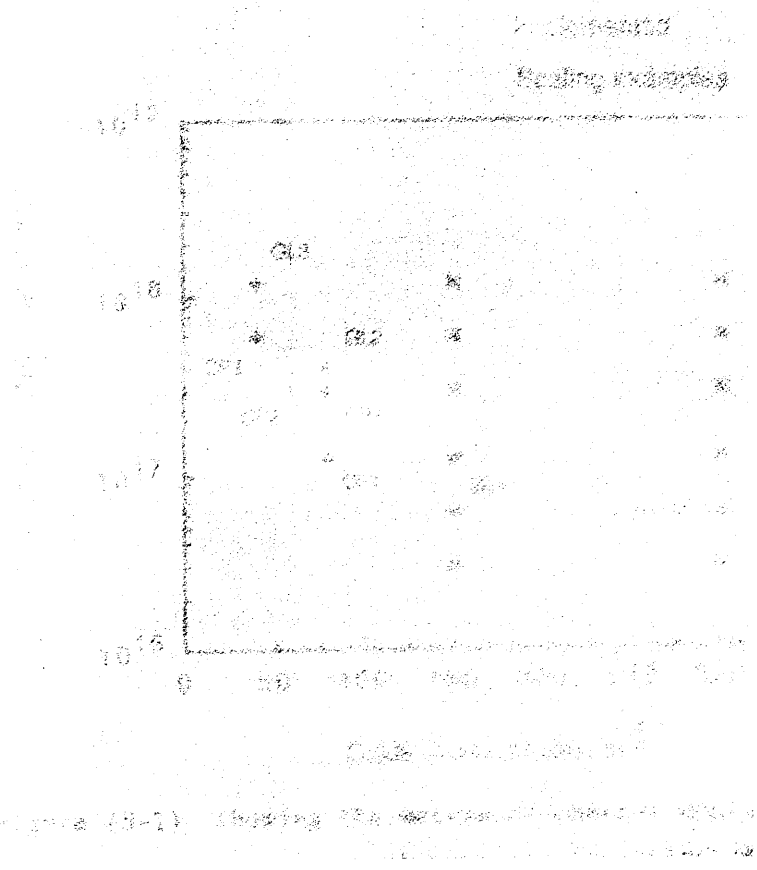
5.2 J. Nulman, J.P. Krusins and A. Gat, 'Rapid thermal processing of thin gate dielectrics,' IEEE Electron Device Lett., vol EDL−6, p205, 1985.

5.3 Wafer batch 3256A − currently under fabrication at B.T.R.L.

5.4 D.A. Antoniadis, S.E. Hansen, R.W. Dutton, 'SUPREM II – a program for IC process modeling and simulation,' Stanford Electronics Laboratories Technical Report #5019–2, Stanford University, June 1978.

5.5 H. Nihira, M. Konaki, H. Iwai and Y. Nishi, 'Anomalous drain current in n–MOSFETs and its suppression by deep ion implantation,' IEEE Tech. Dig., Int. Electron Device Meeting., p487, 1978.

5.6 Discussions with A. Gundlach at the E.M.F. regarding rapid furnace anneal techniques and many other aspects of processing are gratefully acknowledged.



Device Code	Channel length L (um)	Gate oxide thickness t_{ox} (Å)	Channel Doping N_A (atm/cm^3)	S/D junc. depth r_j (um)
CE1	0.4	80	$1.2x10^{17}$	0.12
CP1	0.4	80	$3x10^{17}$	0.12
CP2	0.3	60	$3x10^{17}$	0.08
CP3	0.2	40	$6x10^{17}$	0.06
GL1	0.45	150	$1.2x10^{17}$	0.1
GL2	0.25	80	$4x10^{17}$	0.1
GL2	0.15	40	$1.2x10^{18}$	0.1

Table (5-1) Table showing devices which are predicted to exhibit long channel behaviour. (see chapter 2).

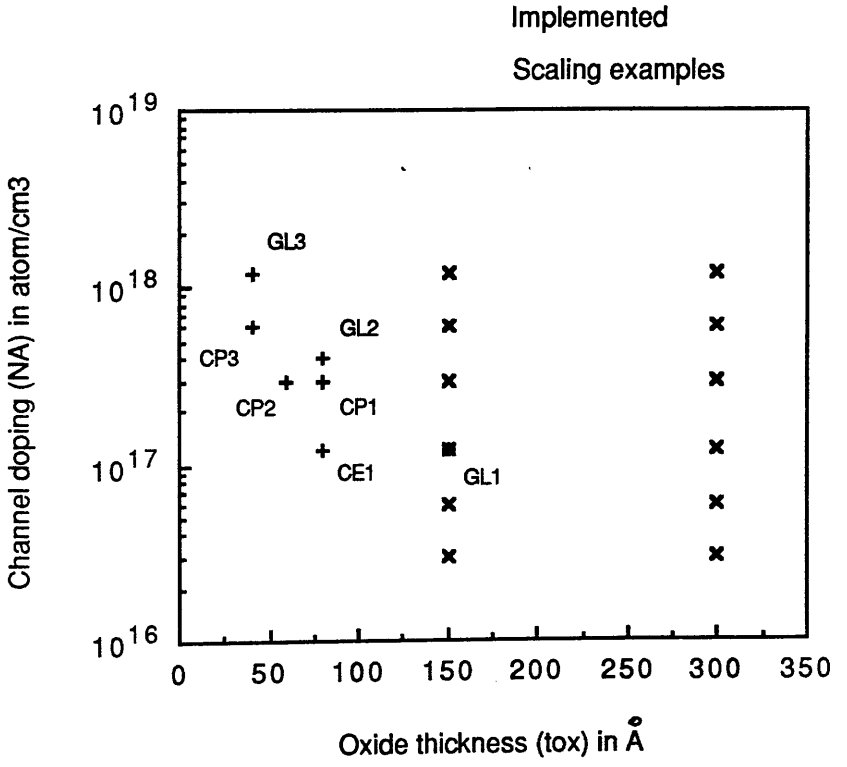


Figure (5-1) Showing the matrix of channel doping levels and oxide thicknesses investigated in this study.

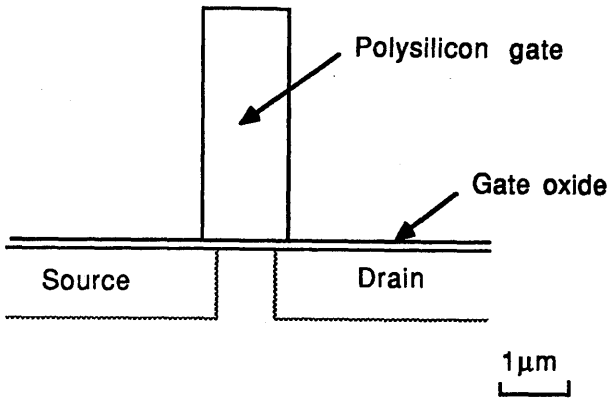


Figure (5-2) Early submicron device (B.T.R.L.)

Nominal thickness in Å	Oxidation time in min	Measured oxide thickness obtained by Ellipsometer in (Å)
300	990	315 ±5
150	300	150 ±5
100	120	105 ±5

Table (5-2) Oxide thicknesses for dry oxidation at 785°C meaured by ellipsometer.

Channel implant code	Nominal channel doping (atm/cm^3)	Boron implant parameters
CH1	$3x10^{16}$	$5x10^{11}$ atm/cm^2 at 30keV & $5x10^{11}$ " " at 80keV
CH2	$6x10^{16}$	$1x10^{12}$ atm/cm^2 at 30keV & $1x10^{12}$ " " at 80keV
CH3	$1.2x10^{17}$	$2x10^{12}$ atm/cm^2 at 30keV & $2x10^{12}$ " " at 80keV
CH4	$3x10^{17}$	$5x10^{12}$ atm/cm^2 at 30keV & $5x10^{12}$ " " at 80keV
CH5	$6x10^{17}$	$1x10^{13}$ atm/cm^2 at 30keV & $1x10^{13}$ " " at 80keV
CH6	$1.2x10^{18}$	$2x10^{13}$ atm/cm^2 at 30keV & $2x10^{13}$ " " at 80keV

Table (5-3) Table of boron channel implant parameters.

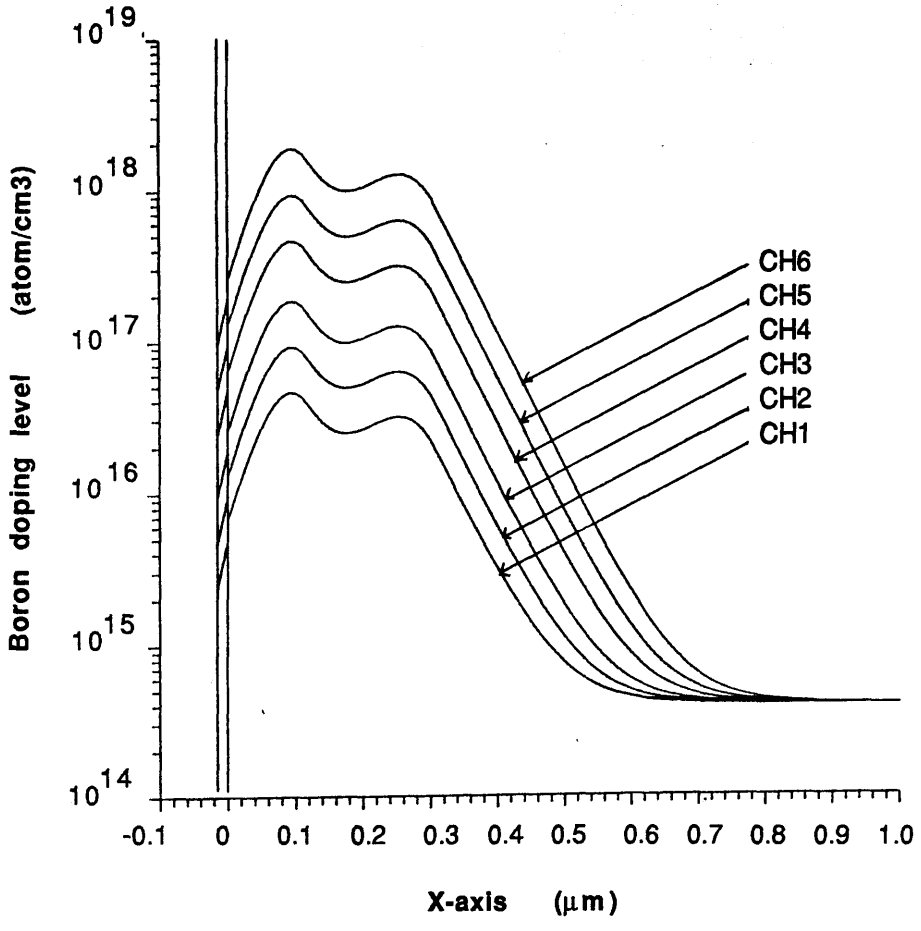


Figure (5-3) SUPREM simulation showing the CH1-CH6 boron channel implants before annealing (t_{ox}=150A).

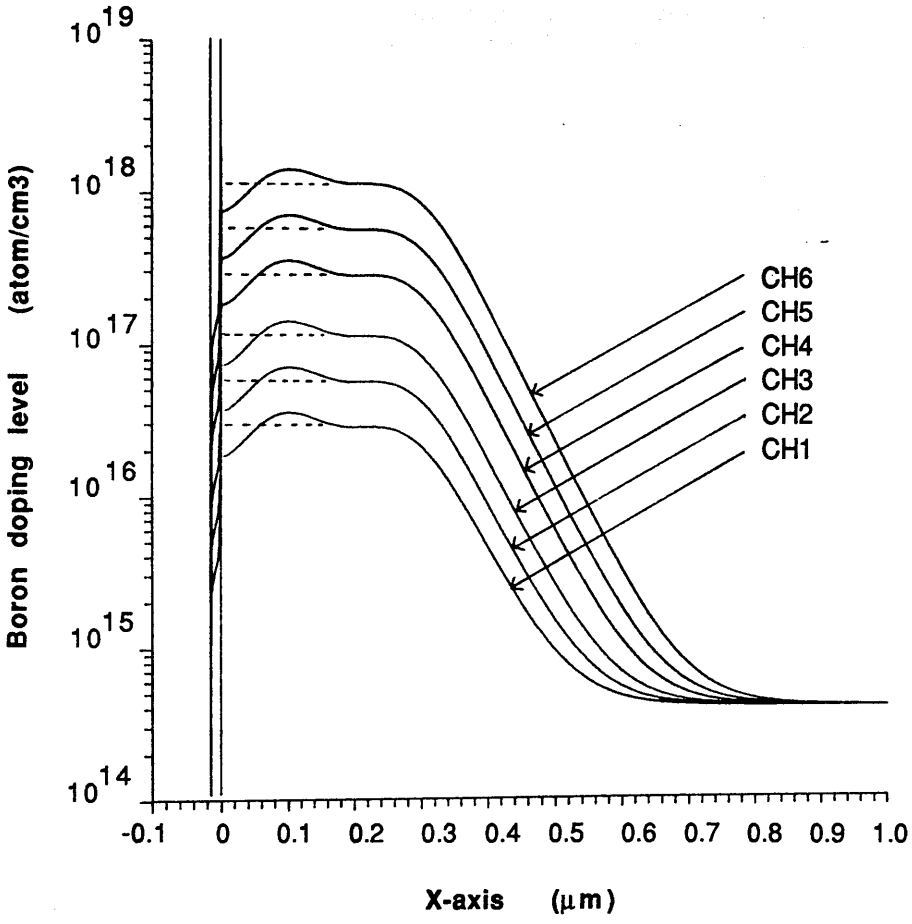
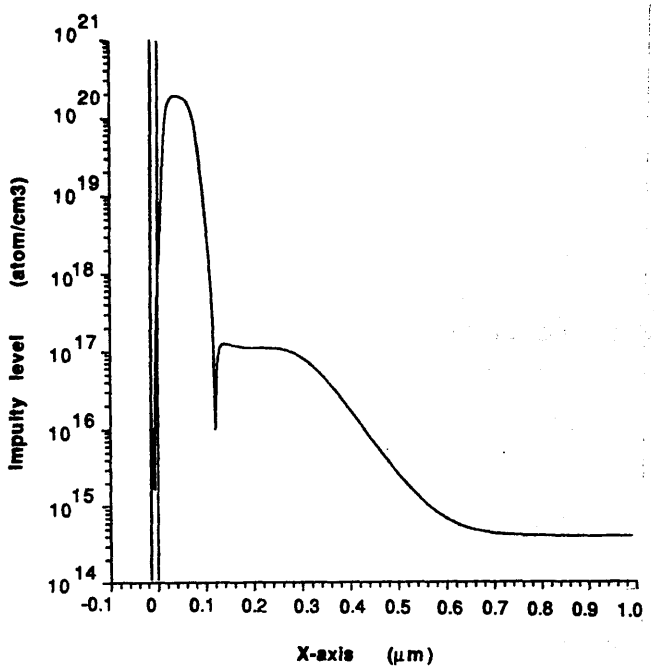
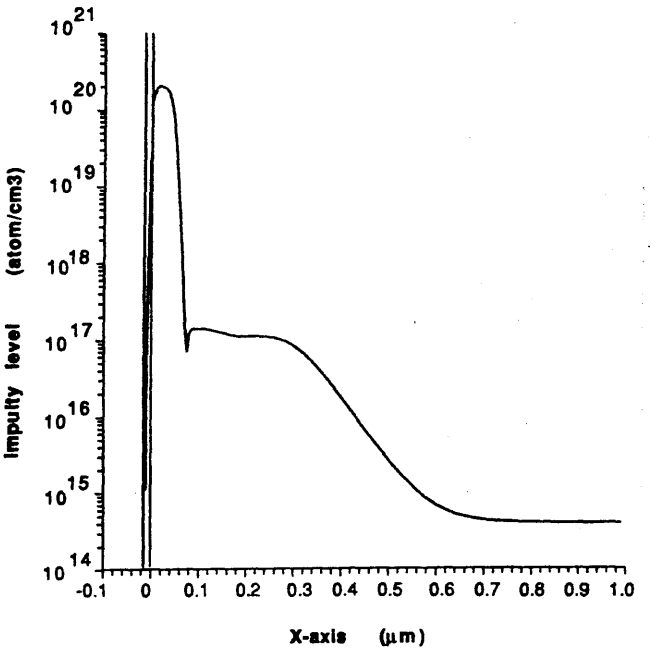


Figure (5-4) SUPREM simulation showing the CH1-CH6 boron channel implants after a 1000°C anneal for 10 minutes.



(a) As 2.5×10^{15} at 70keV.



(b) As 2.5×10^{15} at 40keV.

Figure (5-5) SUPREM simulation showing total impurity profiles for source/drain implanted wafers after R.T.A. (a) is for an arsenic dose of 2.5×10^{15} atoms/cm^3 at 70keV and (b) is for an arsenic dose of 2.5×10^{15} atoms/cm^3 at 40keV. In each case a CH3 channel implant has been included.

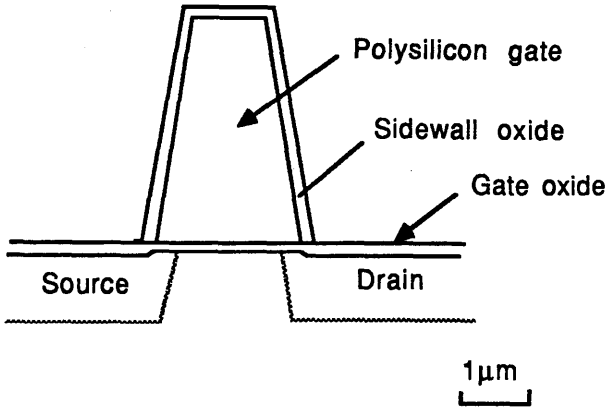


Figure (5-6) Later submicron device (E.M.F.).

Channel implants	Oxide thickness (Å)	
	150	300
CH3	x	x
CH4	x	
CH5	x	

x = implemented

Table (5-4) Showing matrix of channel implants and oxide thicknesses used for later device fabrication (E.M.F.)

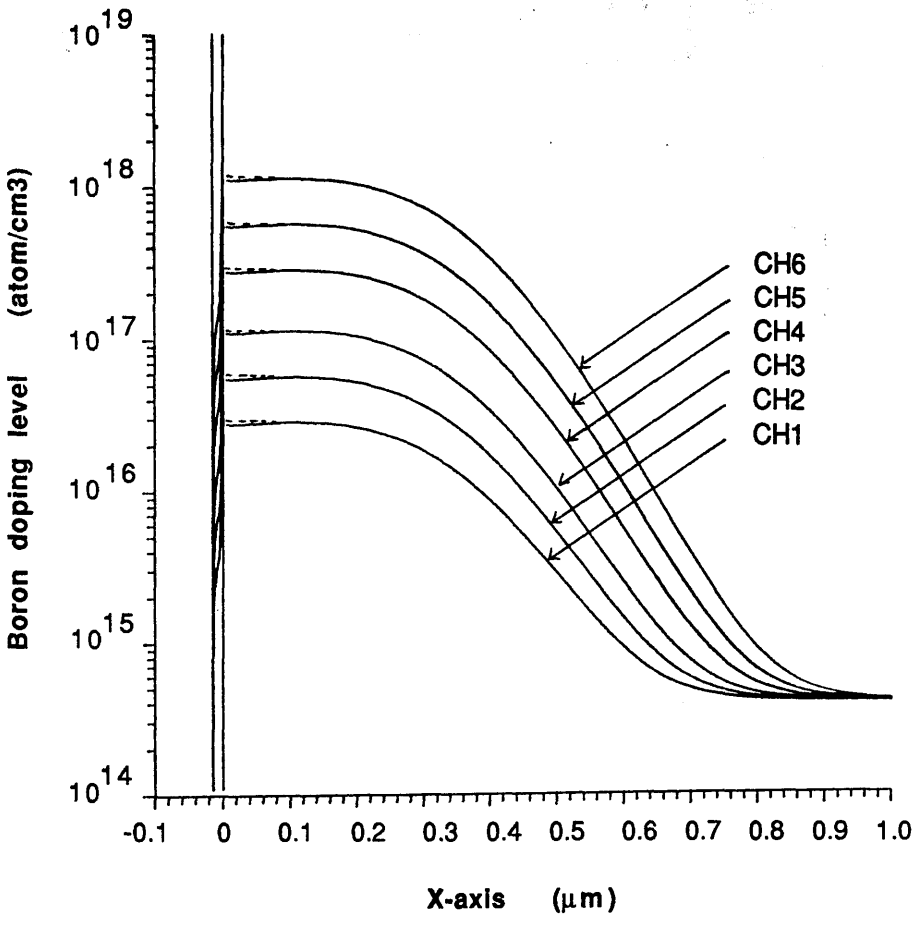


Figure (5-7) SUPREM simulation showing the CH1-CH6 boron channel implants after a 1050°C anneal for 20 minutes.

Asrenic implant parameters	Anneal time (sec)	SUPREM junct. depth (μm)	SUPREM sheet resistivity (Ohm/sq)
2.5×10^{15} atm/cm^2 at 70keV	20	0.103	90
" "	40	0.104	87
" "	60	0.104	84
" "	120	0.109	77
" "	240	0.125	69

Table (5-5) Showing junction depth and sheet resistivity data from SUPREM simulation, for various anneal times at 1000°C.

EXAMPLE DEVICE CODE :- ED-CH3-OX150

Process route:
- BT for early devices
- ED for later devices

Channel implants:
- CH1-CH6
- (see table 5.3)

Gate oxide thickness
- OX150 for 150Å
- OX300 for 300Å

Table (5-6) Device processing code.

Chapter 6

The MOSFET Process Sequence

6.1 Introduction

This chapter describes the n−MOS processing which was used to fabricate the experimental devices. This processing incorporates the critical steps and the matrix details as described in chapter 5.

6.2 Process Overview

The same approach to processing was used for the B.T.R.L. processed wafers and the later E.M.F. processed wafers. In each case, wafer batches were fabricated up to the gate lithography stage and then sent to Glasgow for high resolution patterning. The initial processing also included the formation of suitable registration marks for the Glasgow E.B.L. system. After completion at the source facility, the wafers were returned to Glasgow for testing.

Each wafer was composed of a matrix of chip fields, with each field containing a number of discrete MOSFETs and device arrays. The B.T.R.L. wafers were processed using an existing mask set (423AEG) and the E.M.F. wafers were processed using a new mask set (EU567), designed at Glasgow.

The B.T.R.L. mask set consisted of wide gate inter−digitated devices and arrays of test devices with decreasing gate length (see section 6.4). The mask set included enhancement devices only.

The new E.M.F. mask set was designed with an additional mask level for patterning depletion devices, which are usually required for high speed n−MOS digital circuits[6.1]. Consequently it included all the features of the B.T.R.L. mask set, for both enhancement and depletion mode devices. The mask also included simple n−MOS inverters and 19 stage unloaded ring oscillator circuits. The latter are of special interest when establishing the ultimate switching speed for an experimental device technology. However, it is acknowledged that a device operating in a loaded circuit configuration is unlikely to

reach this limit.

6.3 The Process Sequence

The processing sequence required to fabricate the experimental devices can be divided into seven principal stages:

(1) LOCOS isolation of active areas.
(2) Gate oxide growth and channel implantation.
(3) Gate formation.
(4) Drift region formation.
(5) Dielectric deposition.
(6) Contact window formation.
(7) Interconnect metalisation.

These processing stages are described firstly for B.T.R.L. device wafers, then in modified form for E.M.F. device wafers. Figure (6−1) summarises the B.T.R.L. process which is described next.

6.3.1 LOCOS Isolation of Active Areas

P− type three inch silicon wafers, with a (100) crystal orientation and a nominal resistivity of 14−20 ohm−cm, were initially degreased and cleaned. A LOCOS mask was formed by growing a 500 angstrom silicon dioxide buffer layer and depositing a 700 angstrom silicon nitride layer (see figure (6−2a)). The mask was then used in an unconventional two stage oxidation process whereby the bond pad and field oxide were formed independently for the reasons described earlier (see section 3.4). A resist pattern was defined by optical lithography, leaving only the bond pad regions unprotected. Next, the silicon nitride was dry etched using a CF_4 plasma and the residual resist was removed. A bond pad oxide was grown to a thickness of 2 microns to obtain the structure shown in figure (6−2b).

A second resist pattern was optically defined, leaving only the active device areas protected, and again the exposed silicon nitride was dry etched, this time in preparation for the field oxidation step.

Prior to growing the oxide, a boron field implant was performed to raise the threshold voltage of any parasitic MOS devices formed where polysilicon tracks traverse the field oxide. A low parasitic threshold voltage could lead to carrier inversion under the field oxide which would ruin device isolation. The residual resist was removed and the wafers were introduced into a wet oxidation furnace to grow a 0.85 micron field oxide (see figure (6−2c)). The final LOCOS structure which is shown in figure (6−2d), was obtained by wet etching the silicon nitride and underlying buffer oxide.

6.3.2 Gate Oxide Growth and Channel Implantation

Next, the critical gate oxides were grown using a dry oxidation furnace at 785^{o}C. Thicknesses of 150 and 300 angstroms were obtained, in accordance with the processing conditions and process matrix described in chapter 5. The wafers were then implanted with boron to raise the channel doping level. Six pairs of channel implant parameters (CH1−CH6) were used to obtain the matrix of channel doping levels from $3x10^{16}$ to $1.2x10^{18}$ atoms/cm^3, as shown earlier in table (5−3).

6.3.3 Gate Formation

The gate formation stage required the patterning of suitable registration marks to permit wafer alignment in the Glasgow E.B.L. system. The alignment and exposure strategy is described next, and the gate processing details follow.

6.3.3.1 Alignment and Exposure Strategy

The gate level alignment and exposure strategy for the Glasgow E.B.L. system was chosen in accordance with several considerations.

Firstly, accurate and reproducible gate patterning down to 0.1 microns was required across 73mm (3 inch) wafers. This was necessary to ensure correct interpretation of scaling effects for the

resulting devices. Consequently, an exposure frame of 100 x 80 microns was chosen, in accordance with the 0.1 micron patterning results of chapter 4. Since the B.T.R.L. and E.M.F. chip fields, were respectively 1.5mm square and 2mm square, the chosen frame size required separate patterning of each discrete device and device array within each chip field. This required reduced field registration marks at each device site.

Secondly, linewidth reproducibilty was especially important since no simple non-destructive method was available for directly measuring the gate length of completed devices once concealed under a dielectric layer. Precise focussing and current monitoring of the electron beam before pattern exposure ensured linewidth reproducibility to within 10% for 0.1 micron patterns. This permitted test lines to be patterned onto the wafers for linewidth calibration purposes. The lines were exposed using identical parameters to those of the actual devices, and they were arranged to permit cleaving and profile examination by SEM. To minimise any effects resulting from non-uniform etching across the wafer, all devices and calibration lines were patterned in a region within 20mm of the wafer centre.

Thirdly, a suitable technique was required to ensure correct alignment of the E.B.L. exposure frame to each device site. Positional accuracy to within 0.25 microns was necessary. However, the goniometer stage of the E.B.L. system resulted in position errors of up to 15 microns. Using a single step alignment technique, this large error, together with the small exposure frame size, would leave only a very limited region for pattern exposure. Therefore, a two step alignment technique was chosen, whereby a 400 x 300 micron frame size permitted course alignment to a remote registration feature. Then, fine alignment could be achieved at the desired exposure frame size, using the reduced field registration marks.

Finally, there was an incompatibility between the two MOS processing facilities which use 73mm (3inch) wafers, and the Glasgow E.B.L. system which permitted maximum sample width of only 60mm, as determined by the width of the sample admission gate valve. This rather unfortunate problem was overcome by sawing the wafers to 59mm.

The resulting alignment and exposure strategy is described next. Firstly, the sawn wafers were mounted onto an adjustable wafer chuck

and then loaded into the E.B.L. system. These wafers had both remote and reduced field registration marks etched into the polysilicon gate layer and were coated with HRN. The wafer chuck designed with adjustable arms for supporting sawn wafers is shown in figure (6– 3) and was constructed from non– magnetic brass to avoid any undesirable interaction with the beam. Once in the E.B.L. system, the wafer was tilted perpendicularly to the electron beam and the beam was set to the standard parameters for 0.1 micron patterning (see chapter 4). Next, under manual goniometer control, the wafer was moved to a sacrificial device site, where it was rotated to align the exposure frame to the chip field, achieving global rotational alignment. The goniometer was then moved to align the beam into the chuck mounted Faraday Cup and the current was measured.

Gate level patterning was then undertaken, chip field by chip field, across a selected region of the wafer. The exposure of each chip field required re– focussing of the beam, followed by alignment and patterning at each subsequent device site. Typically twelve chip fields were exposed in this way, with a return to the Faraday cup after every third chip field. This sequence is show in figure (6– 4) and afterwards, the calibration lines were patterned.

A simple sequence was used for device patterning within each chip field. Firstly, the electron beam was focussed at the centre of the chip field using an exposure frame size of 6.3 x 5 microns. The goniometer was then moved to the first device site and a frame size of 400 x 300 microns was selected for course alignment. An alignment pattern was then scanned over the remote registration feature. This feature was located at an adequate distance from the device site to avoid accidental exposure of the active region during alignment. At this frame size, an initial alignment accuracy of 2 microns was achieved by movement of the goniometer. Next, the exposure frame size of 100 x 80 microns was selected and fine alignment performed by scanning patterns over the reduced field registration marks close to the device site. Beam shifting X and Y controls were then used for fine alignment, and an accuracy of better than 0.25 microns was achieved. At the first device site, separate controls for fine X and Y axis magnification correction were also adjusted to ensure a correct frame shape. Figure (6– 5) shows the

two step alignment procedure. The registration marks, in the form of a cross and a series of squares can be seen. Once alignment had been completed, the appropriate gate level pattern was exposed.

6.3.3.2 Gate Processing

Once the gate oxide had been grown and the channel implants performed, a 0.35 micron polysilicon film was deposited and a 400 angstrom polysilicon oxide grown for later use as a dry etch mask (see figure (6−6a). The oxide layer was required for later use as a dry etch mask. The polysilicon sheet resistivity was reduced using a low energy, high dose phosphorous implant step. A dose of $6x10^{15}$ atoms/cm^2 at 40keV was chosen for this purpose.

The wafers were then annealed at high temperature to activate all implants and to anneal the gate oxide, minimising surface states and reducing implantation damage. An anneal temperature of 1000°C for 10 minutes was chosen, as described in chapter 5.

Since the B.T.R.L. mask set did not include a level for the reduced field registration marks, a medium resolution E.B.L. system at B.T.R.L. was used. The wafers were coated with XXL positive resist and then patterned by electron beam. A 1mm square exposure field was used to pattern each chip field with the required registration marks.

The polysilicon oxide layer was used as an intermediate mask for a two step dry etching process to transfer the XXL resist pattern into the polysilicon. Direct use of XXL as a mask was undesirable due to the relatively high dry etch rate of positive resists. A CHF_3 plasma was used to transfer the pattern into the polysilicon oxide, and a Cl_2 plasma was used to etch the polysilicon.

Once the registration marks had been etched, the residual resist and the polysilicon oxide were removed, (see figure (6−6b)). The wafers were sawn to 59mm and then sent to Glasgow for high resolution gate patterning.

On arrival, the wafers were inspected and the HRN resist was spun using the standard preparation conditions to obtain a thickness of 0.36 microns. Each wafer was exposed in the Glasgow high resolution E.B.L. system, using the above strategy together with standard

exposure parameters. Discrete devices were patterned with gate lengths from 0.5 to 0.1 microns and the device arrays were patterned with gate lengths from 3.0 to 0.1 microns. Between 200 to 400 devices were patterned onto each wafer and then developed using standard conditions. The wafers were inspected and returned to B.T.R.L.

As described in chapter 4, the gate patterns were plasma etched in pure Cl_2 using a 15% over-etch time, and the residual resist was stripped by oxygen plasma.

6.3.4 Drift Region Formation

The source and drain drift regions were formed by a self-aligned arsenic implantation step, resulting in the structure shown in figure (6-6c). In accordance with chapter 5, an implant dose of $2.5x10^{15}$ atoms/cm^2 at 70keV was chosen.

6.3.5 Dielectric Deposition

A dielectric layer was required to isolate the interconnect metalisation and polysilicon level. For this purpose, a dielectric layer of highly doped glass, is usually deposited[6.2]. The high doping, typically of phosphorous or arsenic, enables the glass to be reflowed at high temperature to produce a semi-planar topography.

Arsenic doped glass (As-glass) was chosen for these devices in preference to phosphorous doped glass (P-glass), which is more commonly used. This choice was made because As-glass has a more conformal deposition mechanism resulting in improved step coverage. The glass was deposited to a thickness of 0.5 microns and then nitrogen annealed for 1 hour at 550°C. This step, together with a later high temperature anneal were required to densify the glass layer.

6.3.6 Contact Window Formation

Next, contact window patterns were optically defined into resist

and the unprotected As–glass was etched by CHF_3 plasma. Windows were cut to the source/drain drift regions and the polysilicon gate level. In the case of the drift regions, the thin gate oxide was also etched.

Once the residual un–eroded resist had been stripped, a phosphorous implant was introduced through the contact windows to increase the junction depth and thereby suppress junction spiking as described in chapter 3. A double implant of two identical doses of $3x10^{15}$ atom/cm^2 at 80kev and 180keV was chosen. A resulting juntion depth of approximately 0.5 microns was predicted by SUPREM II.

Next, the wafers were annealed by Heatpulse R.T.A., using an 8 second anneal cycle to 950°C, as described in chapter 5. This anneal step was required to accomplish the following:

(1) To activate the drift region implant.
(2) To activate the contact window implants.
(3) To anneal electron beam induced oxide damage.
(4) To complete the As–glass densification.

The electron beam damage indicated above, results in the formation of neutral oxide traps which gradually become charged, leading to long term threshold voltage instability[6.3]. High temperature anneal steps have been demonstrated to reduce this effect[6.4].

6.3.7 Interconnect Metalisation

Interconnect metalisation was undertaken after the R.T.A. step. Firstly, a 0.5 micron thick film of Alusil (Aluminium with 1.5% silicon) was sputter deposited. The silicon component was included to minimise electromigration, junction spiking, and silicon hillock formation[6.5]. The interconnect pattern was optically defined in resist and the metal was etched using a CCl_4 plasma. The resist was stripped and the metal contacts were sintered in forming gas for 2 hours at 350°C.

Finally, a number of non–critical etch steps were used to remove all layers from the wafer–backs in order to permit electrical

contact to the substrate. Figure (6–6e) shows the completed device structure.

6.4 The 423AEG Mask Set

The 423AEG mask set, as used for the B.T.R.L. wafer fabrication, consisted of a 1.5 x 1.5 mm chip field which contained discrete devices together with optical registration and resolution patterns. Figure (6–7) shows an optical micrograph of a fully processed chip field.

Three types of inter–digitated wide gate MOSFET were included, with total gate widths of 100, 300 and 500 microns. The design for the 100 micron wide device is shown in figure (6–8a) and a submicon gate length silicon implementation in figure (6–8b).

The device array consisting of eight MOSFETs with provision for decreasing gate length from 3 to 0.1 microns is shown in figure (6–9a). Since the gate level pattern has a total width greater than 100 microns, two separate alignment and exposure steps were required. Figure (6–9b) shows an electron micrograph of the resulting device array, with the two electrode patterns correctly stitched at the centre. These device arrays are useful for evaluating scaling effects and also for calculating device constants such as channel length reduction, low field mobility and parasitic series resistance as discussed further in chapter 7.

6.5 The Modified Process (E.M.F.)

For the later devices, several changes were made to the process; firstly the gate and channel processing was modified as described in chapter 5; secondly, several refinements associated with the new mask set were incorporated and finally, several changes were required for compatibility with process equipment at the E.M.F.

The new mask set was designed with four additional levels. Firstly, a level was introduced to mask the enhancement devices during implantation of the depletion devices. Secondly, a level was included to replace the electron beam lithography step required for

patterning the reduced field registration marks. Thirdly, a level was added for deep implantation of the drift regions, away from the gate, to prevent junction spiking and to lower the parasitic drift region to substrate capacitance. Finally, another level was included for optical exposure of the gate patterns with 3 to 1 micron gate lengths. This level facilitated rapid process development and comparisons between devices with optically and electron beam defined gates.

Since R.T.A. facilities were not available at the E.M.F. the Heatpulse step was replaced by a rapid furnace anneal as described earlier (see chapter 5). Also, due to the absense of As–glass deposition facilities, a return to the more conventional P–glass was necessary.

The modified process is indicated in figure (6–10) and described below. The majority of the modifications occur during the gate formation stage.

6.5.1 Pre–Gate Processing (E.M.F.)

The initial wafer processing was essentially the same as for the earlier wafers, utilising the same double LOCOS technique. A boron field implant of $1x10^{13}$ atoms/cm^3 at 100keV was used to obtain a doping concentration of approximately $2x10^{16}$ atoms/cm^2 under the field oxide.

Next, a resist pattern was defined with windows at the depletion device active areas and the wafers were implanted with arsenic at 40keV. Doses of $2x10^{11}$, $5x10^{11}$ and $1x10^{12}$ atoms/cm^2 were used corresponding to the CH3, CH4 and CH5 boron channel implants. After arsenic implantation, the resist was stripped and the CH3, CH4 and CH5 boron channel implants were performed as for the earlier devices.

6.5.2 Gate Formation (E.M.F.)

The gate formation stage contained many of the changes, especially the tapered gate and the increased high temperature anneal (see chapter 5). However, the method of junction spiking suppression

was also changed, by incorporating a masking step for the deep phosphorous implant. This implant was performed through the polysilicon layer, using an aluminium mask.

A 0.5 micron polysilicon film was deposited and then oxidised to form a 0.4 micron oxide buffer layer. A one micron aluminium film was then evaporated resulting in a structure as shown in figure (6–11a). The required implant masking pattern was defined in resist and two dry etch steps were used. A CCl_4 plasma was used to etch the aluminium and then a CHF_3 plasma for the polysilicon oxide.

Once the aluminium mask had been formed, the wafers were implanted with phosphorous, at a high dose and energy ($1x10^{15}$ atom/cm^2 at 150keV). The high energy was required to ensure penetration through the 0.3 microns of polysilicon which remained after oxidation. The wafers were then annealed at 1050°C for 10 minutes, thereby driving the implant deeper. Following this, and all subsequent anneal steps, a final junction depth of 0.46 microns was predicted by SUPREM II. Figure (6–11b) shows the structure following implantation.

This implantation through the polysilicon layer was chosen to avoid any contact between the aluminium mask and the critical gate oxide. Aluminium was chosen because no photoresist was available which could withstand a high dose implant.

After the implant step, the aluminium and the polysilicon oxide were wet etched, leaving only the polysilicon layer. Next, the whole polysilicon layer was implanted to lower the sheet resistivity using the same conditions as for the earlier wafers. The wafers then received a further 10 minute anneal at 1050°C.

The wafers were sawn to 59mm and the reduced field registration marks were optically defined into resist. Next, the wafers were sent to Glasgow.

The registration patterns were etched into the polysilicon by $SiCl_4$ reactive ion etching process. The residual resist was stripped and the wafers were dipped into 10% buffered hydrofloric acid to remove any native oxide. HRN resist was spun and pre–baked using standard conditions and then the critical 0.1 micron gate exposures were made using the Glasgow E.B.L. system. The same alignment and exposure strategy was used as for the earier wafers, and once completed, the resist was developed using standard conditions.

The critical gate etching step was then performed using the two stage $SiCl_4$ reactive ion etching process. A 5 minute over–etch was used to obtain a tapered gate profile, as required for these later devices (see figure (4–19b). The HRN residual resist was then stripped using an oxygen plasma followed by immersion in fuming nitric acid. The resulting structure is shown in figure (6–11d).

Next, the wafers were returned to the E.M.F. and the polysilicon sidewalls were grown to a thickness of approximately 200 angstroms.

6.5.3 Post–Gate Processing (E.M.F.)

The drift regions were formed using the same arsenic implant as before and the P–glass was deposited to a thickness of 0.4 microns, by C.V.D. A low temperature anneal step was then undertaken for the first stage of P–glass densification. The wafers were annealed at 450°C for 20 minutes in forming gas. Next, the wafers were annealed at high temperature using the rapid furnace anneal technique in place of the Heatpulse anneal used previously. This final high temperature step was required to:

(1) Activate the shallow drift region implants.
(2) Densify P–glass (2nd stage).
(3) Reflow P–glass.

Contact windows were formed as for the earlier process, using a CHF3 plasma. However, the phosporous contact window implant was omitted since it had been replaced by the earlier deep phosphorous implant step.

Interconnect metalisation was achieved as before, producing the final structure as shown in figure (6–11e). The completed wafers were returned to Glasgow for testing.

6.6 The EU567 Mask Set

The EU567 mask set was designed at Glasgow using GAELIC multiple level mask design software. The design was implemented onto two reticles for the E.M.F. Optimetrix 10:1 reduction wafer stepper. All light field levels were implemented onto one reticle and the dark field levels onto another. Figure (6–12) shows a chip field fabricated using the mask set. Inter–digitated devices with gate widths of 100 microns and device arrays of eight devices were included. Figure (6–13a) shows the design of the device array and figure (6–13b) shows the silicon implementation.

Unloaded ring oscillator circuits were included in the design. These circuits consisted of nineteen n–MOS inverter stages connected in a ring with a three stage output buffer, as discussed further in chapter 9.

Van Der Pauw cross structures[6.6] for measuring sheet resistivities and MOS capacitances were also included, together with suitable chip field registration marks as required for the Optimetrix stepper.

A detailed layout of the mask set is included as an appendix to the thesis.

6.7 Summary

The full n–MOS processing sequence has been described, in two alternative forms, for fabricating the experimental MOSFETs. The alignment and exposure strategy developed for wafer patterning using the Glasgow E.B.L. system has also been described.

References

6.1 A.D. Milne (ed), 'MOS Devices,' Edinburgh University Press, Edinburgh, 1982.

6.2 S.M. Sze (ed), 'VLSI Technology,' Singapore, 1983.

6.3 J.M. Aitken, '1 um MOSFET VLSI technology: part VIII– radiation effects,' IEEE J. Solid–St. Circuits, vol SC–14, p294, 1979.

6.4 M. Shimaya, N. Shiono, O. Nakajima, C. Hastimoto and Y. Sakakibara, 'Electron beam induced damage in poly–Si gate MOS structures and its effect on long–term stability,' J. Electrochem. Soc., vol 126, p945, 1983.

6.5 S. Vaidya, D.B. Fraser and A.K. Sinha, 'Electromigration resistance of fine line Al,' Proc. 18th Reliability Physics Symposium, p165, IEEE, New York, 1980.

6.6 W. Versnel, 'Analysis of the Greek Cross, a Van der Pauw structure with finite contacts,' Solid State Electron., vol 22, p911, 1979

B.T.R.L.-Glasgow University Collaboration.

Submicron Si MOSFET fabrication sequence for use with 423AEG mask set.

Use 3inch wafers : P-type (100) 17-23 ohm-cm

Step	**Locos isolation of active areas**
1.1	Initial clean
1.2	Grow 500A of SiO2
1.3	Deposit 700A of Si3N4
1.4	Photolith - print mask 1 (bond pad oxide)
1.5	Plasma etch Si3N4 (in CF4)
1.6	Plasma strip resist (in O2)
1.7	Wet etch buffer oxide (10:1 buffered HF)
1.8	Grow 2um of SiO2 (bond pad oxide)
1.9	Photolith - print mask 2 (active areas)
1.10	Plasma etch Si3N4 (in CF4)
1.11	Field implant (Boron)
1.12	Plasma strip resist (in O2)
1.13	Grow 0.85um of SiO2 (field oxide)
1.14	Plasma strip nitride from active areas (in CF4)
1.15	Etch-back LOCOS field oxide
1.16	(Formation of chip field registration marks)
	Gate oxide growth and channel implantation
2.1	Pre-gate clean
2.2	Grow gate oxide (dry at 785oC)
2.3	Channel implant (Boron)
	Gate electrode formation
3.1	Deposit 0.35um of Poly-Si
3.2	Grow masking oxide on Poly-Si
3.3	Poly-Si implant (Phos.)
3.4	Neutral ambient anneal
3.5	Spin and pre-bake XXL positive e.beam resist
3.6	E.beam exposure of reduced field registration marks
3.7	Develop XXL
3.8	Wet etch Poly-Si oxide
3.9	Plasma etch Poly-Si (in Cl2)
3.10	Plasma strip resist (in O2)
3.11	Wet etch Poly-Si oxide
3.12	Spin resist,saw wafers to 59mm and strip resist
3.13	Send to Glasgow University
3.14	Spin and pre-bake HRN negative e.beam resist
3.15	E.beam exposure of gate level patterns to 0.1um
3.16	Develop HRN
3.17	Return to B.T.R.L
3.18	Plasma etch Poly-Si (15% over-etch in Cl2)
3.19	Plasma strip resist (in O2)
3.20	Rinse in de-ionised wafer

Figure (6-1) Processing sequence used for the early devices (B.T.R.L.) **(cont. overleaf.)**

Drift region formation

4.1 Source/drain implant (As)

Dielectric deposition

5.1 As-glass deposition

5.2 As-glass densification and 1st stage S/D anneal

Contact window formation

6.1 Photolith - print mask 8 (contact windows)

6.2 Plasma etch As-glass (in CHF3)

6.3 Contact window implant (Phos)

6.4 Plasma strip resist (in 02)

6.5 Heat pulse anneal - 2nd stage S/D anneal (0sec at 950oC)

Inter-connect metalisation

7.1 Pre-metal etch

7.2 Deposit 0.5um Alusil

7.3 Photolith - print mask 9 (metalisation)

7.4 Metal etch

7.5 MIT resist strip

7.6 Resist coat wafer fronts

7.7 Plasma etch wafer backs (in CF4 + 02)

7.8 Etch 2um of Si02 from wafer backs with hot buffered HF

7.9 MIT resist strip

7.10 Sinter in forming gas for 2hrs at 350oC

7.11 Send to Glasgow for testing

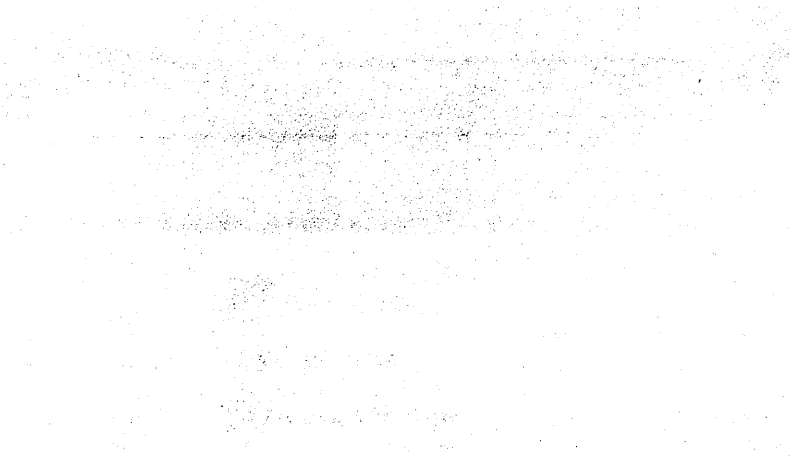
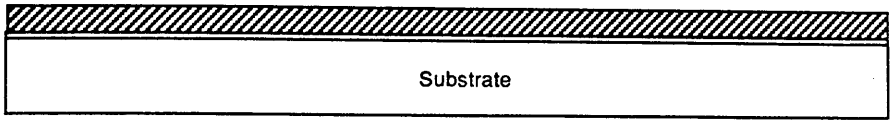
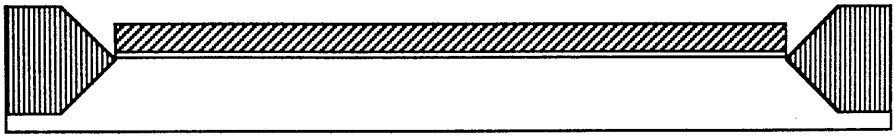


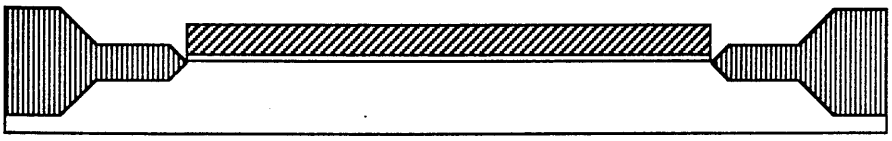
Figure (6-1) cont.



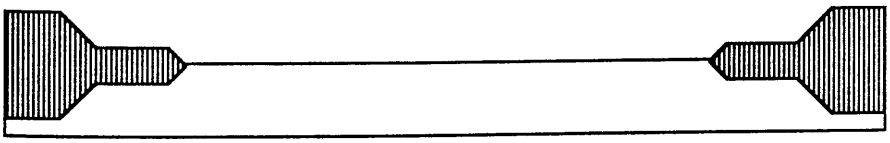
(a) Initial silicon nitride/buffer oxide mask.



(b) Structure after bond pad oxidation.



(c) Stucture following field oxidation.



(d) Final LOCOS structure.

key:

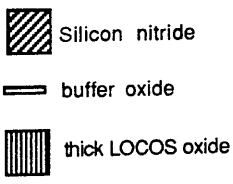
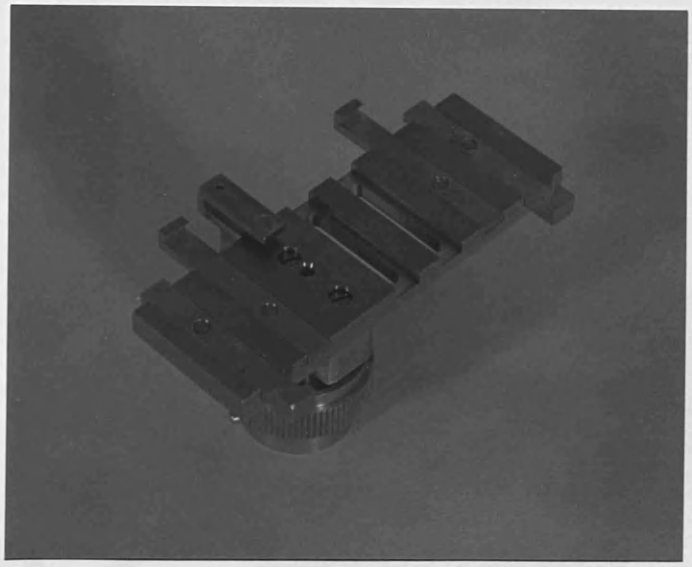
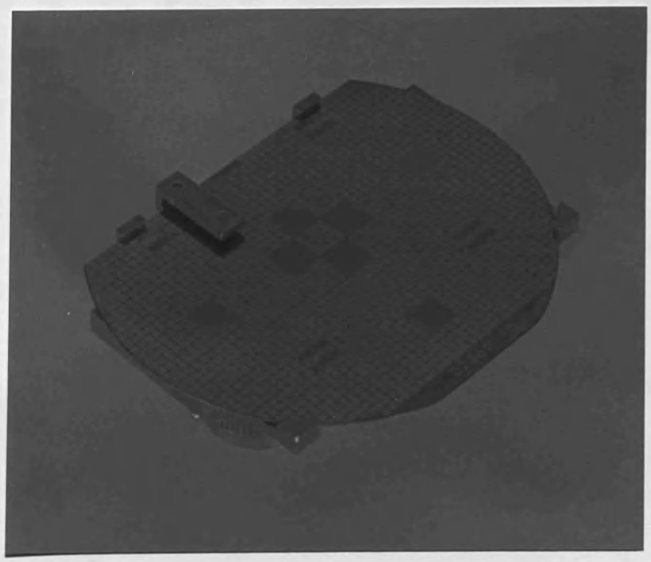


Figure (6-2) Various stages of the double LOCOS process.



(a) unloaded



(b) loaded with sawn B.T.R.L. wafer.

Figure (6-3) Wafer chuck for Glasgow E.B.L. system which was designed to support sawn 73mm (3inch) wafers.

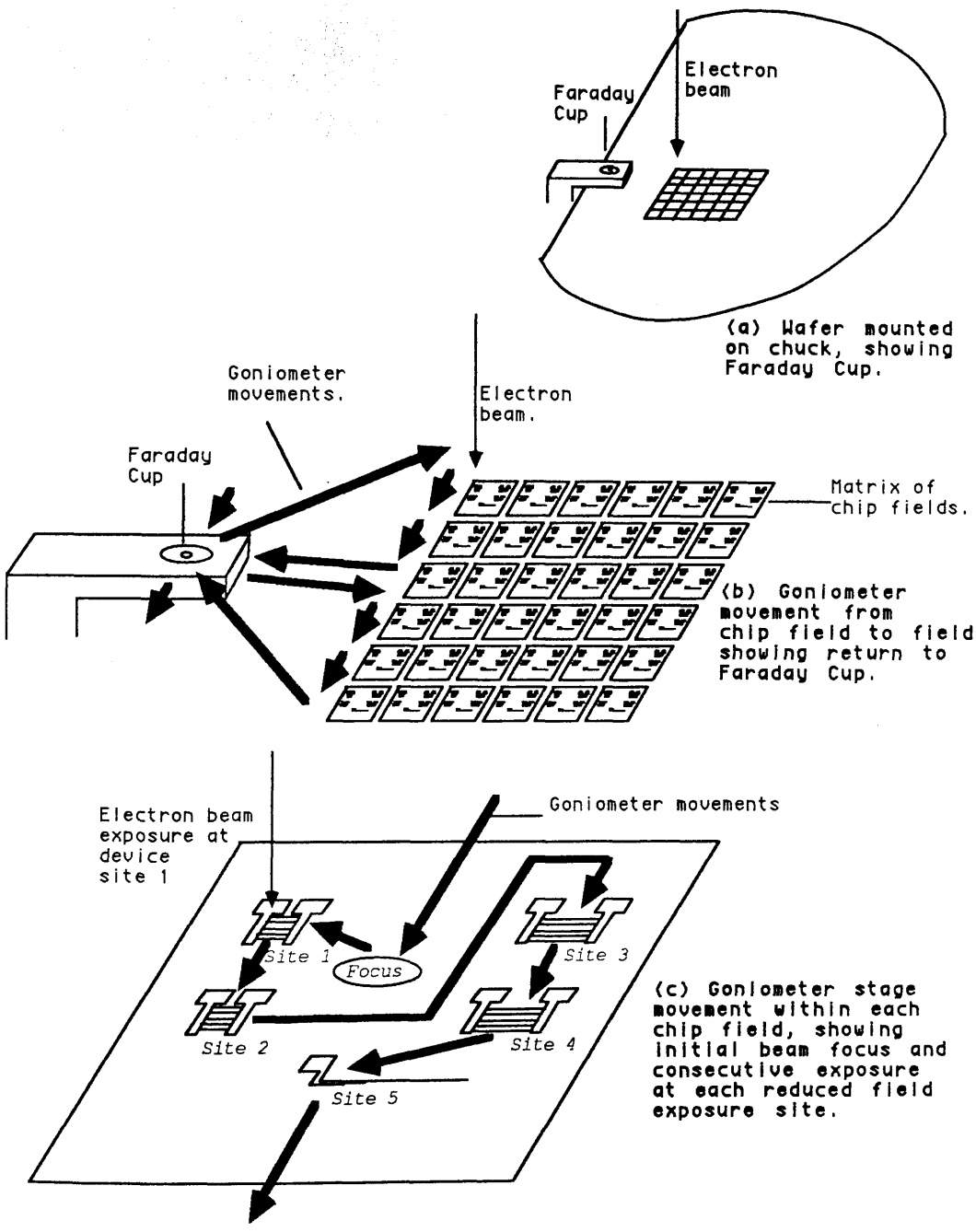
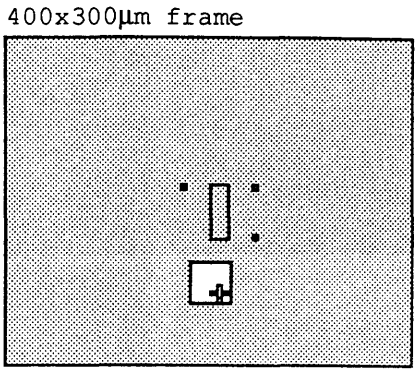
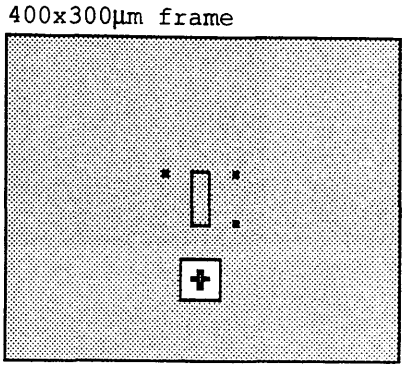


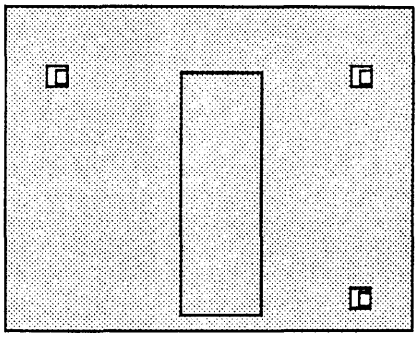
Figure (6-4) Wafer exposure strategy for high resolution electron beam patterning of the gate level.



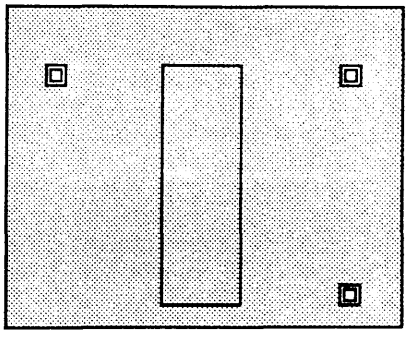
(a) Initial mis-alignment



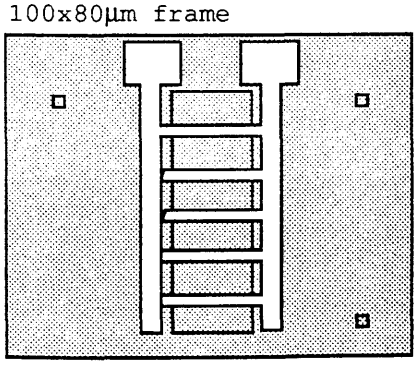
(b) alignment step1 (course)



(c) Slight mis-alignment

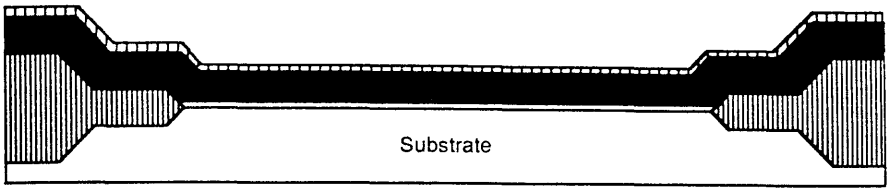


(d) alignment step 2 (fine)

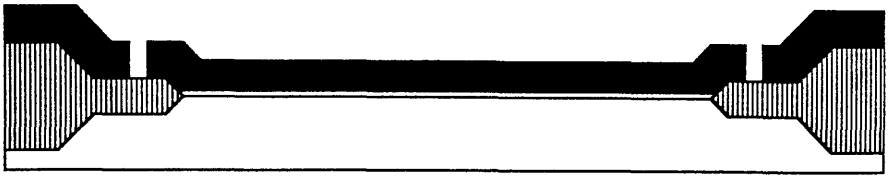


(e) exposure of gate pattern

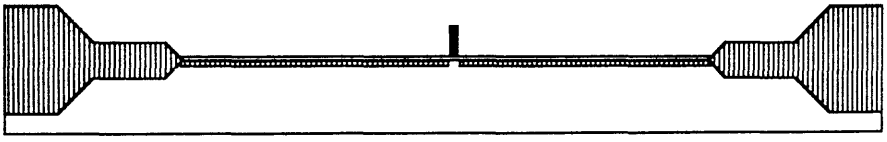
Figure (6-5) Two step technique used for aligning to the reduced field registration marks at each device site.



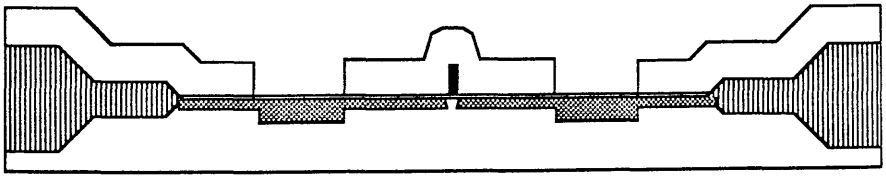
(a) Structure with gate oxide, poly-Si and poly-Si oxide



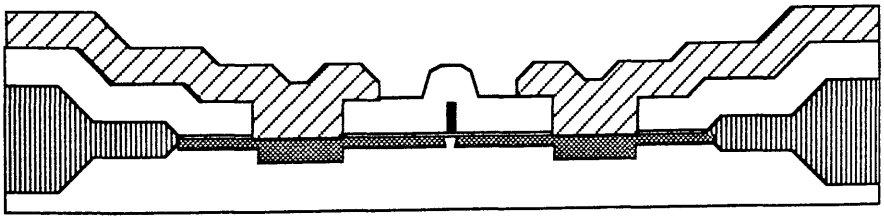
(b) Structure with reduced field registration marks.



(c) Structure after gate etching and S/D implantation.



(d) After etching and implanting the contact windows.



(e) The completed device.

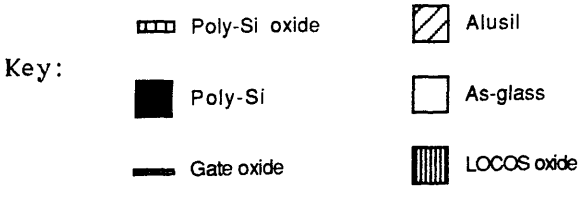
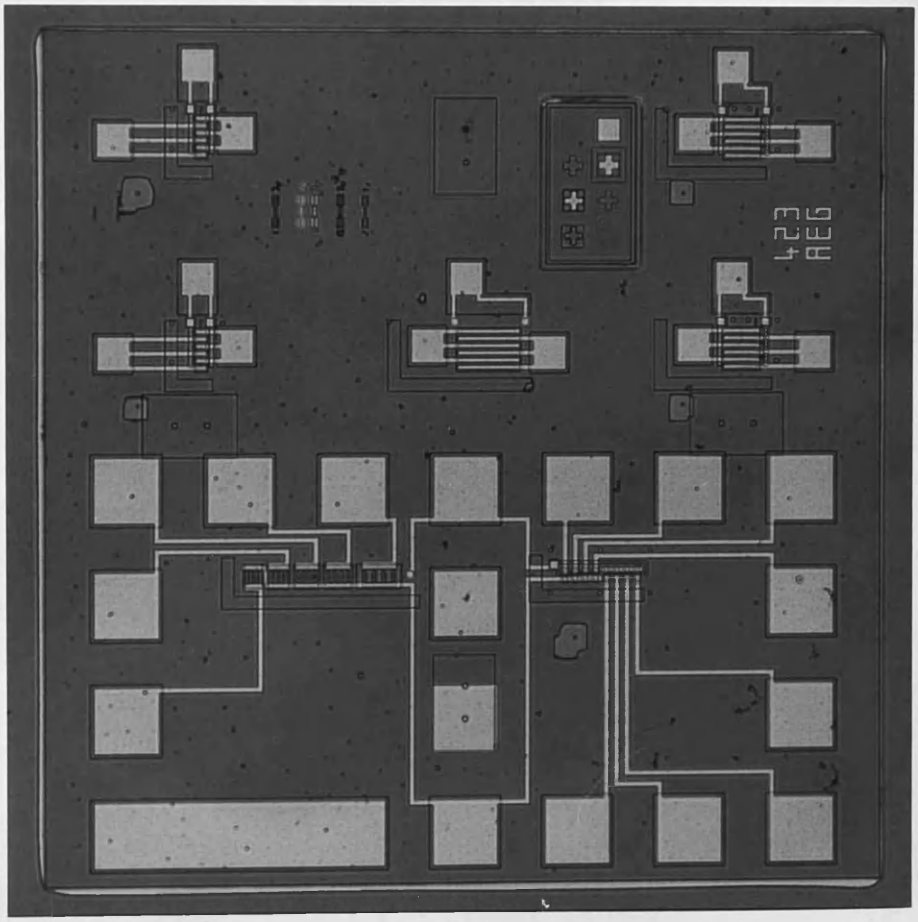
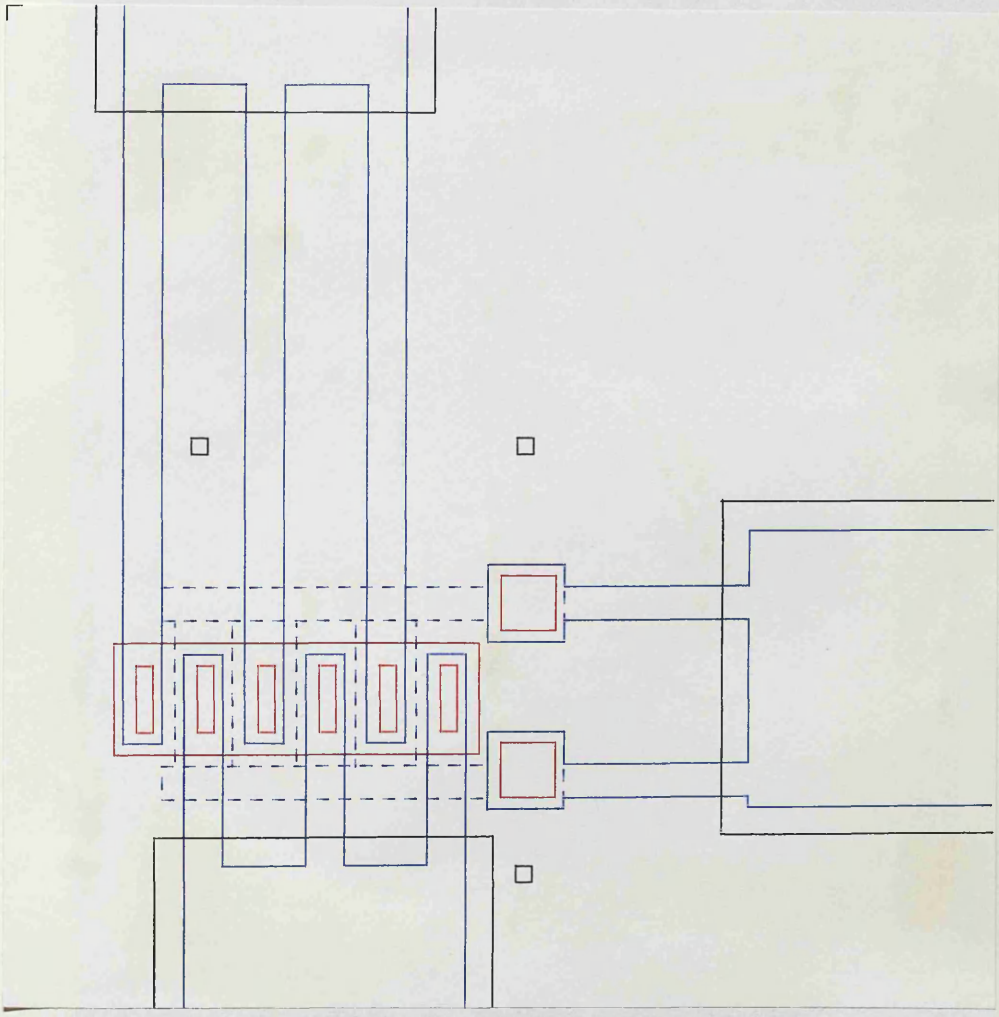


Figure (6-6) Device structure at various stages of fabrication for the B.T.R.L. wafers.



Scale: 250um

Figure (6-7) Fully processed chip field using the 423AEG mask set (designed at B.T.R.L.)



Scale: 50um

Figure (6-8a) Mask layout of an inter-digitated MOSFET structure with a total gate width of 100μm.

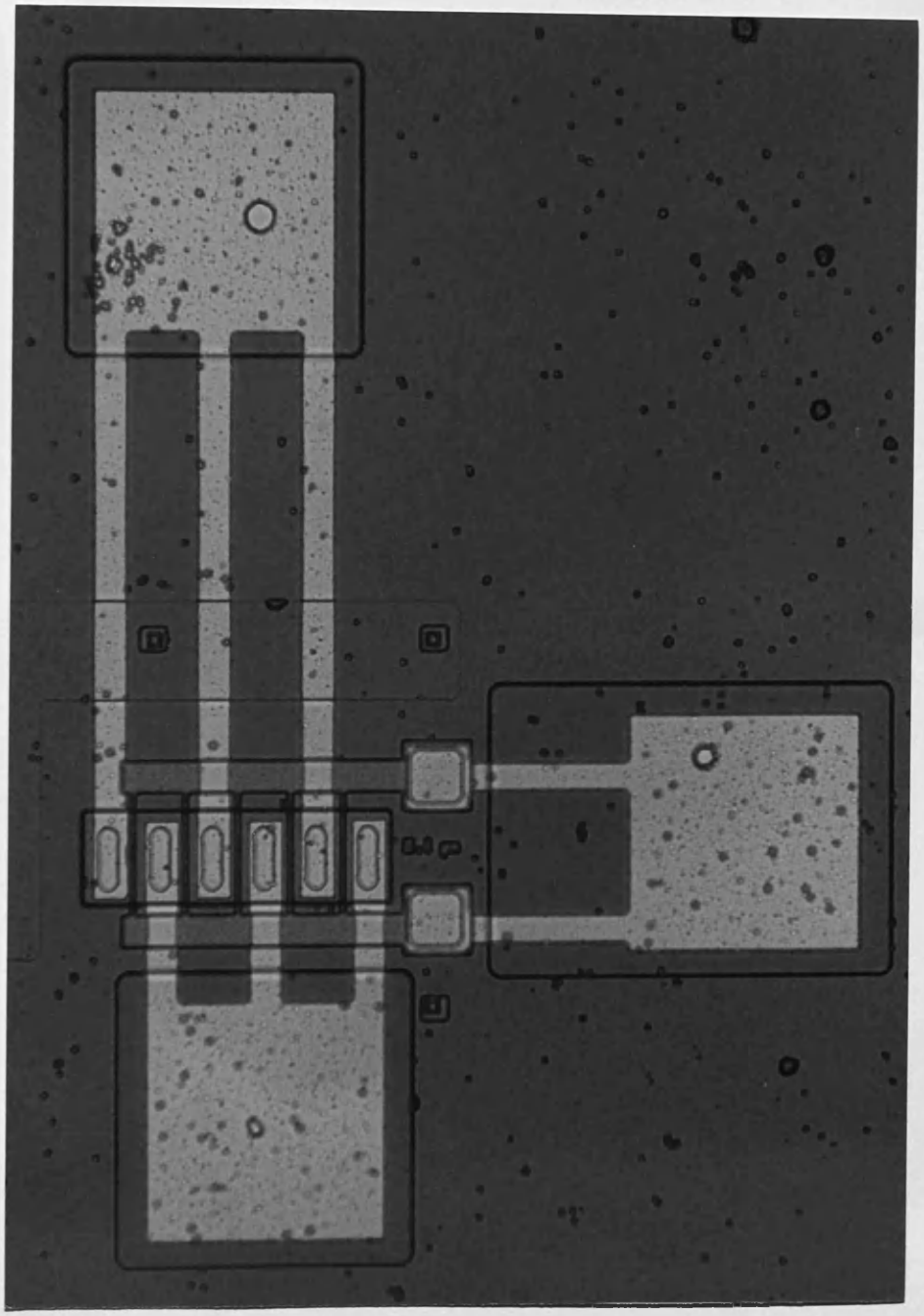


Figure (6-8b) Silicon implementation of an inter-digitated MOSFET structure with a gate length of 0.1μm and a gate width of 100μm.

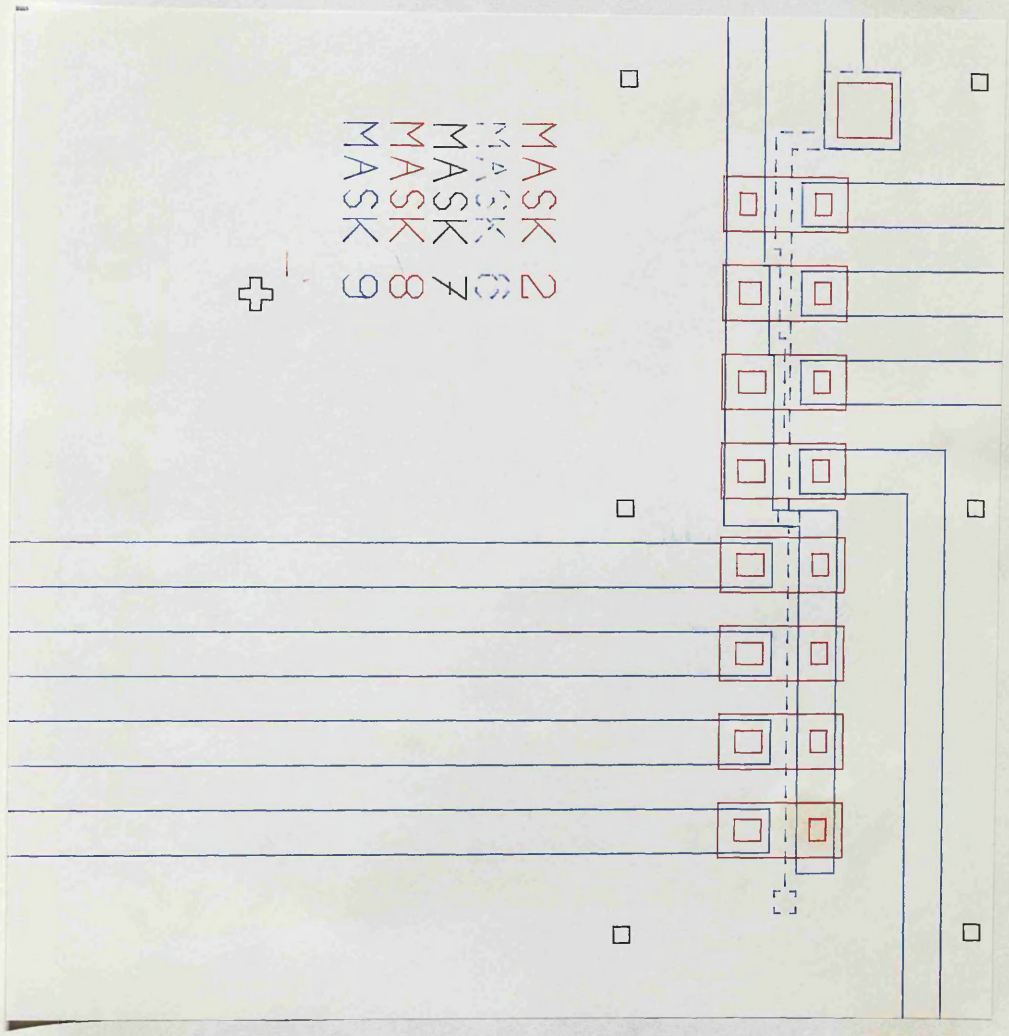
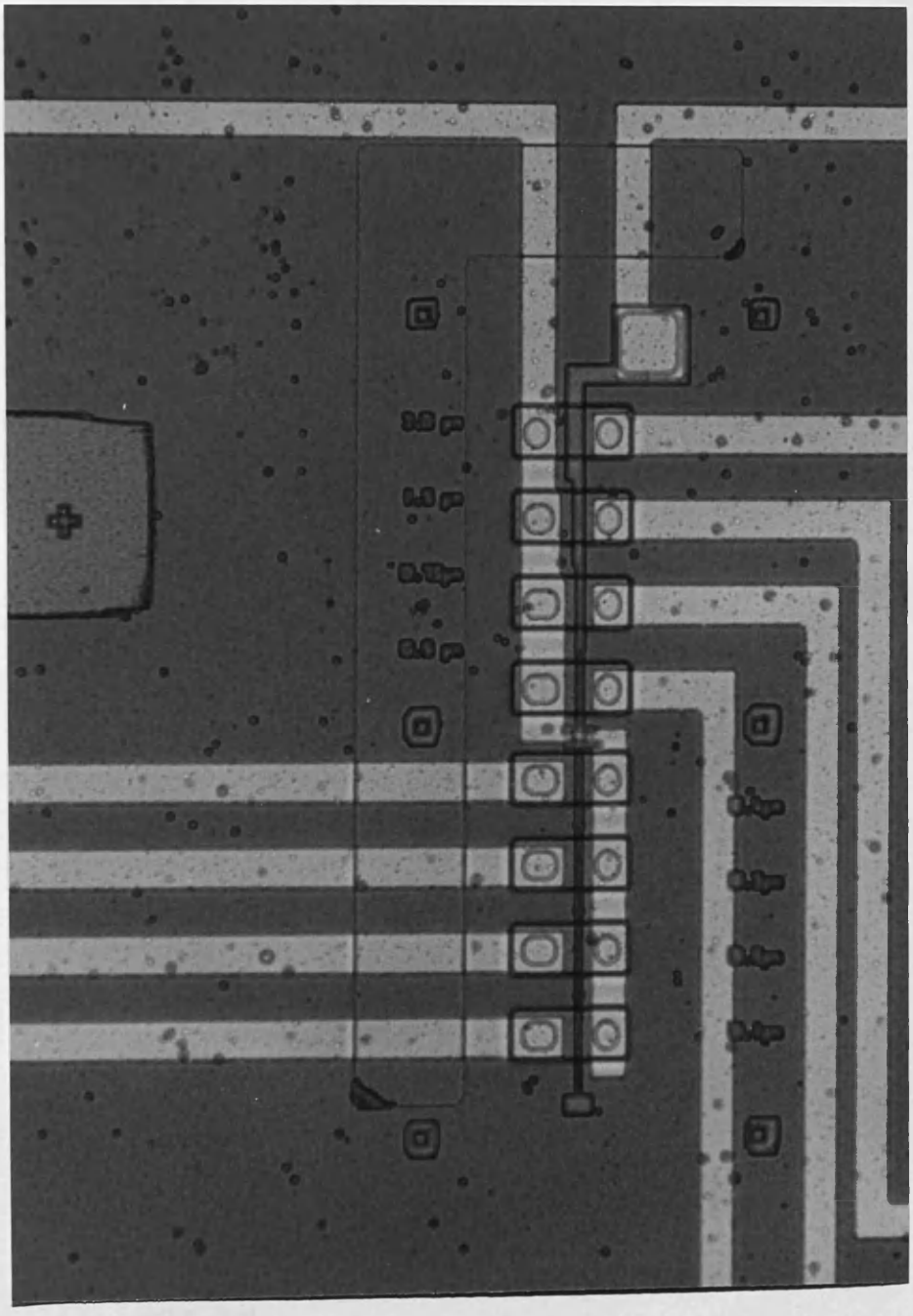


Figure (6-9a) Mask layout of an array of eight devices with decreasing gate lengths.



Scale: 50um

Figure (6-9b) Silicon implementation of a device array with decreasing submicron gate lengths.

Glasgow University-Edinburgh University Collaboration.

Submicron Si MOSFET fabrication sequence for use with EU567 mask set.

Use 3inch wafers : P-type (100) 14-20 ohm-cm

Step	**LOCOS isolation of active areas**
1.1	Initial Clean
1.2	Grow 500A of SiO2 (buffer)
1.3	Deposit 700A of Si3N4
1.4	Photolith - print mask 1 (bond pad oxide)
1.5	Plasma etch Si3N4 (in CF4 & O2)
1.6	MIT strip resist
1.7	Wet etch buffer oxide (10:1 buffered HF)
1.8	Grow 1.4um of SiO2 (bond pad oxide)
1.9	Photolith - print mask 2 (active areas)
1.10	Oxide etch-back (10:1 buffered HF)
1.11	Plasma etch Si3N4 (in CF4 & O2)
1.12	Field implant (Boron)
1.13	MIT strip resist
1.14	Wet etch buffer oxide (10:1 buffered HF)
1.15	Grow 0.85um of SiO2 (field oxide)
1.16	Oxide etch-back (10:1 buffered HF)
1.17	Wet etch Si3N4 from active areas
1.18	Wet etch buffer oxide
	Gate oxide growth and channel implantation
2.1	Photolith - print mask 3 (Dep. devs)
2.2	Depletion implants (As)
2.3	MIT resist strip
2.4	Pre-gate clean
2.5	Grow gate oxide (dry at 785oC)
2.6	Channel implant (Boron)
	Gate electrode formation
3.1	RCA clean
3.2	Deposit 0.5um of Poly-Si
3.3	Grow 0.4 um thick masking oxide on Poly-Si
3.4	Deposit 1um of Al
3.5	Photolith - print mask 5 (Deep implants)
3.6	Plasma etch Al (in CCl4)
3.7	Plasma etch SiO2 on Poly-Si (in CHF3)
3.8	MIT resist strip
3.9	Deep implant (Phos)
3.10	Wet etch Al
3.11	Wet etch SiO2 from Poly-Si
3.12	Neutral ambient anneal
3.13	Poly-Si implant (Phos)
3.14	Neutral ambient anneal
3.15	Dip in 10:1 buffered HF till non-wetting

Figure (6-10) Modified processing sequence used for later devices (E.M.F.) (cont. overleaf).

Select process route A or B:

Route A - Optical gate level exposure

3.16A Photolith - print mask 6 (Gate patterns to 1um)
3.17A Send to Glasgow University
3.18A Plasma etch Poly-Si (in SiCl4 RIE)
3.19A Plasma strip resist (in O2)
3.20A MIT resist strip
3.21A Return to Edinburgh University and continue with step 3.30

Route B - Electron beam gate level exposure

3.16B Spin resist,saw wafers to 59mm and strip resist
3.17B Photolith - print mask 7 (Reduced field registration marks)
3.18B Send to Glasgow University
3.19B Plasma etch Poly-Si (in SiCl4 RIE)
3.20B Plasma strip resist (in O2)
3.21B MIT resist strip
3.22B Dip in 10:1 buffered HF till non-wetting
3.23B Spin and pre-bake HRN negative e.beam resist
3.24B E.beam exposure of gate level patterns to 0.1um
3.25B Develop HRN
3.26B Plasma etch Poly-Si (in SiCl4 RIE)
3.27B Plasma strip resist (in O2)
3.28B MIT resist strip
3.29B Return to Edinburgh University and continue with step 3.30

Continue

3.30 Rinse in de-ionised water
3.31 Poly-Si sidewall oxidation

Drift region formation

4.1 Source/drain implant (As)

Dielectric deposition

5.1 P-glass deposition
5.2 1st stage S/D anneal
5.3 Rapid furnace anneal (wet,30sec at 1000oC) P-glass densification and 2nd stage S/D anneal

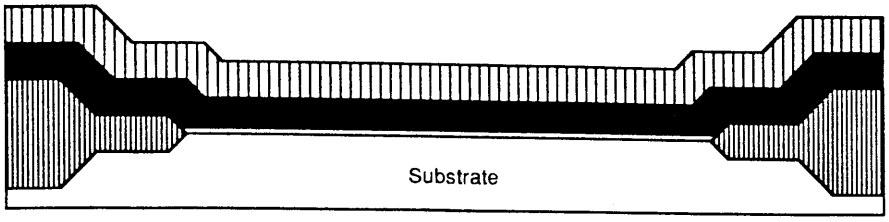
Contact window formation

6.1 Photolith - print mask 8 (contact windows)
6.2 Plasma etch P-glass (in CHF3)
6.3 MIT resist strip

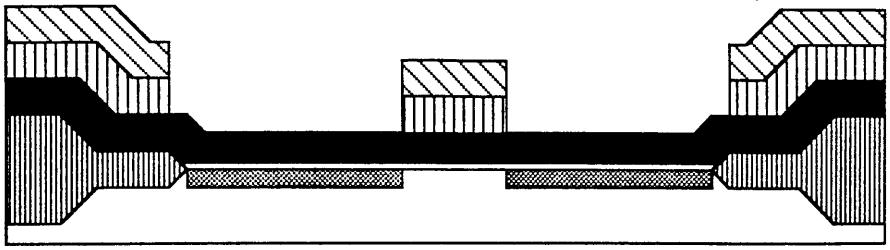
Inter-connect metalisation

7.1 Pre-metal etch
7.2 Deposit 0.5um Alusil (Magnetron sputter)
7.3 Litho - print mask 9 (metalisation)
7.4 Plasma etch metal (in CCl4)
7.5 MIT resist strip
7.6 Sinter in forming gas
7.7 Send to Glasgow for testing

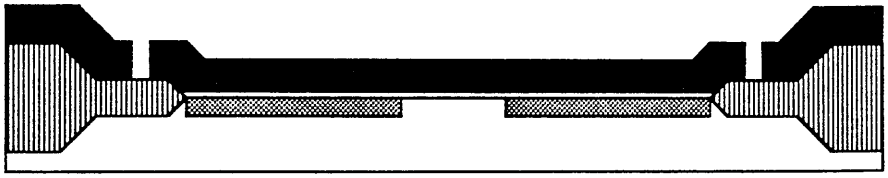
Figure (6-10) cont.



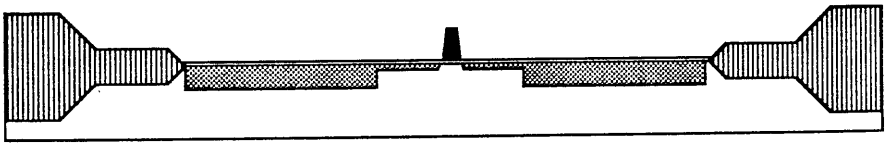
(a) Structure with gate oxide, poly-Si and poly-Si oxide



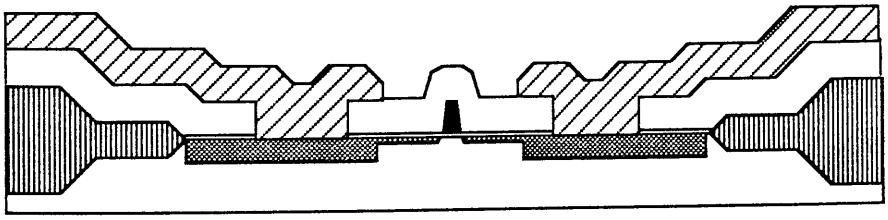
(b) Stucture after implantation through aluminium mask



(c) Structure with reduced field registration marks



(d) Structure after gate etching and S/D implantation.



(e) Completed device structure.

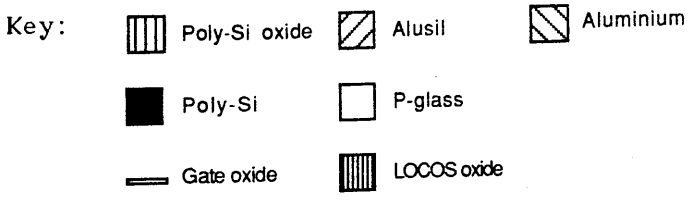
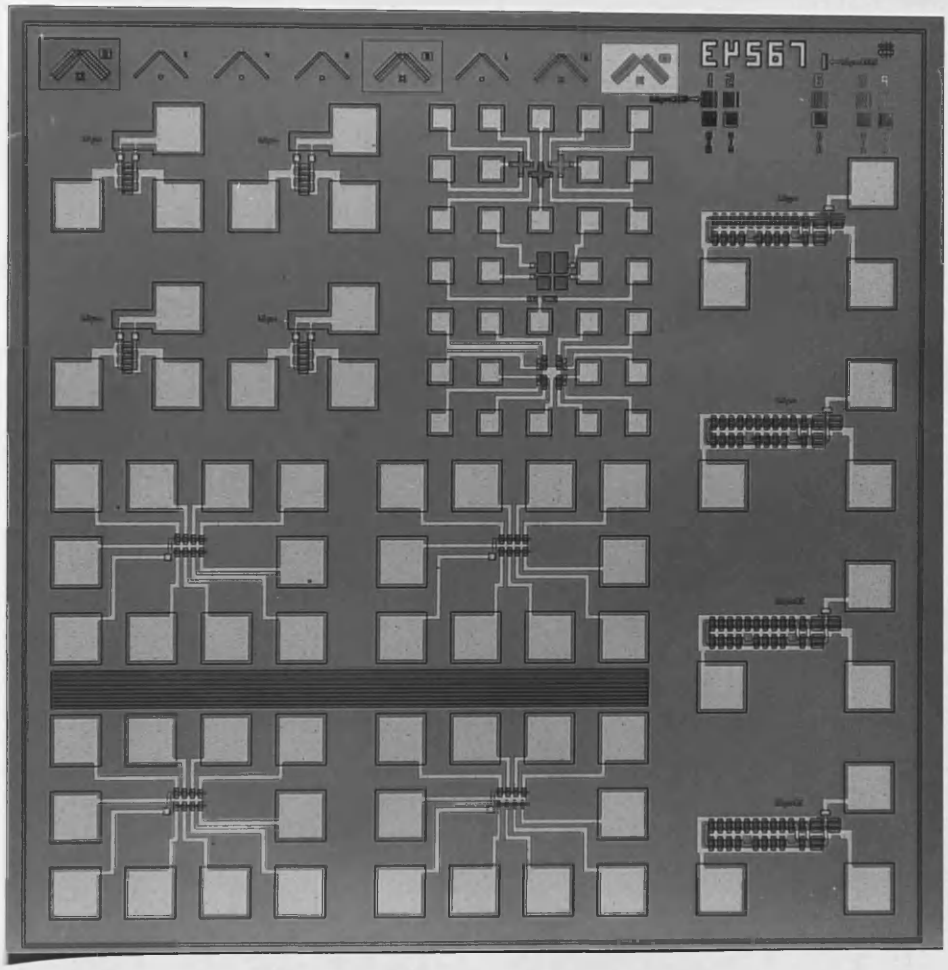


Figure (6-11) Device structure at various stages of fabrication for the E.M.F. wafers.



Scale: 250um

Figure (6-12) Fully processed chip field fabricated using the EU567 mask set (designed at Glasgow).

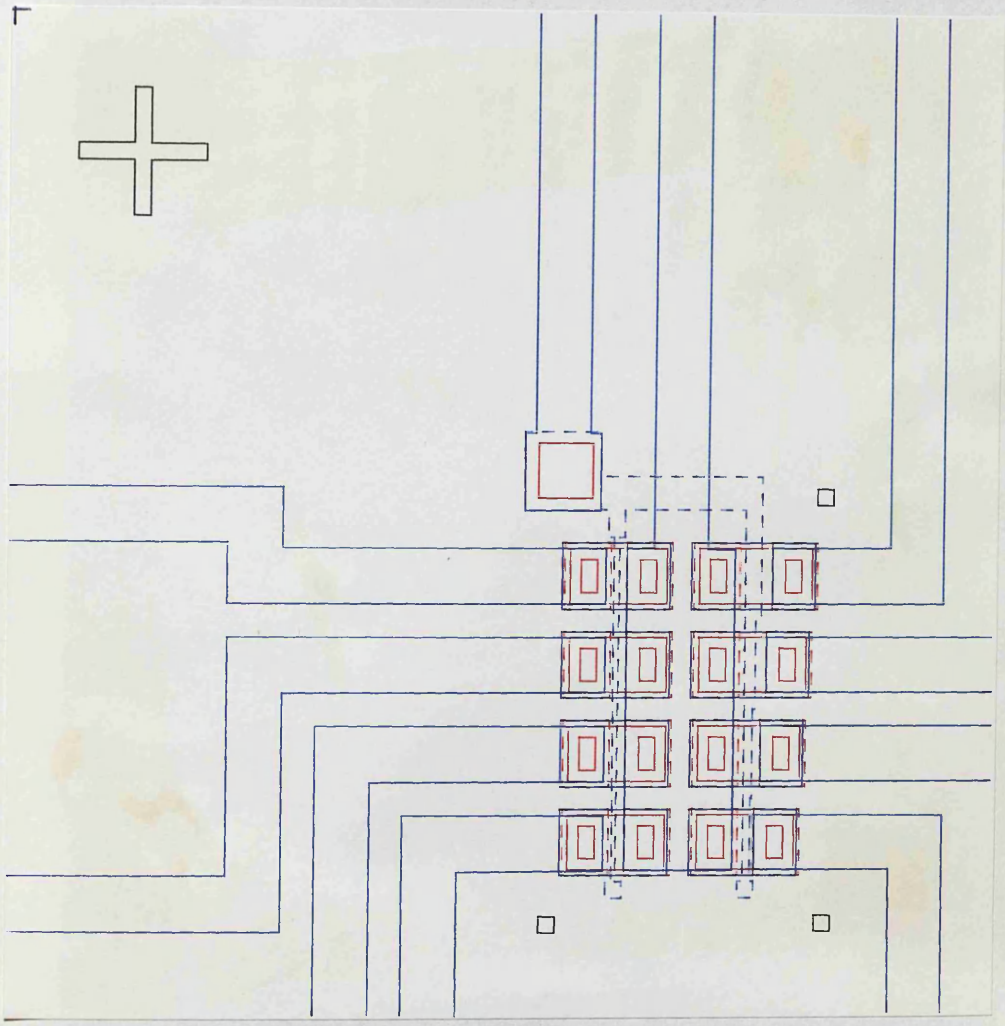
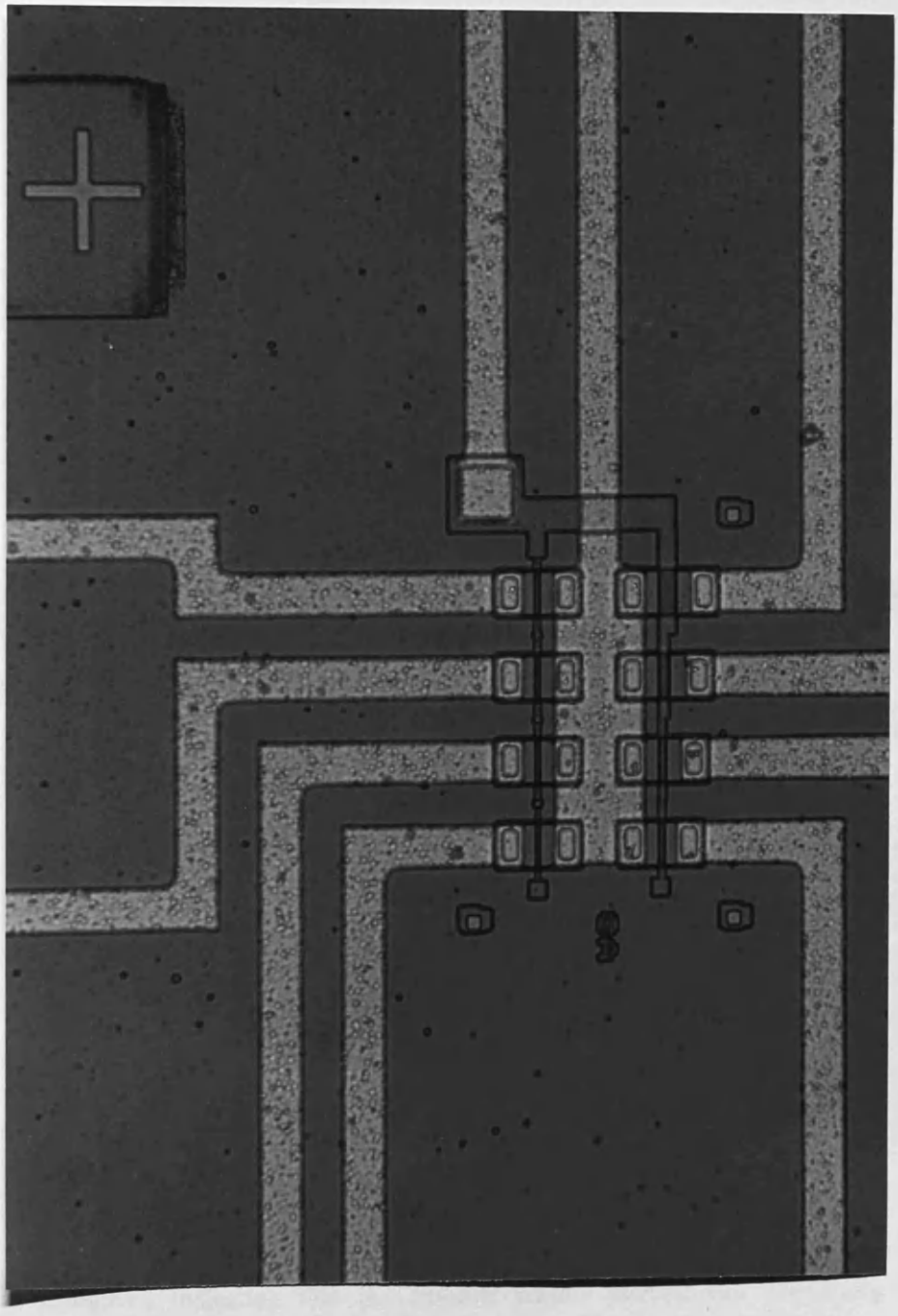


Figure (6-13a) Mask layout of an array of 8 devices with decreasing gate lengths from the EU567 mask set.



Scale: 50um

Figure (6-13b) Silicon implementation of the array of devices shown in figure (6-13a). Gate lengths range from approx. 3 - 0.1 μm.

Chapter 7

Device Results – Scaling

7.1 Introduction

Experimental MOSFETs have been fabricated with gate lengths down to 0.1 microns using the B.T.R.L. process route and the AEG423 mask set. This chapter describes the d.c electrical characterisation of completed devices with 150 angstrom gate oxides.

7.1.1 The Process Matrix

All permutations of the processing matrix, shown earlier in figure (5–1), produced functioning devices. From scaling considerations, the devices with 150 angstrom oxides are of greatest interest and are reported here.

During fabrication, partial wafer masking techniques were used on two wafers to implement the six channel implants. The CH1, CH3 and CH5 conditions were used for one wafer and CH2, CH4 and CH6 were used for the other wafer.

7.1.2 Gate Length Calibration

As described in section 6.3.3.1, gate level calibration lines were exposed onto each wafer using the same parameters as for the gate electrodes. Once completed, the wafers were cleaved and the lines evaluated by SEM. In each case, the polysilicon profiles were highly rectangular, indicating that the chlorine plasma etching had been very successful. Figure (7–1) shows the profile of a 0.11 micron polysilicon line from the CH2/4/6 wafer. The resulting calibration linewidths are shown in table (7–1). Linewidths for the CH1/3/5 wafer were consistently 0.03 microns less than for the CH2/4/6 wafer although the precise cause has not been established.

For each wafer the device gate lengths (L_g) are assumed to be

equivalent to the polysilicon calibration linewidths (see chapter 6). The experimental error associated with the gate length measurements was estimated at less than 10%.

7.2 Electrical Characterisation

Electrical measurements were undertaken at d.c using a Hewlett Packard HP4145A semiconductor parameter analyser which was controlled by an Olivetti computer. For each device under test, electrical connection was achieved manually using an Omniprobe wafer probing station.

Devices were characterised with a range of gate lengths and channel doping levels. The arrays of devices shown earlier in figure (6–9) were used for this purpose. Each device has a nominal gate width (W_g) of 10 microns and a nominal source/drain contact window separation of 8 microns.

Five curves were plotted for each device:

(1) I–V output curves
(2) Subthreshold curves
(3) Turn–on curves
(4) Substrate current curves
(5) Substrate sensitivity curves

Additionally, various discrete measurements were made. These included threshold voltage (V_{TH}), subthreshold swing (S_w), transconductance (g_m) and output conductance (g_d).

With the exception of the substrate sensitivity curves all other measurements were made at a substrate voltage (V_b) of zero volts.

7.2.1 Characterisation of BT–CH3–OX150 Devices

Results are presented for a set of devices with gate lengths (L_g) in the range 0.42 to 0.08 microns. The devices were implanted using the CH3 parameters corresponding to a channel doping level of approximately 1.2×10^{17} atoms/cm^3.

7.2.1.1 I– V Output Curves

Figure (7– 2) shows the I– V output curves for the five devices. The curves are consistent with the gradual onset of short channel behaviour. The curves also indicate carrier velocity saturation, which is seen as a constant transconductance (g_m) for increasing gate voltage (V_g).

The 0.42 micron device in figure (7– 2a) shows good long channel behaviour, indicated by a low output conductance (g_d) in the saturation region. To permit comparisons between devices, g_m and g_d values were measured at a pre– defined operating point. A drain voltage (V_d) of 2 volts and a drain current (I_d) of 0.5 mA were chosen for this purpose. For the 0.42 micron device, a g_m of 80 mS/mm and a g_d of 6 mS/mm were measured, corresponding to a voltage gain (A_v) of 13.

The 0.33 micron device in figure (7– 2b) shows a slightly higher g_m, with a value of 87 mS/mm for the same conditions. The onset of short channel behaviour can be seen via the increased slopes in the saturation region. A g_d of 11 mS/mm was measured giving a voltage gain (A_v) of 8.

An increase in short channel behaviour is seen for the 0.25 micron device in figure (7– 2c). The measured g_m and g_d are 91 and 21 mS/mm resulting in a reduced voltage gain (A_v) of 4.3.

Figure (7– 2d) shows a further decrease in g_d for the 0.17 micron device. In this case, g_m and g_d values of 90 and 29 mS/mm were measured, corresponding to a voltage gain (A_v) of 3. In addition, the severe short channel effect of punch– through is apparent at drain voltages (V_d) above 0.4 volts.

The 0.08 micron device, shown in figure (7– 2e), is punched– through at all drain voltages (V_d). The drain current (I_d) shows a very strong dependence upon drain voltage and cannot be turned off at any gate voltage (V_g).

The relatively small increase in g_m as gate length is decreased from 0.42 to 0.17 microns is consistent with the devices operating under conditions of carrier velocity saturation.

The degradation in output conductance (g_d) can be explained by a drain induced lowering of the threshold voltage. This is predicted

by equation (2−12) where an increased drain voltage (V_d) leads to an increased drain depletion width (y_d) and consequently a reduced short channel threshold voltage ($V_{th(S)}$). At a constant gate voltage, the lower threshold voltage gives rise to a higher drain current (I_d).

7.2.1.2 Subthreshold Curves

Subthreshold curves were obtained for the devices at drain voltages (V_d) of 0.1, 1.0 and 1.9 volts. The resulting curves for the five devices are shown in figure (7−3). These curves can be used to determine the device threshold voltage (V_{TH}) and the subthreshold swing (Sw). As drain voltage increases, a reduction of threshold voltage indicates short channel behaviour whilst an increase in subthreshold swing indicates the presence of a punch−through current component.

For measurement purposes, device threshold voltage (V_{TH}) was defined as the gate voltage required to cause a pre−defined drain current (I_d) to flow. This current, termed the drain threshold current (I_{TH}) was defined by equation (7.1).

$$I_{TH} = 10^{-7} \times \frac{W_g}{L_g} \qquad \text{(in amps)} \qquad (7.1)$$

This definition of threshold voltage was chosen to facilitate measurement, without resort to best−fit and extrapolation algorithms as required by other definitions. A similar definition is used by the MINIMOS device simulation package which is described in section 7.3.

The subthreshold curve for the 0.42 micron device shows long channel behaviour, consistent with the I−V curves described above. There is a slight decrease in threshold voltage (V_{TH}) from 1 to 0.9 volts as the drain voltage (V_d) increases from 0.1 to 1.9 volts. The subthreshold swing (S_w) of 95 mV/decade is independent of drain voltage indicating the absence of a punch−through current component.

For the 0.33 micron device shown in figure (7−3b), the subthreshold swing (S_w) is again constant at a value of 100

mV/decade. However the threshold voltage (V_{TH}) decreases from 0.9 to 0.7 volts as drain voltage increases from 0.1 to 1.9 volts.

The 0.25 micron device shown in figure (7−3c) indicates that the threshold voltage (V_{TH}) decreases from 0.7 to 0.35 volts as drain voltage (V_d) increases from 0.1 to 1.9 volts. Once again the subthreshold swing is relatively constant, increasing from 100 to 115 mV/decade over the same drain voltage range.

The 0.17 micron device shown in figure (7−3d) has a threshold voltage (V_{TH}) of 0.3 volts and a subthreshold swing (S_w) of 160 mV/decade at a drain voltage (V_d) of 0.1 volts indicating the presence of a punch−through current component. Stronger punch−through effects are seen at higher drain voltages.

Severe punch−through is seen in figure (7−3e) for the 0.08 micron device.

7.2.1.3 Turn−On Curves

Plots were obtained for drain current (V_d) as a function of gate voltage (V_g), using a linear current axis. Curves for 0.33 and 0.17 micron devices are shown in figure (7−4). The curves again indicate a reduction of threshold voltage as drain voltage increases. However, for both devices, a non−linear relationship with gate voltage is seen at a drain voltage (V_d) of 0.1 volts. This effect can be attributed to high parasitic source/drain resistance which is described further in section 7.2.5.

7.2.1.4 Substrate Current Curves

Substrate current gives an indication of the extent of impact ionisation taking place within a device. The resulting holes are attracted to the substrate contact and constitute a substrate current (see section 2.5.1.2). A plot of substrate current was obtained for each device as indicated in figure (7−5). The substrate current (I_b) was measured as a function of gate voltage (V_g) for drain voltages (V_d) of 0, 1, 2, 3 and 4 volts at a substrate voltage (V_b) of zero volts.

Due to measurement noise, the substrate current could not be resolved for values below approximately 1nA. The 0.42, 0.33 and 0.25 micron devices show a similar trend between substrate current (I_b) and gate voltage (V_g). The current initially increases rapidly with gate voltage (V_g) and then gradually declines. This can be explained by considering two effects. As the gate voltage increases above threshold, the drain current increases, leading to more carriers being injected into the high field drain region where impact ionisation takes place. However, the increase in gate voltage also leads to an increased surface potential near to the drain. This reduces the electric field strength in the high field region and consequently the rate of impact ionisation decreases[7.1,7.2].

The peak substrate current values are seen to increase rapidly with drain voltage (V_d). For the 0.42 micron device, the peak current increases from about 10 nA to 10 uA as the drain voltage increases from 2 to 4 volts. At this higher voltage, the substrate current becomes significant in comparison to the device drain currents (I_d) (typically in the range 0 to 1mA). These high substrate current effects, together with other hot carrier effects (see chapter 2) have led to the recent interest in Lightly Doped Drain MOSFETs (LDD) [7.3,7.4,7.5,7.6] which attempt to reduce the maximum field strengths in the drain region. Lightly doped drain regions are required for devices with 1 to 0.5 micron channel lengths when operating at a supply voltage (V_{DD}) of 5 volts. However, by reducing the supply voltage to 2 volts or below, hot electron effects are considerably reduced and lightly doped drains are not necessary.

The 0.33, 0.25 and 0.17 micron devices show similar substrate current trends. However peak currents increase for shorter gate lengths. This is best seen for the current peaks at drain voltages (V_d) of 2 volts. Even for the 0.17 micron device the peak substrate current is less than 0.2 uA.

For the 0.08 micron device, drain current is only seen to decrease with gate voltage, showing no current peak. This can be explained by the punched–through operation where drain current is principally controlled by the drain voltage rather than the gate voltage.

7.2.1.5 Substrate Sensitivity Curves

Figure (7−6) shows substrate sensitivity curves for the 0.33 and 0.17 micron devices. All curves were measured at a drain voltage (V_d) of 0.1 volts. Each graph shows plots corresponding to substrate voltages (V_b) of 0, −1, −2, −3, −4 and −5 volts. As expected, the curves indicate an increase in threshold voltage as the substrate voltage is reduced (i.e. becomes more negative). Threshold voltage dependence upon substrate voltage is predicted by equation (2.1). However, it is of interest to note that this dependence is weaker for the 0.17 micron device. This can be attributed to short channel behaviour. The source and drain depletion regions extend significantly under the gate and therefore partly screen the substrate voltage effects.

7.2.2 Characterisation of BT−CH4−OX150 Devices

Device characteristic curves are shown in figures (7−7) and (7−8) for a range of devices with CH4 channel implants. These devices have a channel doping level of approximately $3x10^{17}$ atoms/cm^3, 2.5 times higher than for the CH3 devices. Figure (7−7) shows the I−V output curves and figure (7−8) shows subthreshold curves. As expected, the curves show a reduction in short channel effects in comparison to the CH3 devices. This is seen most clearly for the 0.20 micron device which does not punch through until the drain voltage reaches approximately 2 volts. However the 0.17 micron CH3 device was seen to punch through at drain voltages of about 0.4 volts (see figure (7−3d)).

7.2.3 Characterisation of BT−CH6−OX150 Devices

The CH6 implanted devices have the highest channel doping levels and therefore are most likely to retain long channel behaviour for short physical gate lengths. Figures (7−9) and (7−10) show the resulting I−V output and subthreshold curves for devices with gate lengths of 0.20 and 0.11 microns. The subthreshold characteristic

curves for the longer device indicate only slight short channel behaviour.

The 0.11 micron device has a tranconductance (g_m) of 70mS/mm but shows a degraded operation in the saturation region. The high output conductance (g_d) of 40mS/mm can be attributed to a punch– through current component. It is interesting to note that for a gate voltage (V_g) of zero volts the device does not punch through until the drain voltage (V_d) exceeds 1.5 volts. However at higher gate voltages punch– through occurs at lower drain voltages. This gate voltage modulation of the punch– through current is not fully understood and requires further investigation.

7.2.4 General Matrix Trends

The results presented above have demonstrated the gradual progression from long channel to short channel behaviour as gate length is reduced. This transition is indicated by a threshold voltage (V_{TH}) dependence upon both device gate length (L_g) and drain voltage (V_d). Consequently threshold voltage was chosen as a parameter for identifying the general long channel/short channel trends, for each of the six channel doping levels. The results are shown in figure (7–11). It can be seen that the CH1 and CH2 channel implants are too small to prevent short channel operation for gate lengths of less than 0.5 microns. The CH3 and CH4 channel implants, which correspond to channel doping levels of $1.2x10^{17}$ and $3x10^{17}$atoms/cm^3, are seen to indicate long channel operation for gate lengths down to 0.42 microns and 0.36 microns respectively.

The CH5 channel implant parameters correspond to a channel doping level of approximately $6x10^{17}$ atoms/cm^3 and long channel operation is indicated at gate lengths down to 0.25 microns.

The graph for the CH6 channel implant parameters is more difficult to interpret with the transition from long to short channel behaviour occuring in the region 0.28 to 0.2 microns. The slight increase in threshold voltage for the 0.2 micron device is not understood but may be due to an increased fixed oxide charge at the gate edges due to dry etching damage.

Variations in long channel threshold voltage were observed for

nominally identical devices, located within different chip fields. The variations across the wafers were in some cases as high as 0.3 volts and may have been due to variations in oxide thickness, surface state densities or fixed oxide charge.

Further electrical measurements are shown in table (7−2) for CH3, CH4, CH5 and CH6 devices. The table shows transconductance (g_m), output conductance (g_d) and voltage gain (A_v), measured at a drain voltage (V_d) of 2 volts and a drain current (I_d) of 0.5mA. It can be seen that all devices exhibit a transconductance in the range from 60 to 100mS/mm. Tranconductance remains relatively constant as gate length is reduced but decreases slightly at higher channel doping levels. This latter effect is due to a reduction in saturation velocity resulting from the increased transverse electric fields.

Figure (7−12) shows a plot of voltage gain (A_v) for the CH3, CH4 and CH5 devices from table (7−2). The decrease in gain at shorter channel lengths can be attributed to an increased output conductance (g_d) resulting from the onset of short channel effects. Voltage gain is seen to increase with channel doping level for these devices. The 0.42, 0.36 and 0.25 micron devices, identified above as long channel devices for CH3, CH4 and CH5 channel implants, are seen to have voltage gains (A_v) of 16, 18 and 13 respectively.

7.2.5 Extraction of Device Constants

The low field mobility (μ_o) and the lateral channel length reduction (ΔL) are extremely useful parameters for evaluating devices. Standard electrical measurement methods[7.7] are unsuitable for these devices due to the high parasitic source/drain resistance. Suciu et al.[7.8] and De La Moneda et al.[7.9] have recently reported methods which can be used for devices with high resistance in order to extract low field mobility (μ_o), mobility degredation coefficient (U_o), channel length reduction (ΔL) and series resistance (R_T). Both methods use the same mathematical analysis, with different graphical techniques. A method similar to that reported by De La Moneda has been used to characterise these devices.

A MOSFET can be modelled in the linear region by equation (7.1)[7.10] where the source and drain resistance effects are taken into

consideration using the dashed voltage terms (see figure (7–13)).

$$I_d = \beta \cdot (V_g' - V_{TH} - V_d'/2) \cdot V_d' \qquad (7.1)$$

when:

$\beta = \beta_o/(1+U_o(V_g'-V_{TH}))$

$\beta_o = \mu_o C_{ox} W/L$

μ_o = low field mobility

U_o = mobility degradation coefficient

$V_g' = V_g - I_d R_S$

$V_d' = V_d - I_d R_T$

R_S = source resistance

R_D = drain resistance

$R_T = R_S + R_D$

for:

$L = L_g - \Delta L$

and

ΔL = lateral channel reduction.

De la Moneda et al. derived equation (7.2) from equation (7.1) to obtain an expression for R_{ON}, the effective series resistance of a turned–on device.

$$R_{ON} = V_d/I_d = R_T + \frac{U_o.L}{\mu_o C_{ox} W} + \frac{L}{\mu_o C_{ox} W} \cdot \frac{1}{(V_g - V_{TH})} \qquad (7.2)$$

Equation (7.2) can be expressed in the form of equation (7.3) which indicates a linear relationship between R_{ON} and $1/(V_g - V_{TH})$. This relationship is confirmed in figure (7–14) which shows experimental curves for CH3 devices with gate lengths (L_g) of 0.17 and 0.33 microns. A constant drain voltage (V_d) of 0.1 volts was used.

$$R_{ON} = R_o + S \cdot \frac{1}{(V_g - V_{TH})} \qquad (7.3)$$

where:

$$Ro = R_T + \frac{U_o . L}{\mu_o C_{ox} W} \qquad (7.4)$$

and

$$S = \frac{L}{\mu_o C_{ox} W} \qquad (7.5)$$

The device constants were determined by measuring **S** and R_o (i.e. the slopes and y−axis intercepts) for devices with a range of gate lengths (L_g). The measured **S** values were plotted directly as a function of gate length (L_g). Figures (7−15a) and (7−15b) show the resulting plots for the CH1/3/5 and CH2/4/6 wafers respectively. In each case, the x−axis intercept is equal to the lateral channel length reduction ($\triangle L$). The low field mobility (μ_o) can be calculated from the gradients which are equal to $1/\mu_o C_{ox} W$.

The plots indicate three important features. Firstly, the straight lines indicate that the calibration measurements and the slope values are consistent. Secondly, all plots indicate a lateral channel reduction ($\triangle L$) of approximately 0.06 microns ($\pm$ 0.01 microns). It is therefore confirmed that minimal dopant re−distribution has been achieved using the Heatpulse rapid thermal anneal step. Finally, the gradients are larger for increased channel doping levels indicating a reduction in low field mobility.

Low field mobility (μ_o) was calculated for each channel doping level and the results are shown in figure (7−16). The decrease in mobility at higher channel doping levels is attributed to impurity scattering effects and the increased transverse fields required to establish carrier inversion[7.11]. The mobility is seen to decrease from approximately 400 to 200 $cm^2/V-s$ as the channel doping levels increase from $3x10^{16}$ to $1.2x10^{18}$ atoms/cm^3. This reduction will

degrade the performance of the scaled devices when operating in the linear region.

R_O can be plotted as a function of S for each set of devices with decreasing gate length and a straight line is predicted from equations (7.3) and (7.4). The y−axis intercept is equal to the total source/drain parasitic resistance (R_T) and the gradient is equal to the mobility degradation coefficient (U_O). This analysis indicated a source/drain parasitic resistance of 300−350 ohms for all the devices tested. However, a relatively small scatter in the R_O measurements prevented an accurate determination of the mobility degradation coefficient (U_O).

The source/drain parasitic resistance can be attributed to three components; firstly, the resistance associated with the source and drain contacts; secondly the resistance associated with the source/drain drift regions and finally the speading resistance[7.12,7.13] associated with current crowding at each end of the channel. For these devices, the drift region resistance will be high due to the shallow junction depth (approx. 0.12μm) and the large source/drain contact spacing (8μm). The high series resistance of these devices results in an extrinsic transconductance ($g_{m(ext)}$) which is significantly lower than the internal transconductance ($g_{m(int)}$) (see equation (5.1)). For example, a device with a width of 10 microns, a source resistance (R_S) of 150 ohms and an intrinsic transconductance ($g_{m(int)}$) of 100mS/mm, will have a reduced extrinsic transconductance ($g_{m(ext)}$) of 87mS/mm.

A device wafer with an arsenic source/drain implant of $2.5x10^{15}$ atoms/cm^2 at a reduced energy of 40 keV rather than the 70 keV used above, gave rise to an extremely high parasitic resistance of 800 ohms. A junction depth of only 0.073 microns is predicted by SUPREM (see section 5.4.3), but the resulting parasitic resistance is unacceptably high.

7.2.6 Determination of Channel Doping Profile

The channel doping profiles for the six different channel implant parameters were calculated using the Beuhler method[7.14]. Long channel devices with gate lengths of approximately 3 microns were used for these measurements.

For this analysis the devices were operated in the linear region and the gate voltage was incremented. At each gate voltage, the substrate voltage was adjusted to maintain a constant drain current (I_d) and hence constant carrier density in the channel region. The increased substrate voltage (V_b) gives rise to an increased depletion depth. Both the junction depth (X) and the channel doping level ($N_A(X)$) are calculated from gate voltage (V_g) and substrate voltage (V_b) measurements using equations (7.6) and (7.7).

$$X = \frac{\epsilon_s}{C_{ox}} \cdot \frac{dV_b}{dV_g} \qquad (7.6)$$

$$N_A(X) = C_{ox}^2 \cdot \frac{d^2V_b}{dV_g^2}^{-1} \qquad (7.7)$$

A drain voltage (V_d) of 0.1 volts and a constant drain current (I_d) of 1 uA were chosen for these measurements. A Hewlett Packard parameter analyser, controlled by an Olivetti computer, was used to increment the gate voltage (V_g) and to establish the required substrate voltage (V_b) using a four step iteration algorithm. Each consecutive iteration determined the required substrate voltage to an additional significant digit. After four iterations, the substrate voltage was established to an accuracy of 0.001 volts and the drain current (I_d) was maintained to an accuracy ± 0.5 percent. After completion of the measurements, the Olivetti computer performed all necessary calculations.

Figure (7−17) shows the experimental results, indicated by crosses, for the six channel implant parameters. The SUPREM II simulation results presented in figure (5−4) are also included. It should be noted that the restricted span of the experimental data is due to the limited substrate voltage range. At voltages above $+0.5$ volts the source and drain junctions become turned on, whilst for voltages of less than -15 volts the junctions break down. In both cases, the drain current increases abnormally.

The experimental data is in good agreement with the SUPREM

simulation results. In most cases, the experimental and SUPREM data agree to within 10%. However, the agreement is least good for the high doping levels of the CH6 implants and the low doping levels of the CH1 implants. The cause of these two anomalies is not known, but may be due to inaccuracies in the electrical measurements or the SUPREM II simulations.

7.3 MINIMOS Modelling

The MINIMOS2.2, MOSFET simulation package[7.15,7.16] was used to model devices with CH3 channel implants. This software package uses finite difference algorithms to solve Poisson's equation and the continuity equations for MOSFETs with channel lengths (L) down to 0.2 microns.

The MINIMOS THRES model was used to predict threshold voltages (V_{TH}) for devices with channel lengths from 0.5 to 0.2 microns at drain voltages (V_d) of 0.1 and 1.0 volts. Figure (7−18) shows the MINIMOS results together with the experimental data from figure (7−11c). Both sets of data are shown as a function of channel length (L). The MINIMOS results assume a work function difference (Φ_{ms}) value of −0.83 volts. The figure shows that the experimental devices generally have a higher threshold voltage than predicted by MINIMOS, possibly due to surface states and fixed oxide charge. The figure also shows that the experimental devices show a more rapid onset of short channel behaviour with decreasing channel length than predicted by MINIMOS. This may be due to a difference in the shape of the source/drain junctions for the experimental and modelled devices and requires further investigation.

7.4 Implications for Device Scaling

It has been shown that short channel effects can be suppressed by increasing the channel doping levels. However, it has not been possible to suppress these effects for gate lengths (L_g) of 0.11 microns even at channel doping levels as high as 1.2×10^{18} atoms/cm^3 (corresponding to CH6 implants).

Long channel threshold behaviour has been observed for devices with gate lengths down to 0.25 microns and channel doping levels of approximately $6x10^{17}$ atoms/cm^3 (see section 7.2.4). These devices have a corresponding channel length (L) of approximately 0.19 microns and appear promising for future VLSI applications. Figure (7–19) shows the I–V characteristic of such a device with a gate width of 100 microns. The large gate width was obtained using the inter–digitated gate structure shown earlier in figure (6–8). A drain voltage range from zero to 4 volts is shown and the onset of avalanche breakdown can be seen for drain voltages above 3.5 volts. The device shows a maximum transconductance of 8 mS (80mS/mm). For useful application, these devices will require the oxide thickness to be reduced from 150 angstroms to approximately 50 angstroms in order to reduce the threshold voltage. It is interesting to note that the channel length (L) and channel doping level (N_A) correspond approximately to the CP3 scaling example of chapter 2.

Whilst it might be possible to suppress short channel effects at shorter dimensions by reducing source/drain junction depths, it appears unlikely that long channel behaviour will be obtained for devices with gate lengths of 0.1 microns. However, refinement of the basic MOSFET structure may help to reduce short channel effects. Figure (7–20) shows a structural enhancement whereby the channel doping level is increased near to the drain junction. This enhancement should retain an acceptably low threshold voltage and high carrier mobility whilst confining the drain depletion region and hence suppressing short channel effects. Low voltage operation would be essential for these devices in order to avoid hot carrier effects.

7.5 Summary

MOSFETs have been fabricated with physical gate lengths down to 0.11 microns. Results have been presented for a wide range of devices with sub 0.5 micron gate lengths. Increased channel doping levels have been shown to reduce short channel effects at these dimensions. The performance of the devices has been limited by the effects of high parasitic source/drain resistance and the effects of carrier velocity saturation. Electrical methods have been used to

determine the lateral channel reduction, the low field mobility and the source/drain parasitic resistance. SUPREM predictions for the channel doping profiles have been verified by electrical measurements.

Long channel behaviour has been demonstrated for devices with gate lengths of 0.25 microns. The channel doping level for these devices corresponds to the CP3 scaling example of chapter 2. The need for a refined device structure has been identified if long channel behaviour is to be attained for smaller physical gate lengths.

References

7.1 C.M. Wu and K.W. Yeh, 'Safety drain voltage avoiding avalanche breakdown in MOSFET,' IEEE Electron Dev. Letts., vol EDL–3, p245, 1982.

7.2 Y. Nissancohen, G.A. Franz and R.F. Kwasnick, 'Measurement and analysis of hot–carrier–stress effect in NMOSFETs using substrate current characterisation,' IEEE Electron Device Lett., vol EDL–7, p451, 1986.

7.3 S. Ogura, P.J. Tsang, W.W. Walker, D. Critchlow and F.J. Shepard, 'Design and characteristics of the lightly doped drain–source (LDD) insulated gate field effect transistor,' IEEE Trans. Electron Devices, vol ED–27, p1359, 1980.

7.4 P.J. Tsang, S. Ogura et al., 'Fabrication of high–performance LDDFETs with oxide sidewall–spacer technology,' IEEE Trans. Electron Devices, vol ED–29, p590, 1982.

7.5 H. Katto, K. Okyama, S. Meguro, R. Nagai and S. Ikeda, 'Hot carrier degradation modes and optimisation of LDD MOSFETs,' IEDM Tech. Dig., p774, San Francisco, 1984.

7.6 J. Hui, F.–C. Hsu and J. Moll, 'A new substrate and gate current phenomenon in short–channel LDD and minimum overlap devices,' IEEE Electron Device Lett., vol EDL–6, p135, 1985.

7.7 A. S. Grove, 'Physics and technology of semiconductor devices,' Wiley, New York, 1967.

7.8 P. I. Suciu and R. L. Johnson, 'Experimental derivation of the source and drain resistance of MOS transistors,' IEEE Trans. Electron Devices, vol ED-27, p1846, 1980.

7.9 F. H. De La Moneda, H. N. Kotecha and M. Shatzkes, 'Measurement of MOSFET constants,' IEEE Electron Device Lett., vol EDL-3, p10, 1982.

7.10 G. Merckel, J. Borell and N.Z. Cupcea, 'An accurate large-signal MOS transistor model for use in computer-aided design,' IEEE Trans. Electron Devices, vol ED-19, p681, 1972.

7.11 R.W. Coen and R.S. Muller, 'Velocity of surface carriers in inversion layers on silicon,' solid State Electron., vol23, p35, 1980.

7.12 G. Baccarani, G.A. Sai-Halasz, 'Spreading resistance in submicron MOSFET's,' IEEE Electron Device Lett., vol EDL-4, p27, 1983.

7.13 K.K. Ng, R.J. Bayruns and S.C. Fang, 'The spreading resistance of MOSFETs,' IEEE Electron Device Lett., vol EDL-6, p195, 1985.

7.14 M.G. Buehler, 'Dopant profiles determined from enhancement mode MOSFET dc measurements,' Appl. Phys. Lett., vol 31, p848, 1977.

7.15 S. Selberherr, A Schutz and H,W, Potzl, 'A two-dimensional MOS transistor analyser,' IEEE Trans. Electron Devices, vol ED-27, p1540, 1980.

7.16 A. Schutz, S. Selberherr and H.W. Potzl, 'A two-dimensional model of the avalanche effect in MOS transistors,' Solid State Electron., Vol 25, p177, 1982.

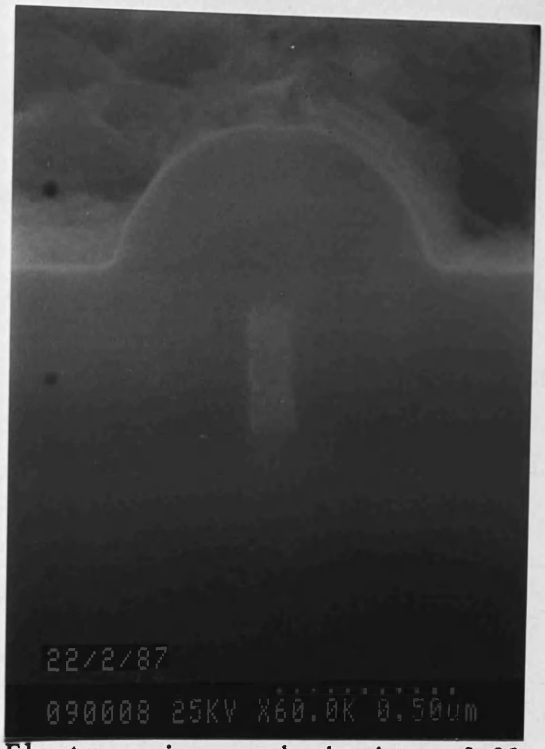
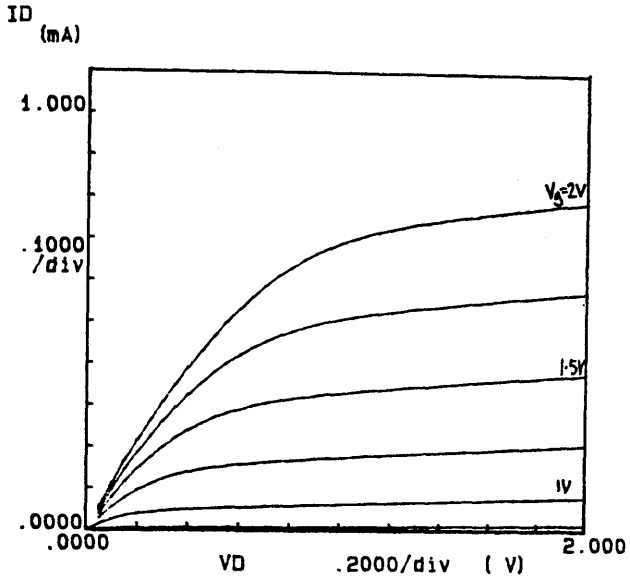


Figure (7-1) Electron micrograph showing a 0.11 micron polysilicon calibration line.

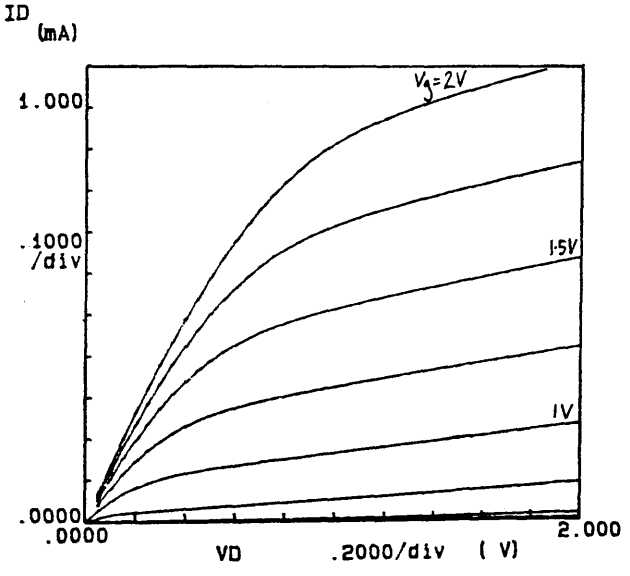
EBL Exposure pixel number at 100x80um frame size	Calibration linewidths for CH1/3/5 wafer (um)	Calibration linewidths for CH2/4/6 wafer (um)
5	0.08	0.11
10	0.17	0.20
15	0.25	0.28
20	0.33	0.36
25	0.42	0.45

Table (7-1) Showing polysilicon linewidths from the calibration patterns of BT wafers with 150 angstrom gate oxides.



(a) L_g=0.42um

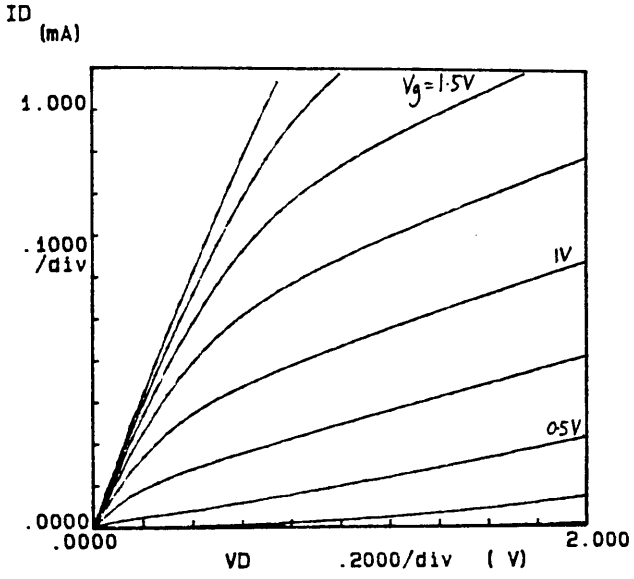
BT-CH3-OX150



(b) L_g=0.33um

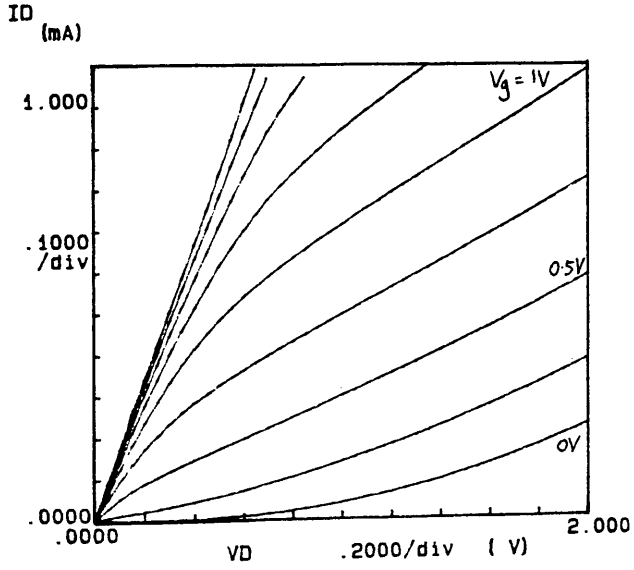
BT-CH3-OX150

Figure (7-2) I-V output curves for a set of five devices fabricated using the BT-CH3-OX150 process conditions. (Cont. overleaf)



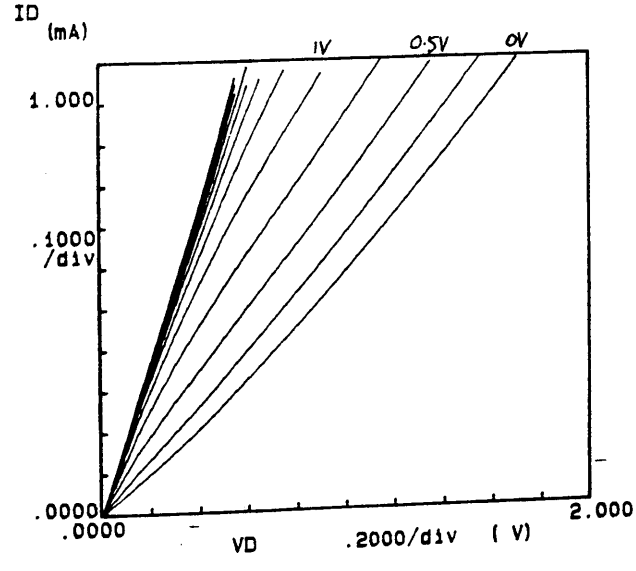
(c) L_g=0.25um

BT-CH3-OX150



(d) L_g=0.17um

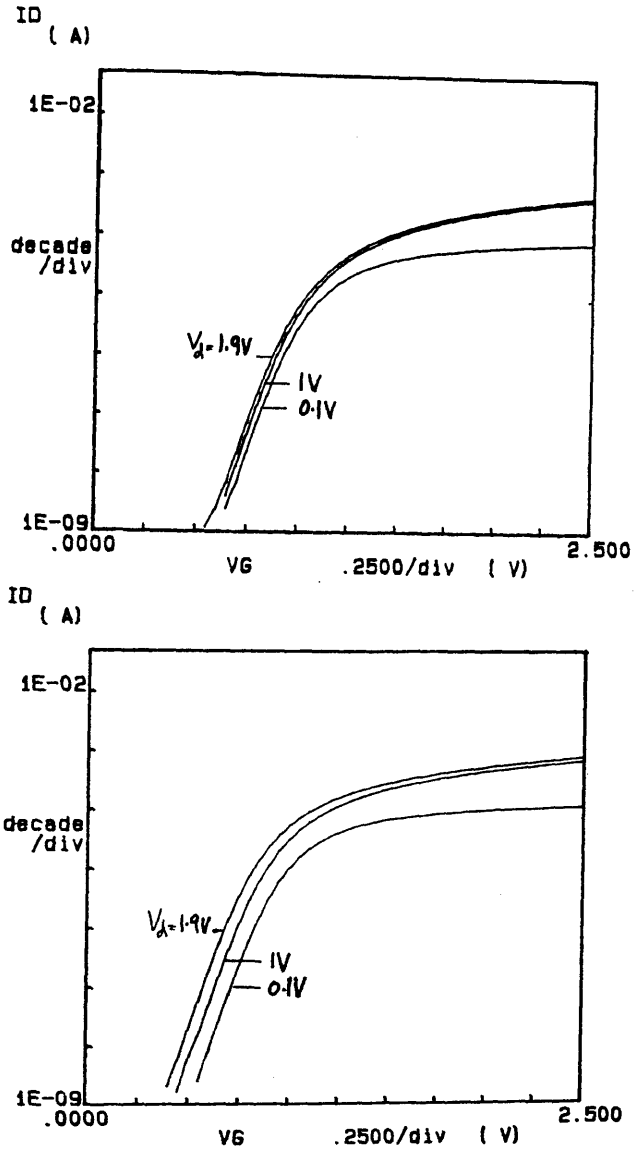
BT-CH3-OX150



(e) L_g=0.08um

BT-CH3-OX150

figure (7-2) cont.



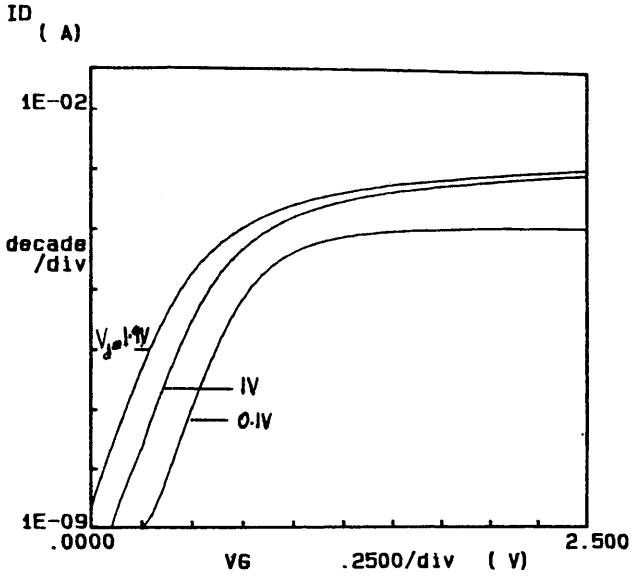
(a) L_g=0.42um

BT-CH3-OX150

(b) L_g=0.33um

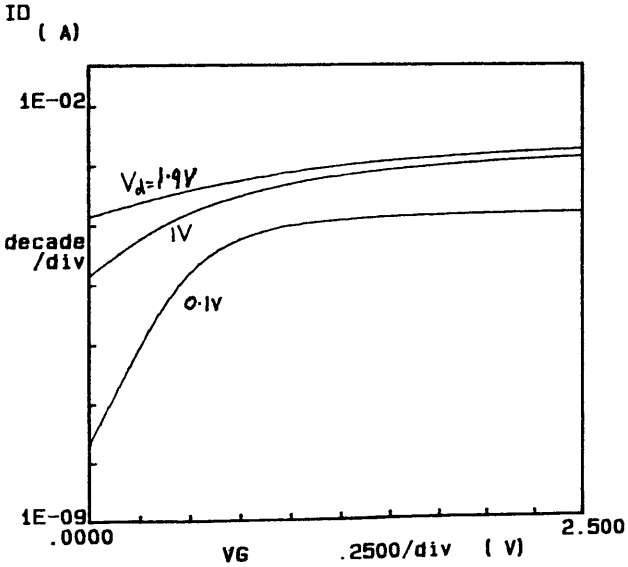
BT-CH3-OX150

Figure (7-3) Subthreshold curves for a set of five devices fabricated using the BT-CH3-OX150 processing conditions. (Cont. overleaf)



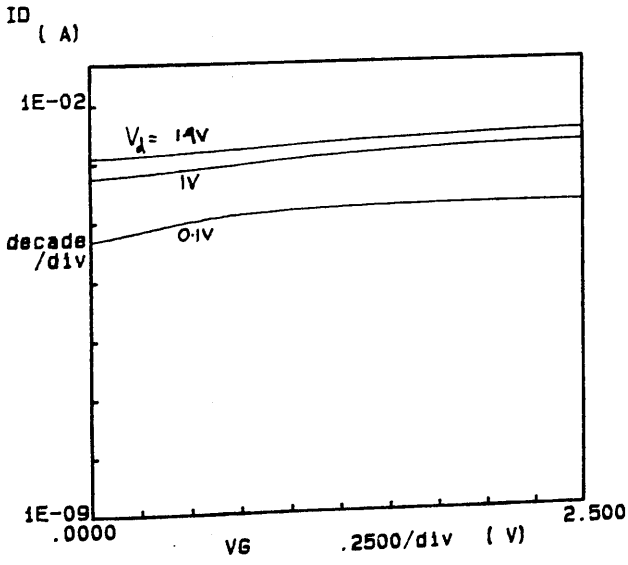
(c) L_g=0.25um

BT-CH3-OX150



(d) L_g=0.17um

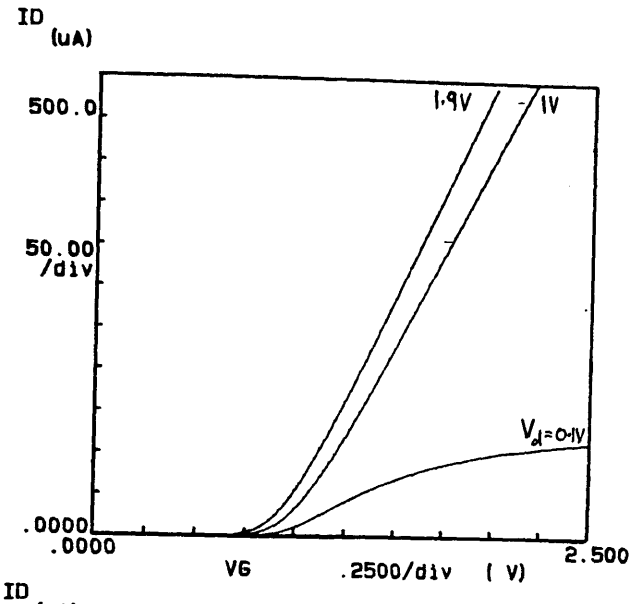
BT-CH3-OX150



(e) L_g=0.08um

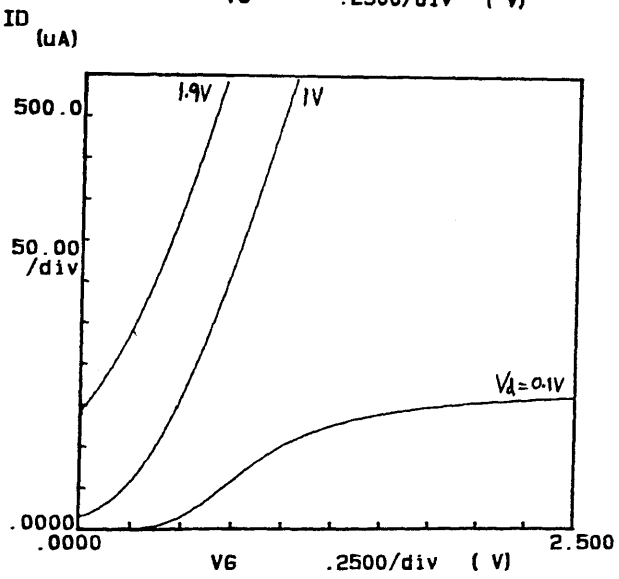
BT-CH3-OX150

figure (7-3) cont.



(a) L_g=0.33um

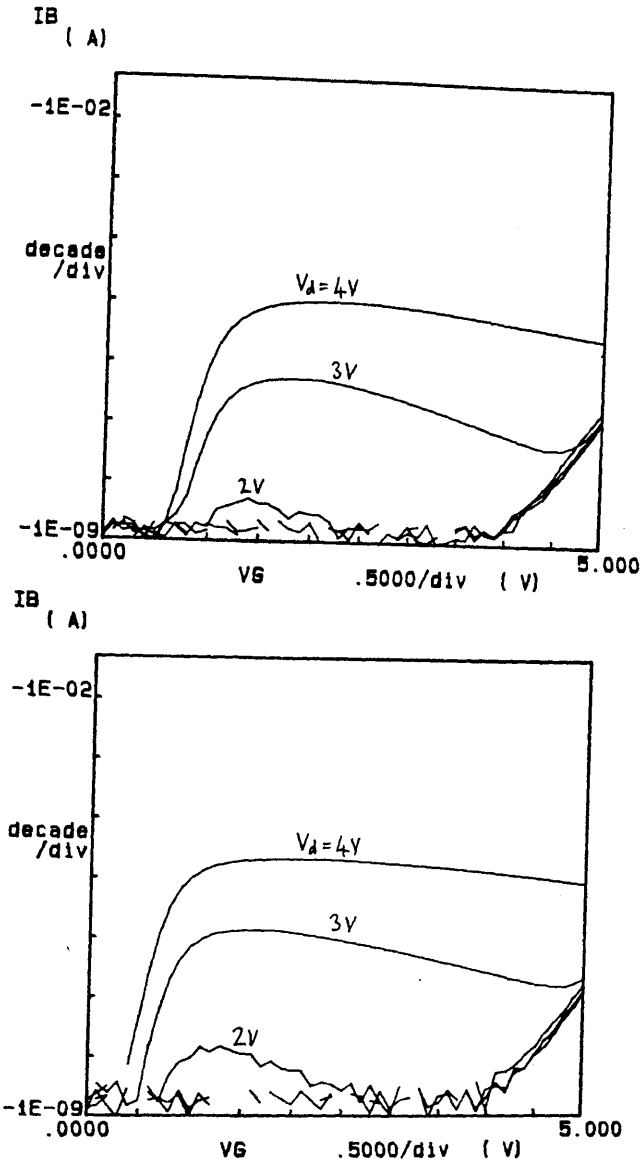
BT-CH3-OX150



(b) L_g=0.17um

BT-CH3-OX150

Figure (7-4) The threshold region curves for two devices with gate lengths (L_g) of 0.33 and 0.17 microns fabricated using the BT-CH3-OX150 processing conditions.



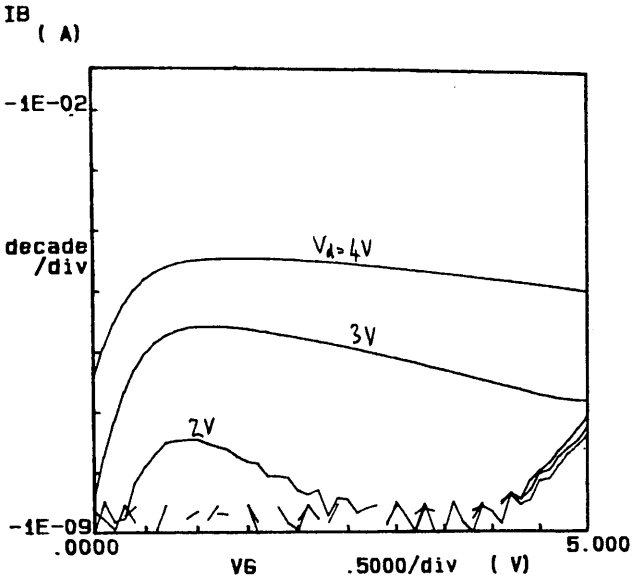
(a) L_g=0.42um

BT-CH3-OX150

(b) L_g=0.33um

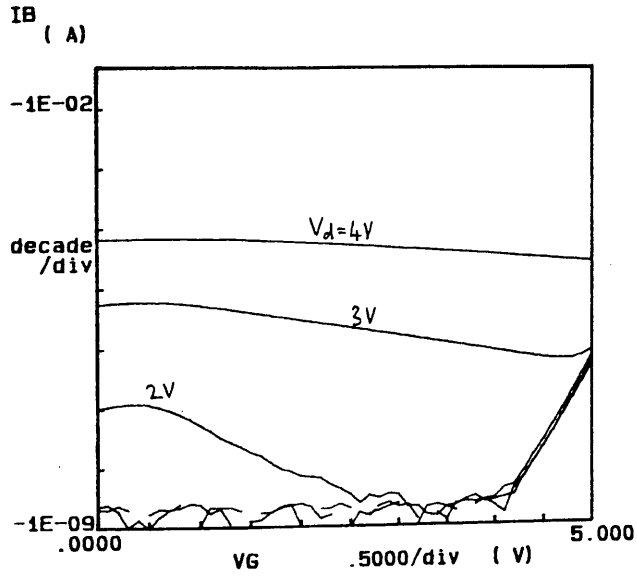
BT-CH3-OX150

Figure (7-5) Substrate current curves for a set of five devices fabricated using the BT-CH3-OX150 processing conditions. (Cont. overleaf).



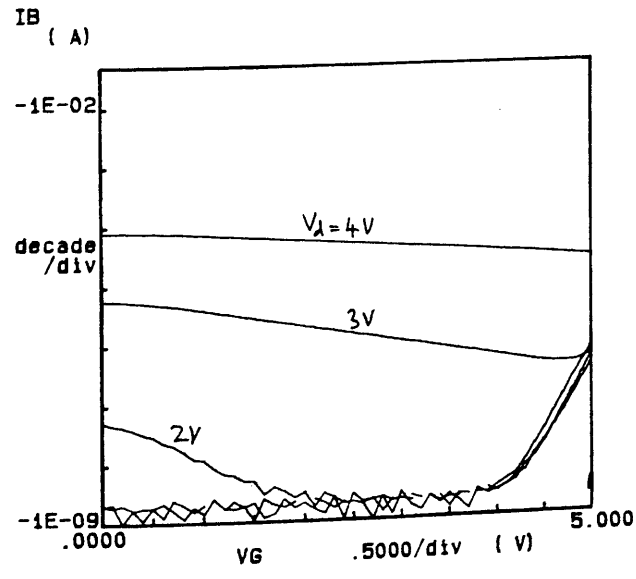
(c) L_g=0.25um

BT-CH3-OX150



(d) L_g=0.17um

BT-CH3-OX150



(e) L_g=0.08um

BT-CH3-OX150

figure (7-5) cont.

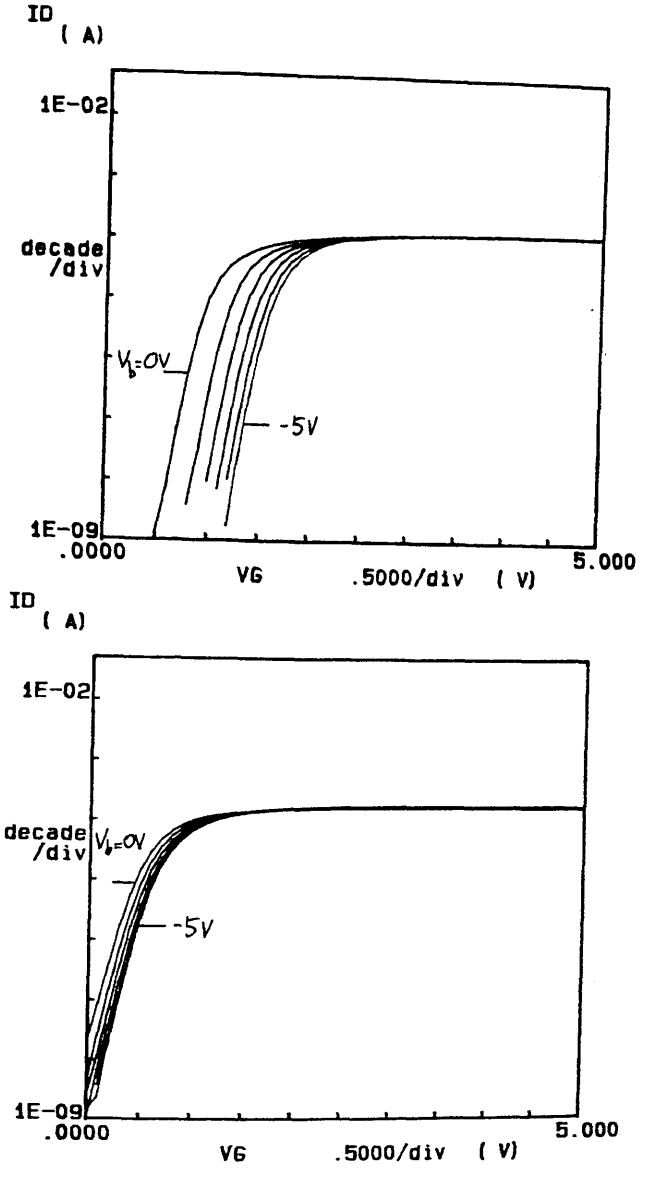
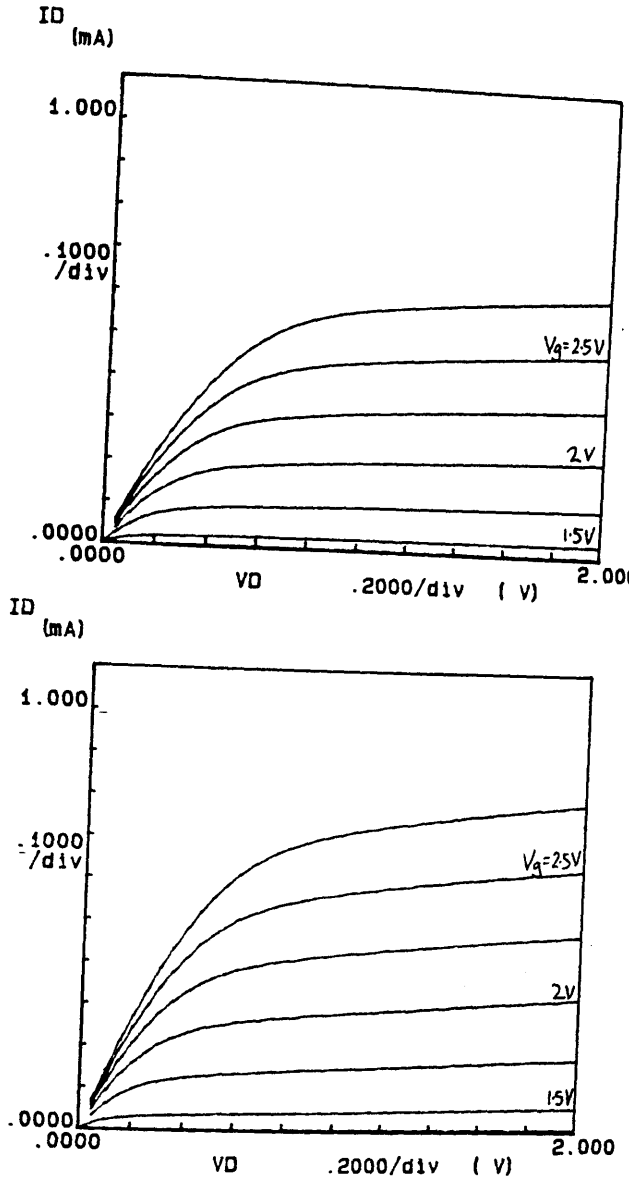


Figure (7-6) Substrate sensitivity curves for two devices with gate lengths (L_g) of 0.33 and 0.17 microns fabricated using the BT-CH3-0X150 processing conditions.



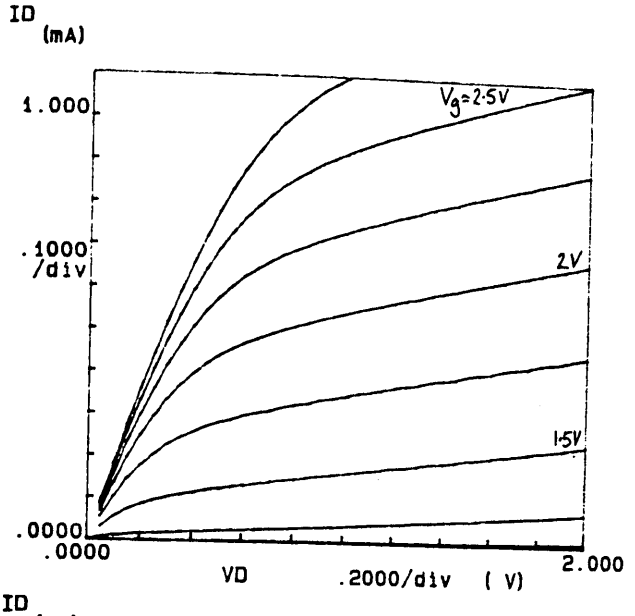
(a) L_g=0.45um

BT-CH4-0X150

(b) L_g=0.36um

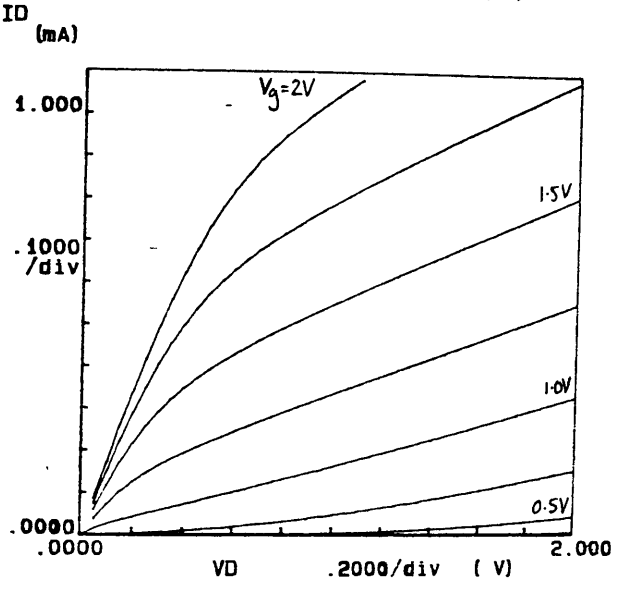
BT-CH4-0X150

Figure (7-7) I-V output curves for five devices which were fabricated using the BT-CH4-0X150 process conditions. (Cont. overleaf)



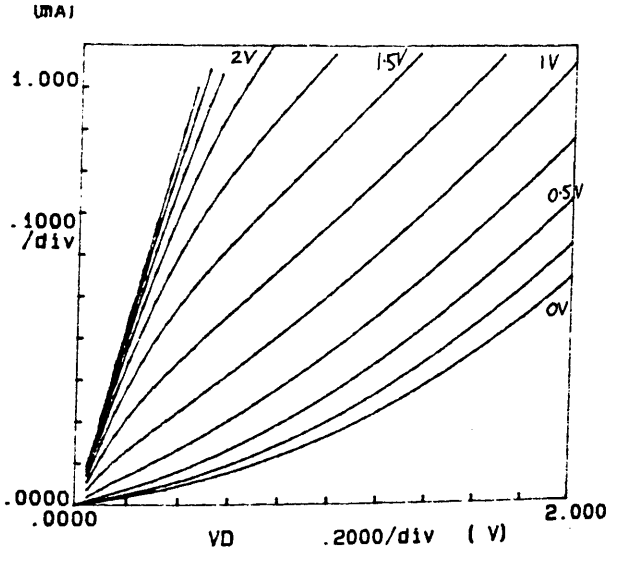
(c) L_g=0.28um

BT-CH4-OX150



(d) L_g=0.20um

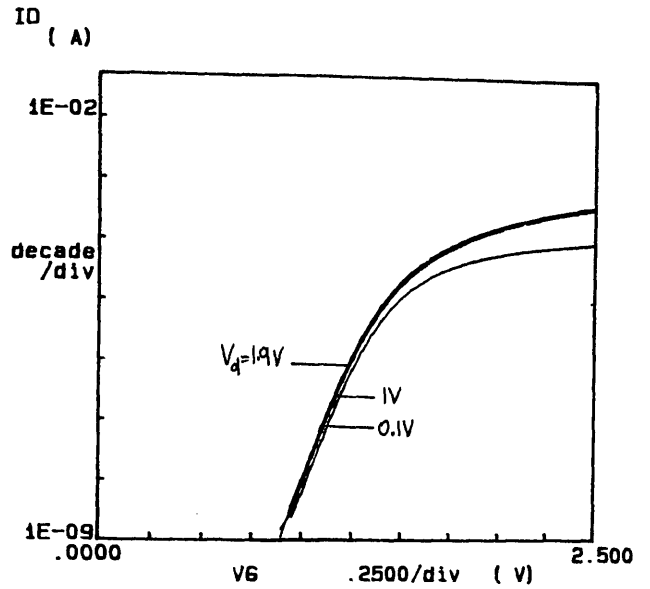
BT-CH4-OX150



(e) L_g=0.11um

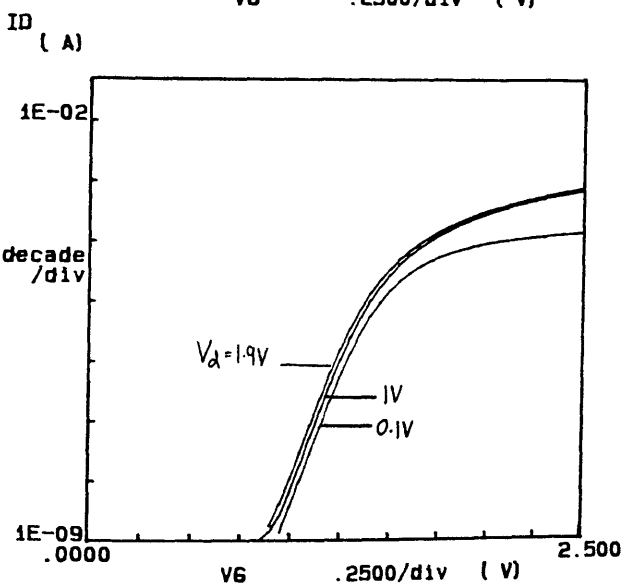
BT-CH4-OX150

figure (7-7) cont.



(a) L_g=0.45um

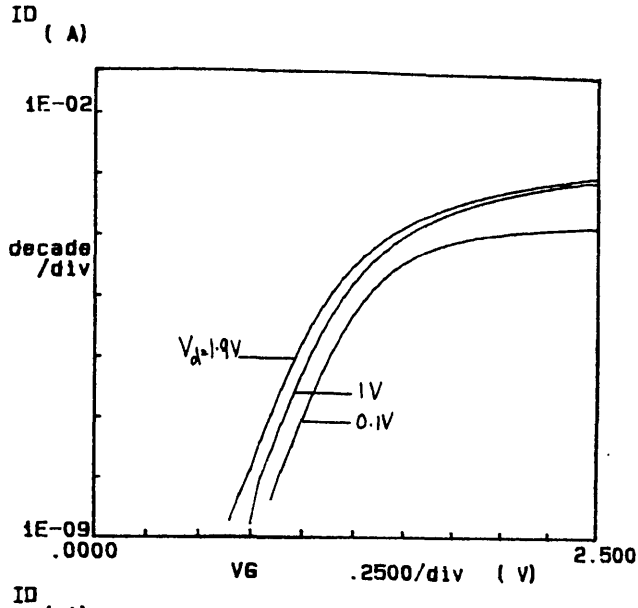
BT-CH4-OX150



(b) L_g=0.36um

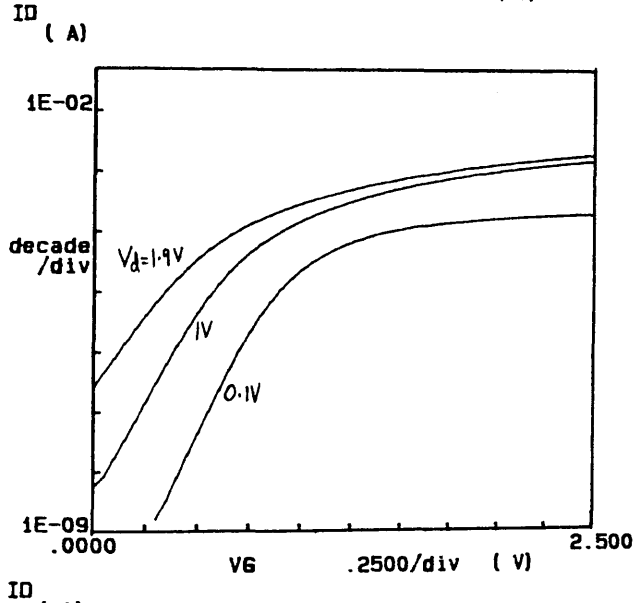
BT-CH4-OX150

Figure(7-8) Subthreshold curves for a set of five devices fabricated using the BT-CH4-OX150 processing conditions. (Cont. overleaf).



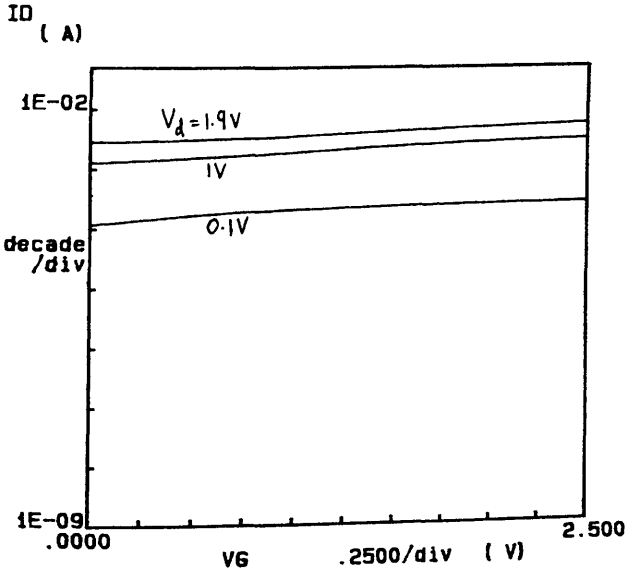
(c) L_g=0.28um

BT-CH4-OX150



(d) L_g=0.20um

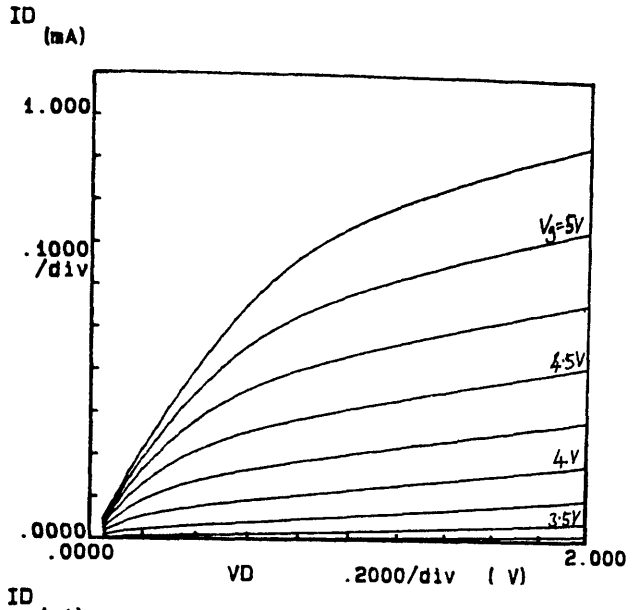
BT-CH4-OX150



(e) L_g=0.11um

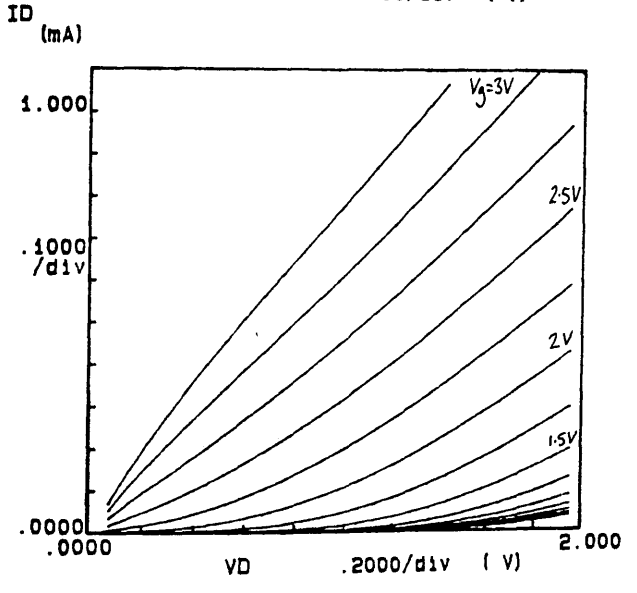
BT-CH4-OX150

figure (7-8) cont.



(a) L_g=0.20um

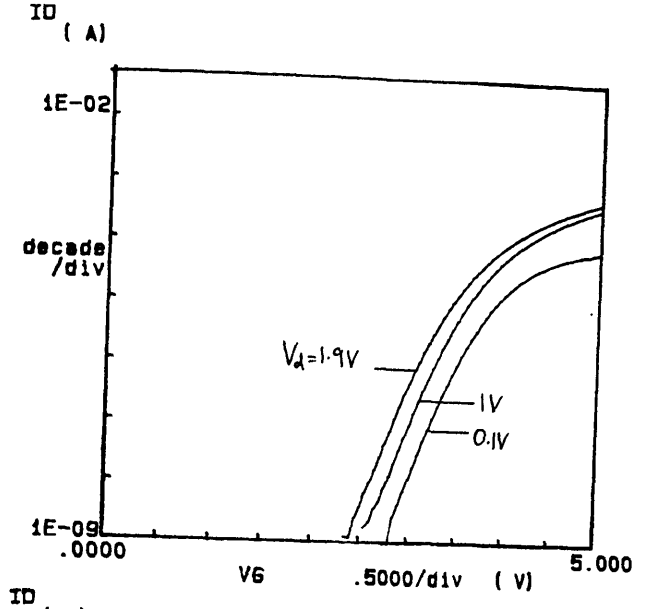
BT-CH6-OX150



(b) L_g=0.11um

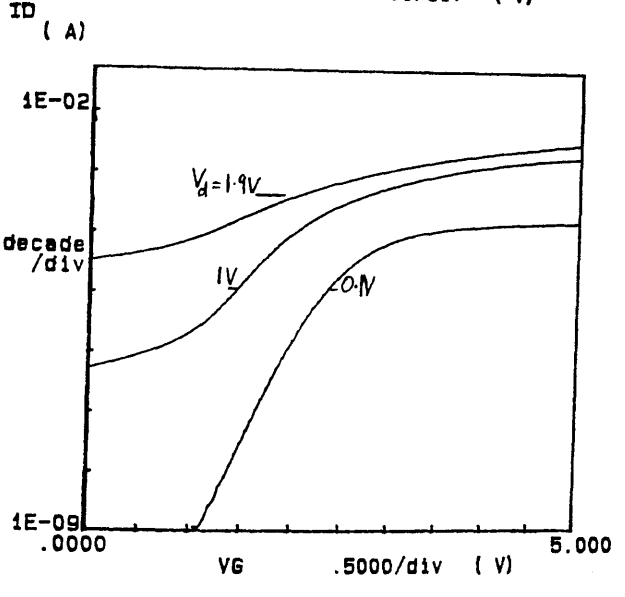
BT-CH6-OX150

Figure (7-9) I-V output curves for two devices which were fabricated using the BT-CH6-OX150 processing conditions.



(a) L_g=0.20um

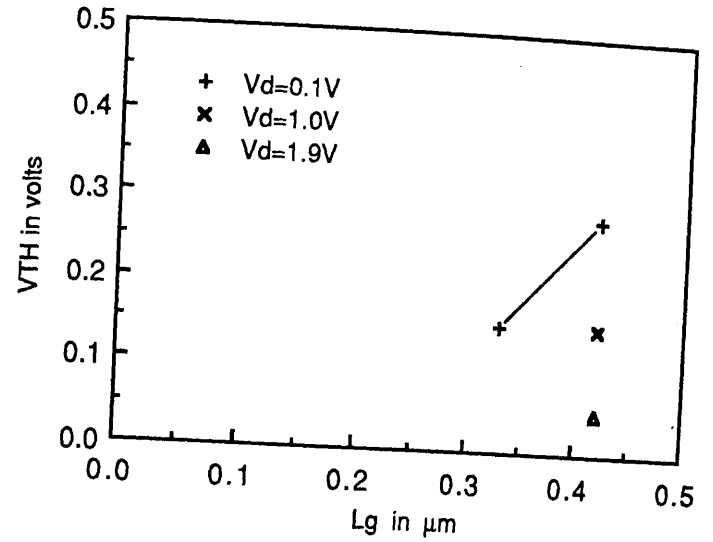
BT-CH6-OX150



(b) L_g=0.11um

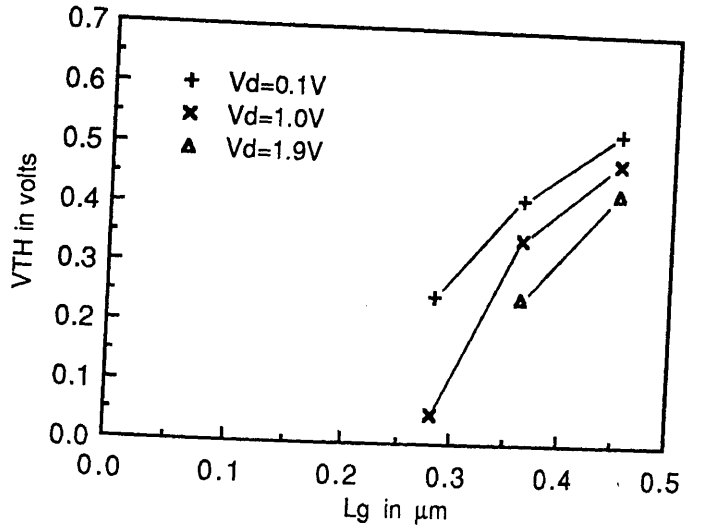
BT-CH6-OX150

Figure (7-10) Subthreshold curves for two devices which were fabricated using the BT-CH6-OX150 processing conditions.



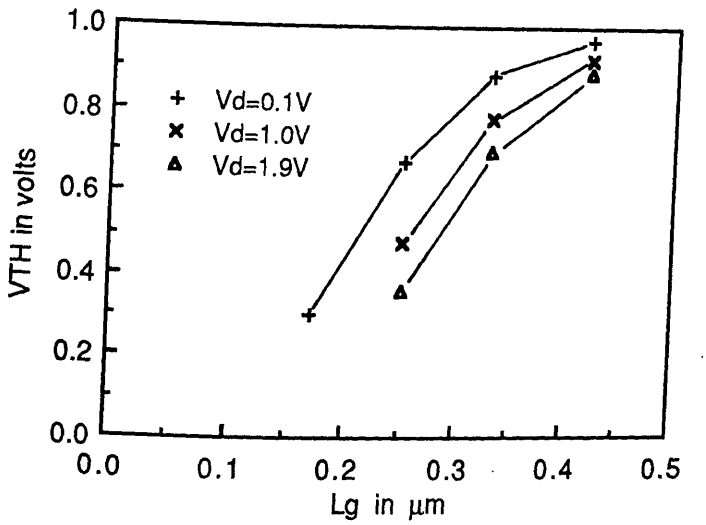
(a)

BT-CH1-OX150



(b)

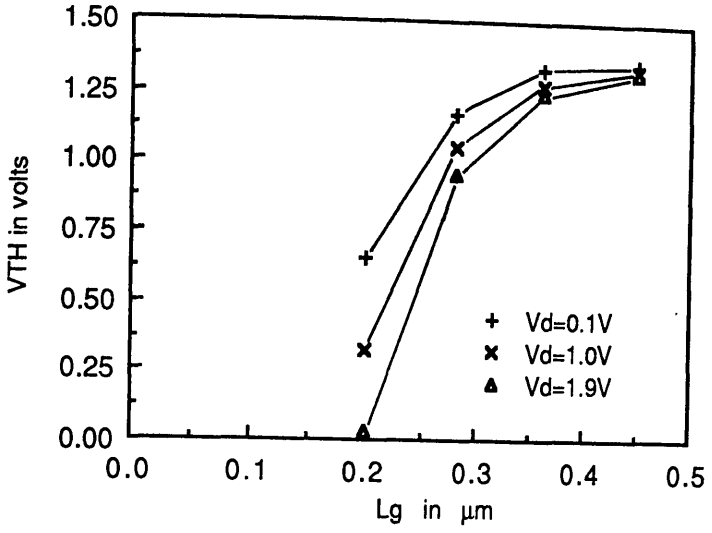
BT-CH2-OX150



(c)

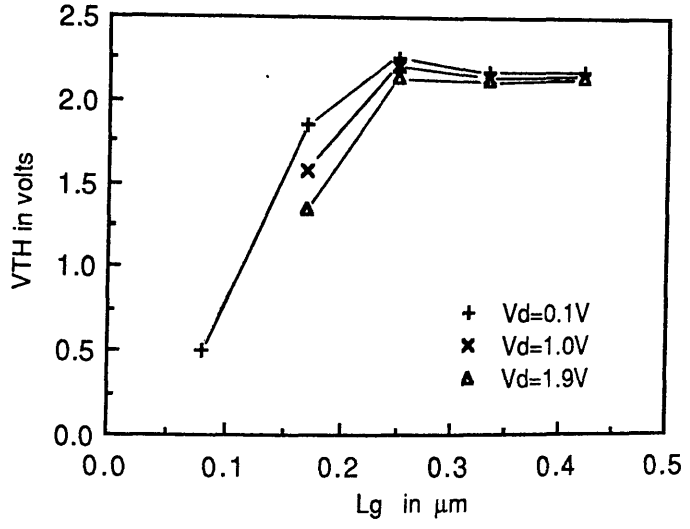
BT-CH3-OX150

Figure (7-11) Six graphs which show measured threshold voltage (V_{TH}) as a function of gate length (L_g) for various drain voltages (V_d). The graphs correspond to the CH1-CH6 implants.



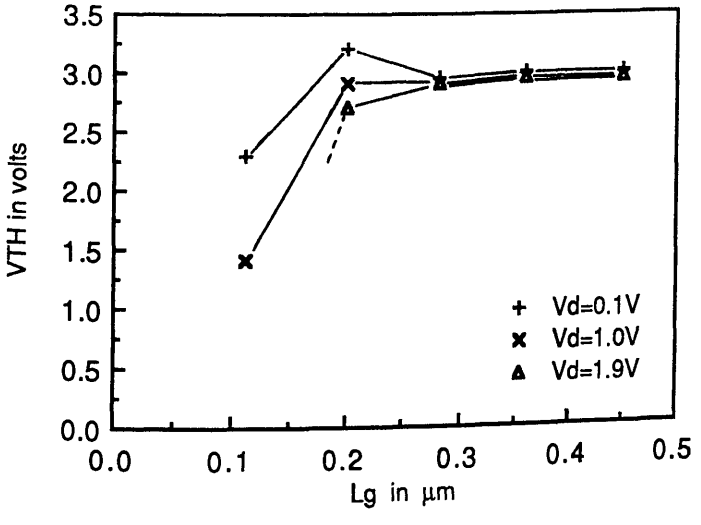
(d)

BT-CH4-OX150



(e)

BT-CH5-OX150



(f)

BT-CH6-OX150

figure(7-11) cont.

	L_g (um)	g_m (mS/mm)	g_d (mS/mm)	A_V
BT-CH3-OX150	0.42	86	5.5	16
"	0.33	90	11	8
"	0.25	96	21	5
"	0.17	88	32	2.7
"	0.08	-	-	-
BT-CH4-OX150	0.45	87	2.5	35
"	0.36	91	5	18
"	0.28	100	10	10
"	0.20	99	27	3.7
"	0.11	-	-	-
BT-CH5-OX150	0.42	65	2	33
"	0.33	70	2.5	28
"	0.25	80	6	13
"	0.17	85	12	7.1
"	0.08	-	-	-
BT-CH6-OX150	0.45	59	2.6	23
"	0.36	60	2.5	24
"	0.28	61	3.8	16
"	0.20	73	13	5.6
"	0.11	72	40	1.8

Table (7-2) Typical electrical parameters for devices with various gate lengths (Lg) and channel implants g_m, g_d & A_V were measured at V_d=2V & I_d=0.5mA.

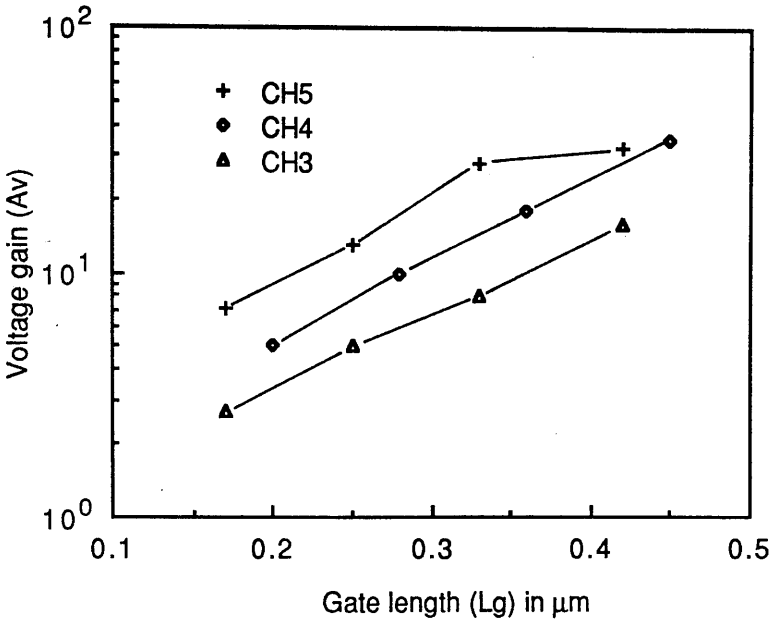


Figure (7-12) Graph showing voltage gain as a function of gate length (L_g) for a set of devices with CH3, CH4 and CH5 channel implants. (data from table (7-2)).

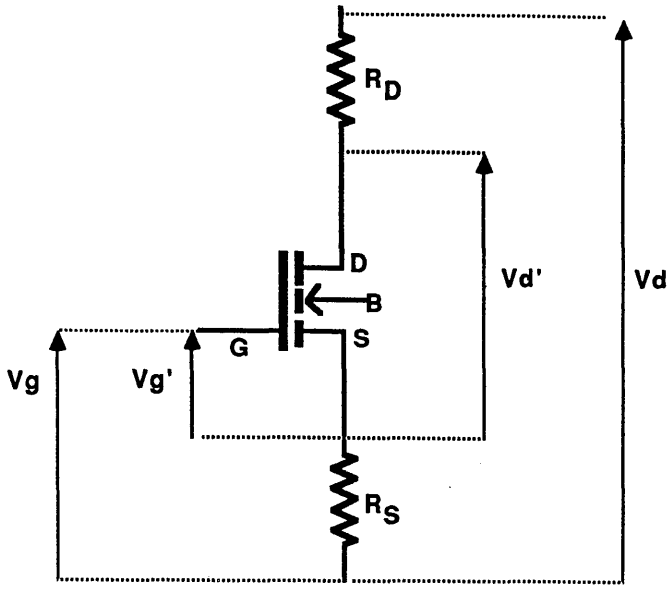


Figure (7-13) Diagram showing MOSFET with parasitic source/drain resistance. Dashed voltage terms are also indicated.

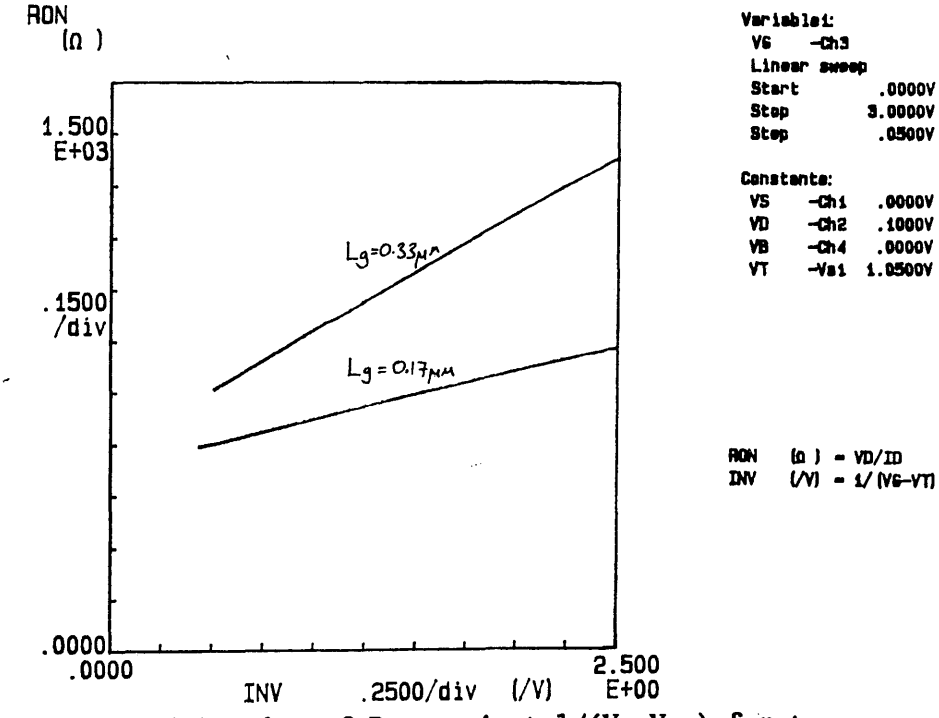
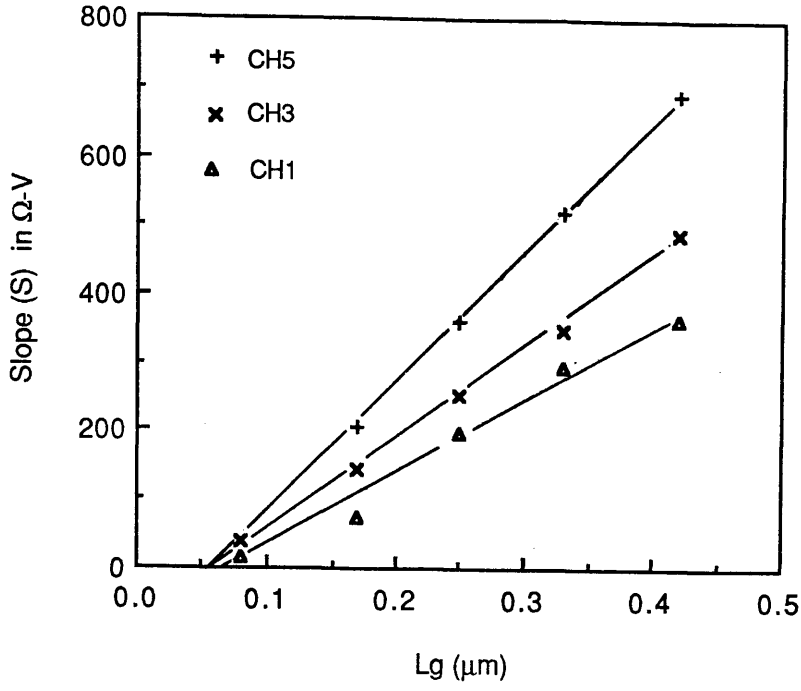
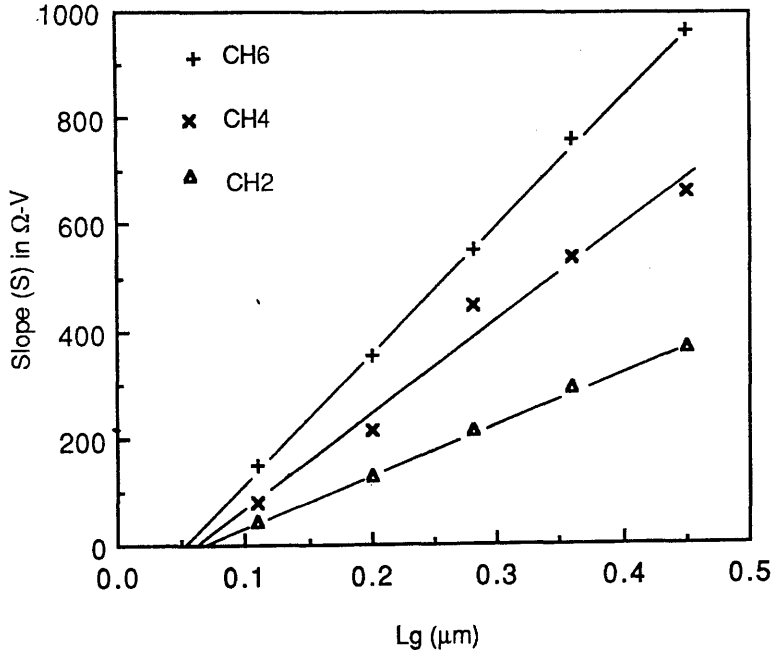


Figure (7-14) Plot of R_{ON} against $1/(V_g-V_{TH})$ for two CH3 devices with gate lengths (L_g) of 0.17 and 0.33 microns.



(a) Results for wafer CH1/3/5



(b) Results for wafer CH2/4/6

Figure (7-15) Two graphs showing the measured slope parameter (S) as a function of gate length (L_g). Figures (a) & (b) show the results for the CH1/3/5 & CH2/4/6 wafers respectively.

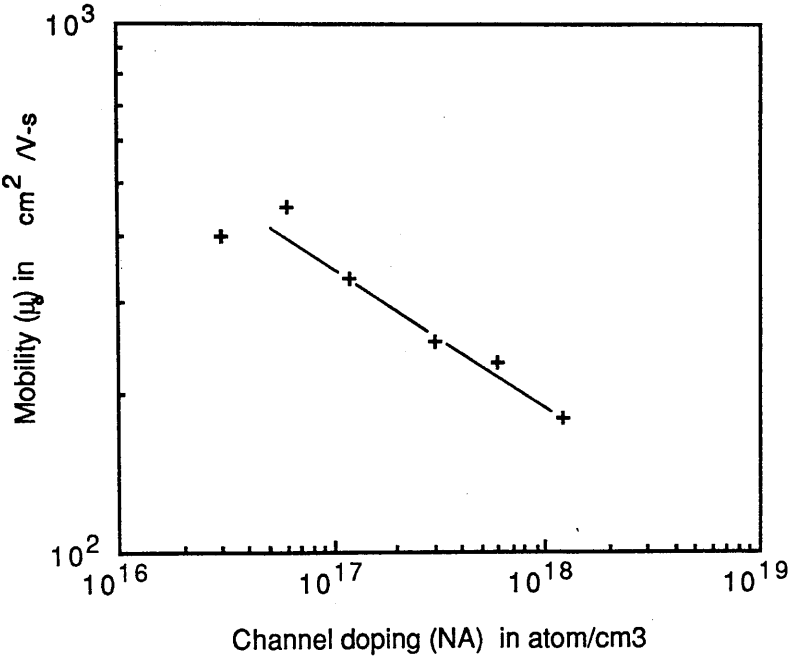
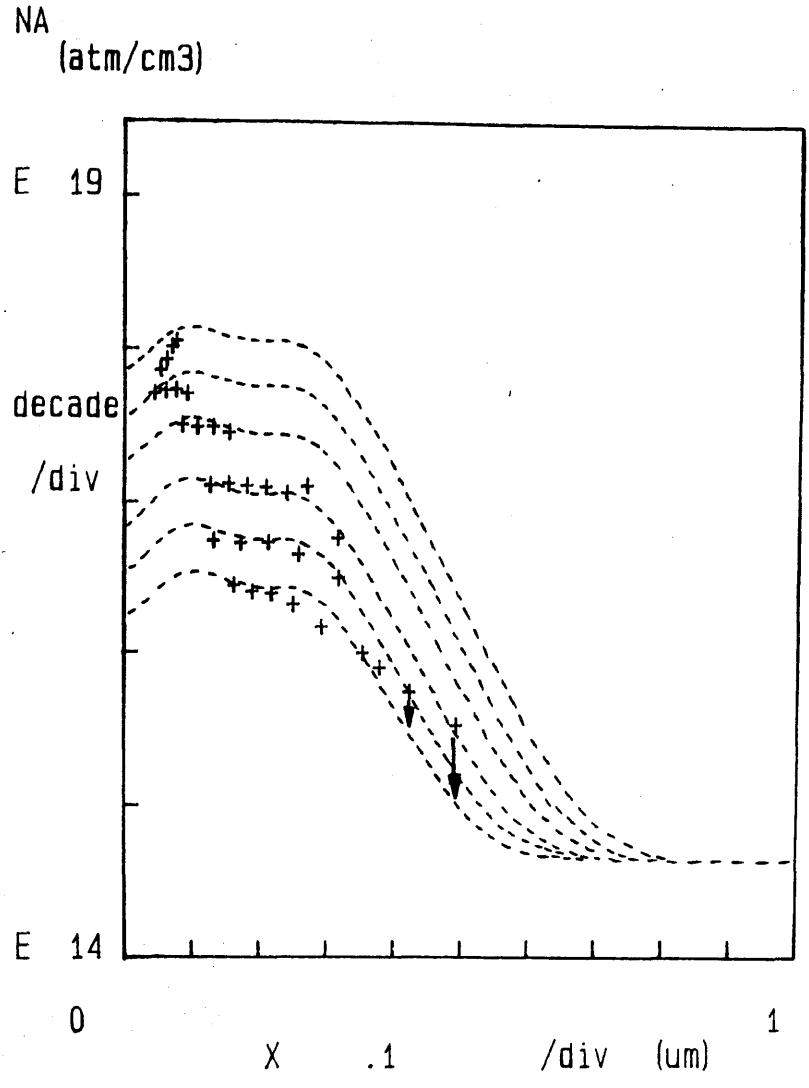


Figure (7-16) Low field mobility (μ_o) as a function of channel doping level.



+ from Beuhler method

---- from SUPREM II

Figure (7-17) Channel doping profiles for the (CH1-CH6) determined from electrical measurements using the Beuhler method (for t_{ox}=150A). The results of SUPREM II simulations are shown for comparison.

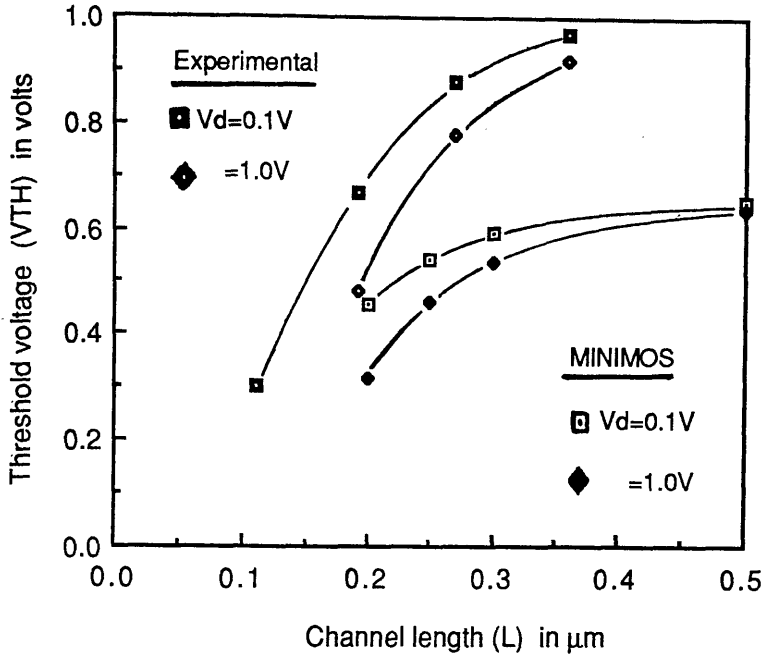


Figure (7-18) Graph showing threshold voltage as a function of channel length for the CH3 devices. MINIMOS results are shown together with the experimental results from figure (7-11).

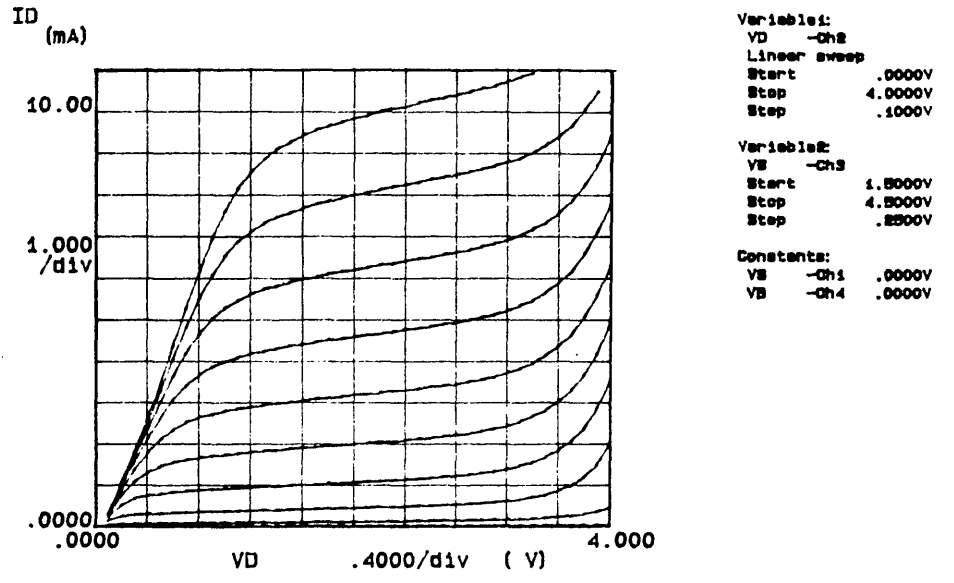


Figure (7-19) I-V characteristic for BT-CH5-OX150 device with a gate length (L_g) and gate width (W_g) of 0.25 and 100 um respectively. The wide gate was obtained using the inter-digitated structure shown earlier in figure (6-8). The channel length is approximately 0.19 um.

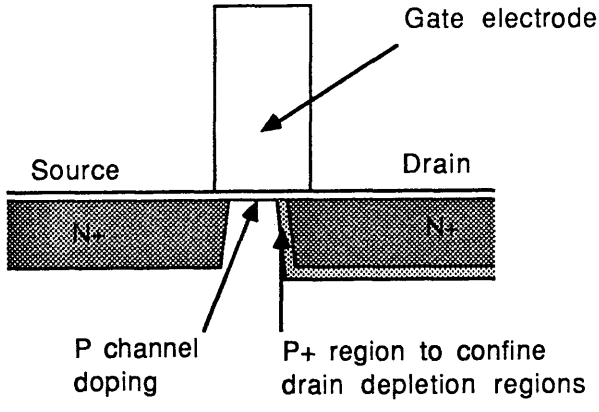


Figure (7-20) Proposed structural enhancement for reducing short channel effects in MOSFETs with short physical gate lengths.

Chapter 8

E.M.F. Device Results

8.1 Introduction

A second set of MOSFETs have been fabricated with sub 0.5 micron gate lengths. These devices were fabricated using the alternative E.M.F. process route and the new EU567 mask set. The gates were patterned using the new $SiCl_4$ reactive ion etching process. Both discrete devices and ring oscillator circuits have been implemented.

8.1.1 Objectives of the E.M.F. route

The E.M.F. process route was initiated 24 months after the start of the project for four reasons.

Firstly, it was considered too risky for the whole investigation to depend upon the successful processing of a single wafer batch, particularly in view of the extended fabrication timescales involved.

Secondly, the new process route could be used, together with a new mask set, to implement ring oscillator circuits.

Thirdly, the route could be used to test the new polysilicon $SiCl_4$ reactive ion etching process, developed at Glasgow.

Additionally, the process route was established with future collaborative research in mind.

Unfortunately, due a serious laboratory fire at Glasgow, the $SiCl_4$ reactive ion etching system was non-operational for a period of approximately six months. This interrupted processing of the E.M.F. wafers which was finally completed 39 months after the start of the project. Consequently only a limited evaluation has been possible.

8.1.2 The Process Matrix

Results are presented for devices with 150 angstrom gate oxides and CH3, CH4 and CH5 channel implants. These implants have corresponding channel doping levels of $1.2x10^{17}$, $3x10^{17}$ and $6x10^{17}$ atoms/cm^3.

8.1.3 Physical Gate Length Measurements

The calibration test patterns were cleaved and examined by SEM to determine physical gate lengths. Tapered gate structures were obtained, with linewidths approximately 0.7 microns wider at the base than at the top, consistent with the etching experiments described in chapter 4. Figure (8−1a) shows an electron micrograph of the resulting gate structure together with the P−glass dielectric layer. Unfortunately, the thin sidewall oxide is not resolved.

Table (8−1) shows the linewidths of the polysilicon calibration lines. The measurements correspond to the polysilicon widths at the top and base of the tapers. The latter values were used to define the device gate lengths (L_g) as indicated by figure (8−1b). No linewidth or profile variation was observed between the three wafers and once again the experimental error associated with the gate length measurements was estimated at less than 10%.

8.2 ED−CH3−OX150 Wafer Results

Figures (8−2) and (8−3) show the I−V output and subthreshold characteristic curves for three devices fabricated with CH3 channel implants. Each device has a gate width (W_g) of 10 microns and a series source/drain resistance of 200−250 ohms. The gradual onset of short channel behaviour occurs as gate length decreases from 0.39 to 0.20 microns. It will be seen in chapter 9 similar devices with 0.39 and 0.23 micron gate lengths were implemented in the ring oscillator circuits.

A comparison between the 0.20 micron device and the earlier B.T.R.L. processed 0.17 micron device (see figure (7−2d)) indicates

that both devices have similar I– V characteristics, showing punch– through at drain voltages (V_d) of approximately 0.5 volts. This demonstrates a high degree of consistency between the device results despite many processing variations.

The attainment of functioning devices demonstrates that two previously untested techniques have been applied successfully to submicron device fabrication. Firstly, the new $SiCl_4$ reactive ion etching process developed at Glasgow is suitable for submicron gate patterning and secondly rapid furnace anneal technique developed at the E.M.F. can be used to anneal the source/drain drift regions for devices at these dimensions.

8.3 ED– CH4– OX150 Wafer Results

Figures (8– 4) and (8– 5) show I– V and subthreshold curves for devices with CH4 channel implants and gate lengths of 0.39 and 0.23 microns. A general comparison with the earlier B.T.R.L. devices of similar gate lengths and channel implants (see figure (7– 7)) indicates three differences. Firstly, the later devices show a slightly improved output conductance in the saturation region for comparable gate lengths. Secondly, the devices show degraded transconductance with values approximately 30% lower than for the B.T.R.L. devices. This occurs despite a lower parasitic source/drain resistance and therefore suggests a lower carrier saturation velocity (see equation (2.20)). Finally, the later devices show a more gradual transition from the linear region to the satuaration region indicating a reducted carrier mobility.

The reduction in carrier saturation velocity and mobility can be attributed to the change in channel doping profiles for the earlier and later devices. SUPREM simulation results indicate an increased doping level at the oxide interface for the later devices due to the increased high temperature anneal step (see section 5.5.2). The channel carriers are confined within a very narrow layer close to interface and hence increased interface doping levels will reduce both mobility and saturation velocity.

8.4 ED– CH5– OX150 Wafer Results

Figures (8– 6) and (8– 7) show I– V and subthreshold curves for devices with CH5 channel implants. The 0.23 micron device shows long channel behaviour with a transconductance (g_m) of 55 mS/mm and an output conductance (g_d) of 4.5 mS/mm corresponding to a voltage gain (A_v) of 12. This is consistent with the results for the earlier devices where it was demonstrated that a gate length of 0.25 microns resulted in long channel behaviour for the same channel implant parameters.

Results are also shown for a device with a gate length (L_g) of only 0.15 microns. This device exhibits a transconductance (g_m) of 50 mS/mm and an output conductance (g_d) of 19 mS/mm giving a voltage gain (A_v) of 2.6. It is interesting to note that no punch– through current is indicated despite the extemely short gate length.

Due to time limitations it has not been possible to accurately determine the channel length reduction ($\triangle L$) for the later devices and so a detailed comparison with the earlier devices has not been possible.

8.5 Discussion

One preliminary fabrication run has been completed using the E.M.F. process route and functioning devices have been demonstrated. This has confirmed the use of two experimental processing techniques. Firstly,the new $SiCl_4$ reactive ion etching process has been applied successfully to device fabrication and minimum gate lengths (L_g) of 0.15 microns have been obtained. Secondly, the rapid furnace anneal technique undertaken at the E.M.F. has been demonstrated as a suitable alternative to Heatpulse annealing of shallow arsenic source/drain implants.

The effects of tapered gate profiles and gate sidewall oxides upon device performance have not been identified and will require further investigation under more controlled conditions.

The results suggest that a reduced channel doping level at the oxide interface is desirable for optimising carrier mobility and

saturation velocity for MOSFETs with short physical gate lengths. However, a rapid increase in doping level with depth is required in order to suppress short channel effects.

The E.M.F. device results should be considered in the context of a wider research programme. Having demonstrated the successful implementation of the new process and the new mask set a further study may be undertaken to investigate the use of thinner gate oxides and silicide techniques[8.1]. Thinner oxides are required to reduce threshold voltages whilst silicide source/drain and gate structures can be used to reduce parasitic resistances which are seen to be relatively high for the devices fabricated here. Silicide formation at the source, drain and gate can be achieved using a self–aligned sidewall spacer technique[8.2,8.3,8.4].

References

8.1 S.P. Murarka, 'Silicides for VLSI applications,' Academic Press, 1983.

8.2 H. Okabayashi et al., 'Low resistance MOS technology using self–aligned refactory silicidation, IEDM Tech. Dig., p556, 1982.

8.3 M.E. Alperin et al., 'Development of the self–aligned titanium silicide process for VLSI applications,' IEEE J. Solid–State Circuit, vol SC–20, p61, 1985.

8.4 M.E. Alperin, T.C. Holloway, R.A. Haken, C.D. Gosmeyer, R.V. Karnaugh and W.D. Parmantie, 'Development of the self–aligned titanium silicide process for VLSI applications, IEEE Trans. Electron Devices, vol ED–32, p141, 1985.

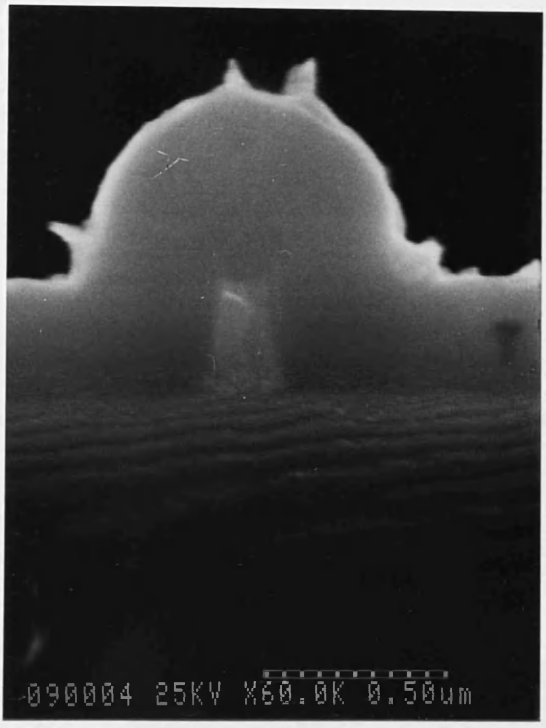


Figure (8-1a) Electron micrograph showing a tapered polysilicon calibration line from the ED-CH3-OX150 wafer, for 8 pixel exposure.

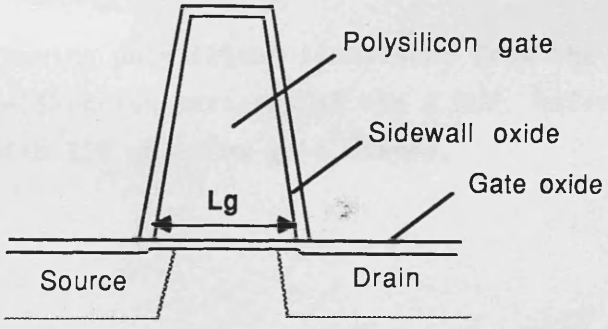
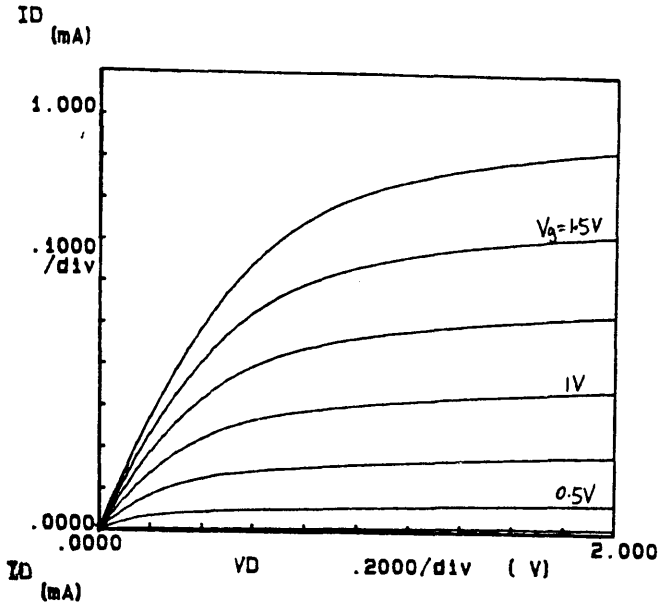


Figure (8-1b) Diagram showing tapered gate structure together with definition of device gate length (L_g).

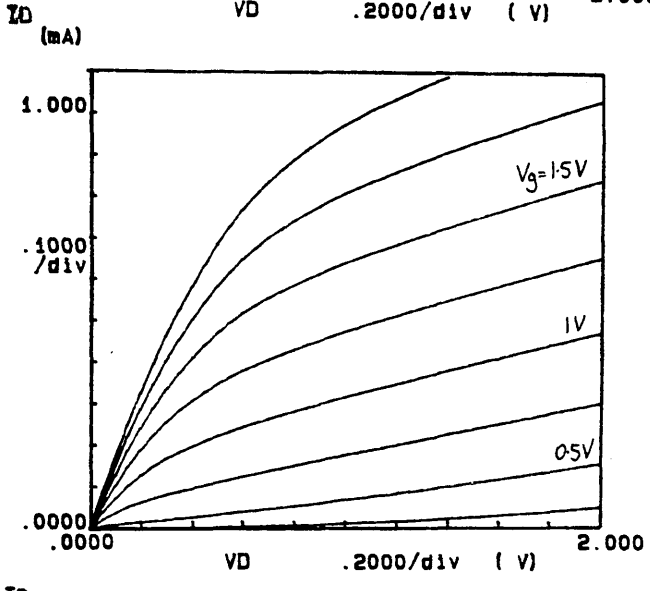
EBL Exposure pixel number at 100x80um frame size	Calibration linewidth at base of taper (um)	Calibration linewidth at top of taper (um)
5	0.15	0.08
8	0.20	0.13
10	0.23	0.16
15	0.31	0.24
20	0.39	0.32
27	0.49	0.42

Table (8-1) **Showing polysilicon linewidths from the calibration patterns of the E.M.F. wafers with 150 angstrom gate oxides.**



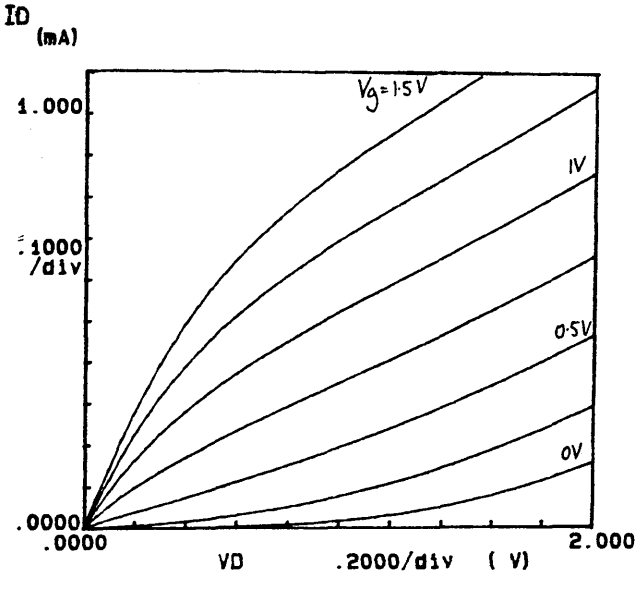
(a) L_g=0.39um

ED-CH3-OX150



(b) L_g=0.23um

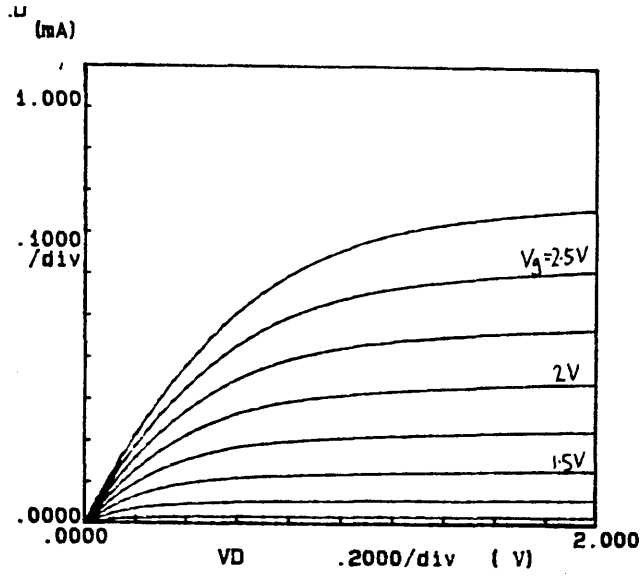
ED-CH3-OX150



(c) L_g=0.20um

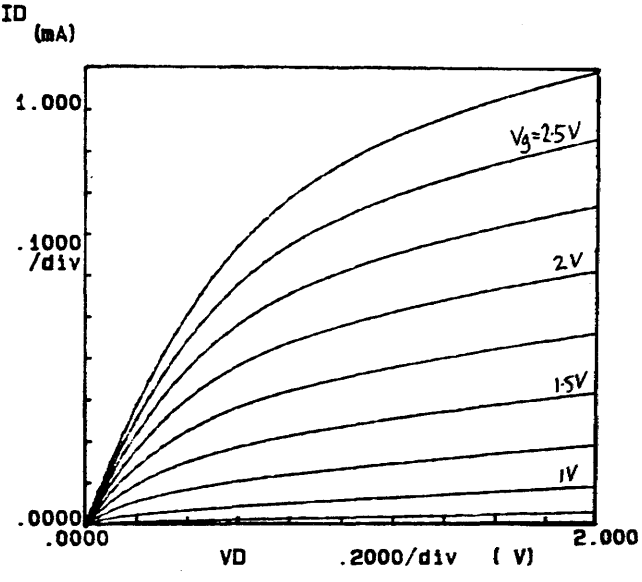
ED-CH3-OX150

Figure (8-2) I-V output curves for a set of three devices from the ET-CH3-OX150 wafer.



(a) L_g=0.39um

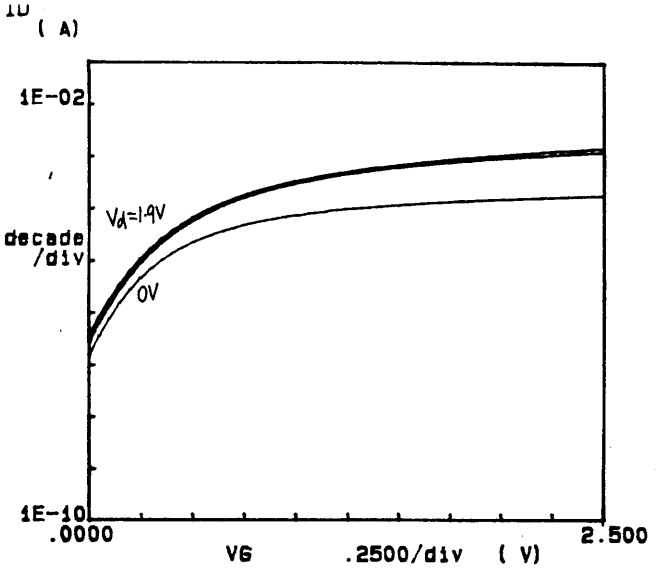
ED-CH4-OX150



(b) L_g=0.23um

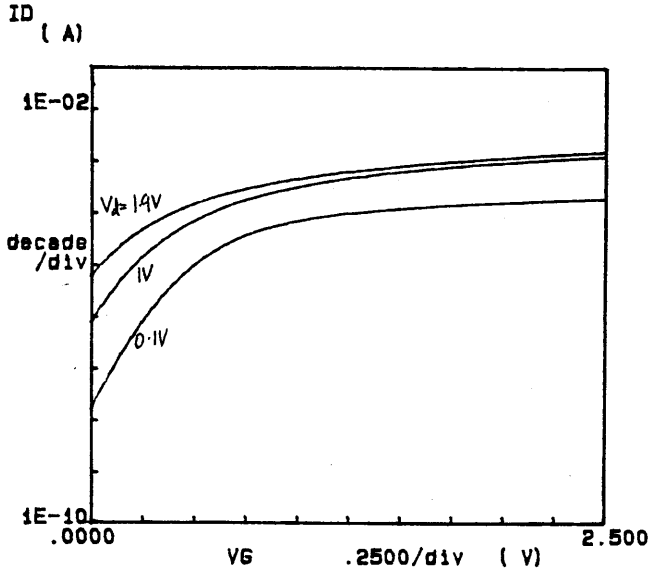
ED-CH4-OX150

Figure (8-4) I-V curves for two devices fabricated on the ED-CH4-OX150 wafer.



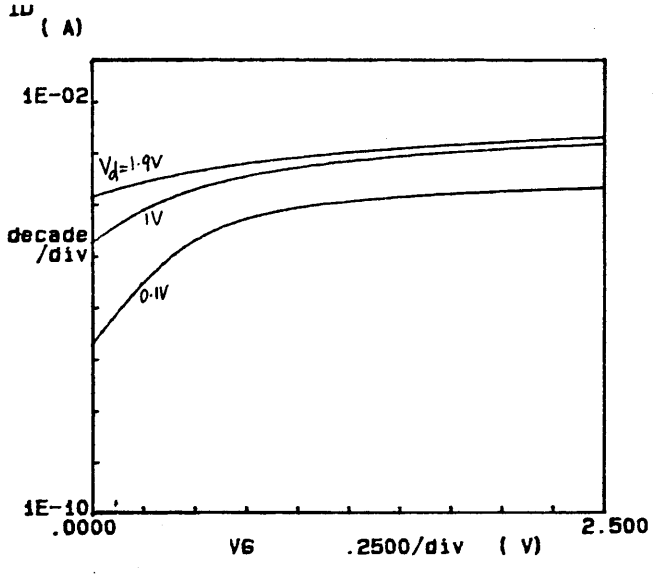
(a) L_g=0.39um

ED-CH3-OX150



(b) L_g=0.23um

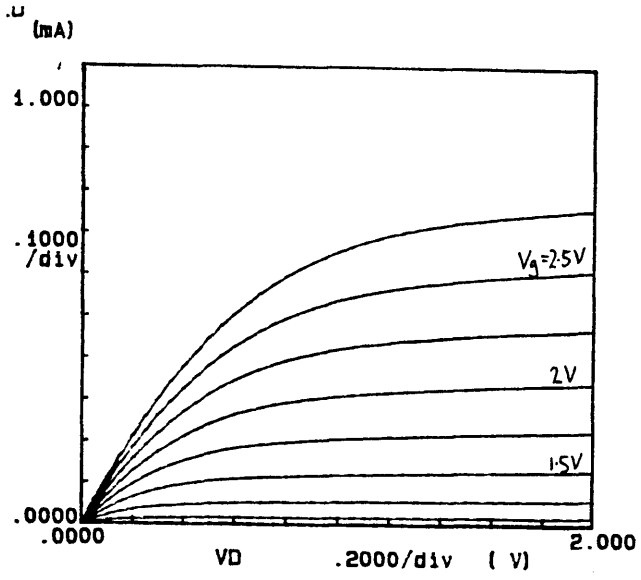
ED-CH3-OX150



(c) L_g=0.20um

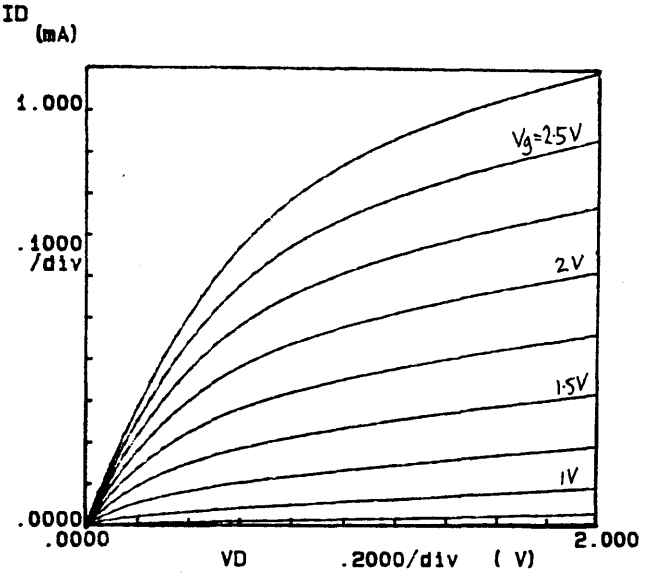
ED-CH3-OX150

Figure (8-3) Subthreshold curves for a set of three devices from the ED-CH3-OX150 wafer.



(a) L_g=0.39um

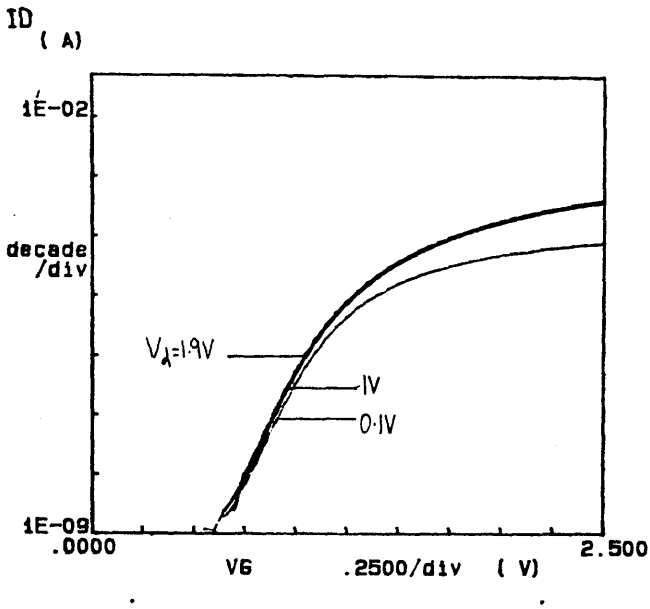
ED-CH4-OX150



(b) L_g=0.23um

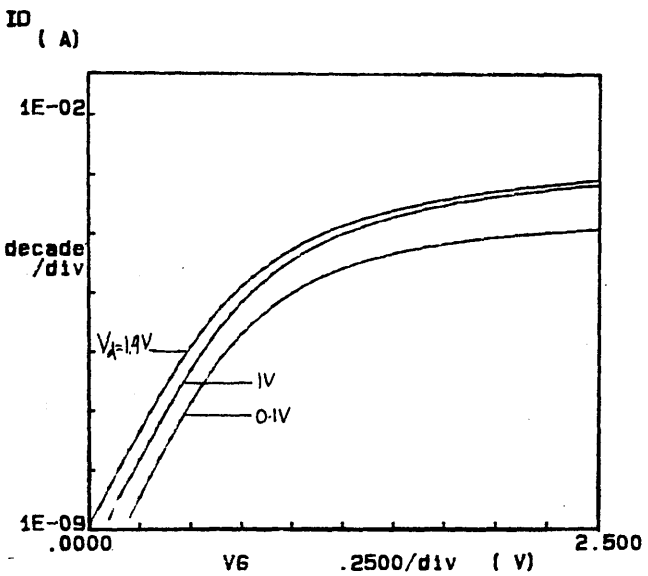
ED-CH4-OX150

Figure (8-4) I-V curves for two devices fabricated on the ED-CH4-OX150 wafer.



(a) L_g=0.39um

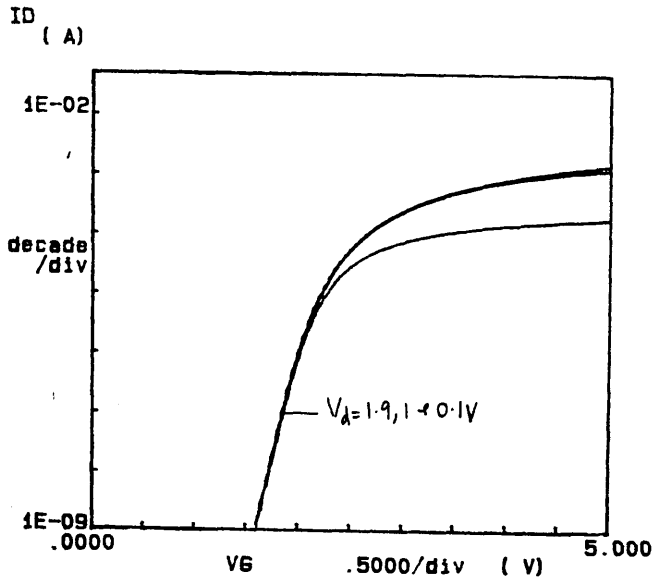
ED-CH4-OX150



(d) L_g=0.23um

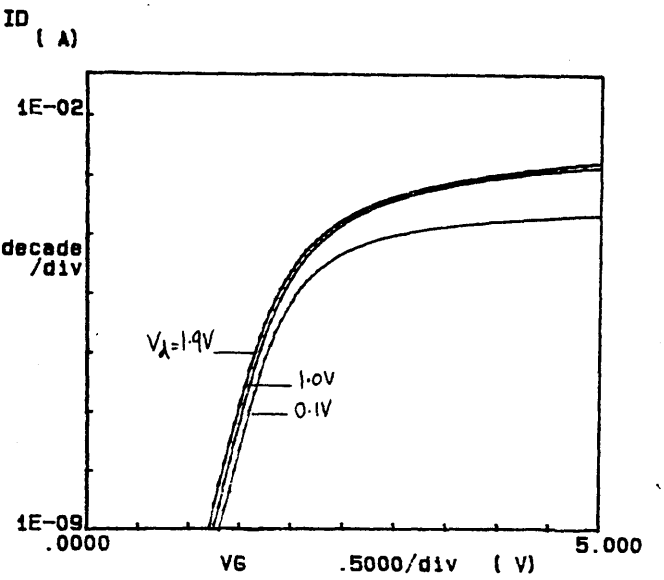
ED-CH4-OX150

Figure (8-5) Subthreshold curves for two devices fabricated on the ED-CH4-OX150 wafer.



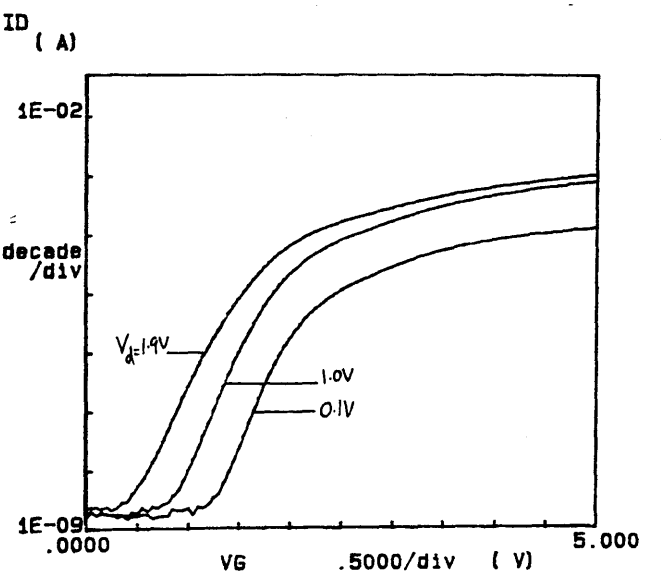
(a) L_g=0.39um

ED-CH5-OX150



(b) L_g=0.23um

ED-CH5-OX150



(c) L_g=0.15um

ED-CH5-OX150

Figure (8-7) Substrate curves for a set of three devices from the ED-CH5-OX150 wafer.

Chapter 9

E.M.F. Results – Ring Oscillators

9.1 Introduction

Operational ring oscillator circuits have been fabricated using the E.M.F. process route. The resulting high speed measurements are presented in this chapter for circuits with active device gate lengths of 0.23 microns and 0.39 microns, implemented on the ED–CH3–OX150 wafer.

9.2 Ring Oscillator Design

Ring oscillators are extremely useful for establishing minimum stage delays for an experimental device technology[9.1]. The circuits fabricated for this study consisted of nineteen unloaded inverter stages and a three stage buffered output circuit as indicated in figure (9–1). Each inverter stage was designed for enhancement mode operation of the active devices and depletion mode operation of the load devices. A single inverter stage is show in figure (9–2).

Figure (9–3) shows the mask layout of the oscillator circuit together with the output buffer circuit. It can be seen that the nineteen inverter stages have gate widths of 10 microns whilst the final output stage has a total gate width of 50 microns, giving a five–fold increase in current drive. This increase is required to ensure that the oscillator can drive the parasitic capacitance of the bond pad and the input impedance of monitoring equipment. Figure (9–4) shows an optical micrograph of a completed ring oscillator with gate lengths of 0.23 microns for active devices and 0.20 microns for the load devices. Three E.B.L. exposure fields were used to pattern each ring oscillator, requiring three sets of reduced field registration marks. Single inverter stages were also implemented to permit d.c. evaluation.

The inverters and ring oscillators were fabricated with active device gate lengths of 0.39 and 0.23 microns. Unfortunately, the

arsenic depletion implant dose chosen for the ED–CH3–OX150 wafer was too small to achieve depletion mode operation for the load devices. This problem was identified prior to the gate lithography stage and so an unconventional approach was chosen whereby the load devices were designed with 0.20 micron gate lengths in order to operate in a light punch–through mode. Load device gate widths of 10 and 5 microns implemented resulting in a total of four different ring oscillator designs. These designs, labelled RO1 to RO4, are identified in table (9–1).

9.3 Inverter Characterisation at d.c.

Functioning inverters were fabricated on the ED–CH3–OX150 wafer. Figure (9–5) shows the voltage transfer characteristic of an inverter as used in the RO2 oscillator design. The active gate length is 0.23 microns and the active load has a gate length and width of 0.20 and 5 microns respectively. The transfer characteristic was measured for a positive supply voltage (V_{DD}) of 3 volts and a substrate voltage (V_b) of zero volts. This characteristic shows a decrease in output voltage as the input voltage (V_{IN}) increases from zero to 1 volt. Over this range the output voltage (V_{OUT}) switches from 1.7 volts to 0.1 volts. However, at low input voltages the output does not reach the supply voltage level and at high input voltages the output does not reach zero volts. Both of these effects can be attributed to punch–through of the load device.

The d.c. transfer characteristics of the inverters require further optimisation. This may be achieved by increasing the arsenic depletion implants and by adjusting the load gate lengths.

9.4 Ring Oscillator Characterisation

The ring oscillator bond pad configuration was designed to permit direct wafer probing using a Cascade 50 ohm waveguide probe–card, which permits probing at frequencies to 18GHz[9.2]. A standard 50 ohm co–axial waveguide was used to connect the probe–card to either a 1 GHz sampling oscilloscope or a 3GHz

frequency meter, each with input impedance of 50 ohms. A variable voltage d.c. power supply was connected as shown in figure (9–6).

Measurements were then taken for each ring oscillator circuit at a range of positive supply voltages (V_{DD}). For each supply voltage, the oscillation frequency (f_{osc}) was measured using the frequency meter and then verified by oscilloscope. The oscilloscope was also used to measure the maximum and minimum voltage of the output waveform.

The switching delays per stage (τ) were determined using the relationship $\tau = 1/38f_{osc}$.

Figure (9–7) shows stage delay (τ) as a function of supply voltage (V_{DD}) for the RO1 and RO2 ring oscillators. Both have active device gate lengths of 0.23 microns and corresponding gate widths of 10 microns and 5 microns. The graph shows that the RO1 circuit with a wider load has a shorter stage delay for a given supply voltage (V_{DD}). The minimum stage delay for this circuit was measured to be 80 psec, whilst for the RO2 circuit the minimum delay was 100 psec.

Figure (9–8) shows a graph of power consumption per stage as a function of stage delay (τ) for the RO1 and RO2 circuits described above. It can be seen that the faster RO1 circuit also consumes less power for a given stage delay. The power–delay product of an inverter is a useful figure of merit and is shown in figure (9–9) as a function of stage delay for the RO1 and RO2 circuits. The RO1 circuit has a power–delay product of 210 fJ/stage at the maximum operating speed, whilst the RO2 circuit has a slightly higher power–delay product of 250 fJ/stage.

Figure (9–10) shows a plot of the maximum and minimum output voltage as measured by the sampling oscilloscope for the 80 psec RO1 circuit. The relatively small maximum output voltage can be attributed to two factors. Firstly, the inverters have been demonstrated to produce a reduced output voltage range at d.c. and secondly the output stage is loaded by the 50 ohm input impedance of the test equipment. This output loading is significant since the transconductance of the output device is approximately 3.5mS.

The minimum output voltage from the oscillator circuit is also shown in figure (9–10). It can be seen that the minimum voltage approaches the maximum value as the positive supply voltage is

increased.

Similar graphs were obtained for the RO3 and RO4 circuits which have active device gate lengths of 0.39 microns. These circuits achieved minimum stage delays of 150 psec and 210 psec for load widths of 10 and 5 microns respectively. The corresponding power–delay products were 400 fJ and 440 fJ.

The circuits with 0.23 micron active device gate lengths are faster and have lower power–delay products than circuits with 0.39 micron gate lengths.

The gate capacitance of an inverter stage with a 0.23 micron device was calculated to be approximately 5 fF (ignoring fringing effects). The depletion capacitance of a large area ($1200 \mu m^2$) source/drain drift region was measured using a Hewlett Packard HP4275A LCR meter, and found to decrease from 600 fF to 300 fF as the applied d.c voltage was increased from zero to 2 volts. The larger value corresponds to an inverter stage output capacitance of approximately 50 fF which is significantly higher than the calculated input gate capacitance. Consequently the maximum switching speed of the ring oscillator circuits will be determined by the loading effect of the output capacitance of the n^{th} stage rather than by the input capacitance of the $(n+1)^{th}$ stage.

It is extremely likely that improved stage delays could be obtained by reducing the inverter output capacitances and by improving the active loads to operate in the conventional depletion mode.

Several n–MOS ring oscillator results have been reported for devices with sub micron channel lengths. Lepselter[9.3] reported an enhancement/depletion mode ring oscillator with 0.3 micron channel lengths which achieved a minimum stage delay of 30 psec. Kobayashi et al.[9.4] have reported an enhancement/enhancement mode ring oscillator with similar 0.3 micron channel lengths which achieved a minimum stage delay of 50 psec. Fitchner et al.[9.5] have demonstrated a fully scaled 0.75 micron n–MOS technology with a minimum stage delay of 50 psec for an enhancement/depletion mode ring oscillator.

9.5 Discussion

Several conclusions can be drawn from the results presented in this chapter.

Firstly, the study has progressed from the stage of discrete device fabrication and d.c characterisation to the implementation of ring oscillator circuits consisting of 44 MOSFETs for measuring device high speed performance. These functioning circuits have been obtained with the preliminary wafer batch from a process involving many previously untested techniques.

Secondly, the circuits have demonstrated minimum stage delays of only 80 psec. However, due to the non-scaling of contacting structures, the parasitic output capacitance of the inverter stages is relatively high and consequently the optimum switching performance of the submicron devices has not been measured. In the absence of all parasitic impedences, this limit would be determined by the carrier transit time across the channel region. A stage delay of only 5 psec is predicted for a device with a 0.25 micron channel length and a carrier saturation velocity of $5x10^6$cm/s.

Finally, to obtain optimum switching delays for submicron n-MOS circuits, it is likely that scaling the whole device structure will be necessary in order to minimise parasitic impedances. Consequently these results represent significant progress towards the implementation of ultra fast ring oscillator circuits using 0.25 micron design rules.

References

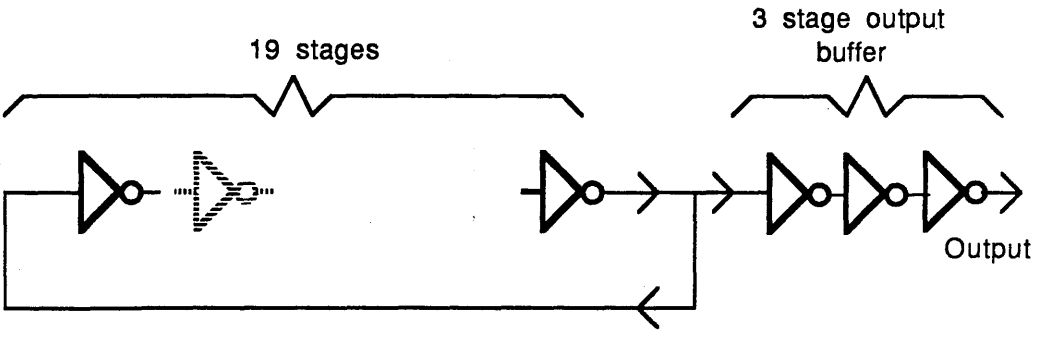
9.1 S.S. Liu, C-H Fu, G.E. Atwood, J. Langston, E. Hazani, H. Dun, I. Beinglass, S. Sachdev and K. Fuchs, 'HMOS III technology,' IEEE Solid-Sate Conf. Dig. Tech. Papers, p234, 1982.

9.2 Supplied by Cascade Microtech inc., Beaverton OR, USA.

9.3 M.P. Lepselter, 'Scaling the micron barrier with X-rays,' in IEDM Tech. Dig. p42, 1981.

9.4 T. Kobayashi, S. Horiguchi and K. Kiuchi, 'Deep– submicron MOSFET characteristics with 5nm gate oxide,' IEDM Tech. Dig., p414, San Francisco, 1984.

9.5 W. Fichtner, E.A. Hofstatter, R.K. Watts, R.J. Bayruns, P.F. Bechtold, R.L. Johnson and D.M. Boulin, 'A submicron nMOS technology suitable for low power high speed circuits,' IEDM Tech. Dig., p264, Washington, 1985.



Figure(9-1) Circuit diagram showing a 19 stage unloaded ring oscillator circuit.

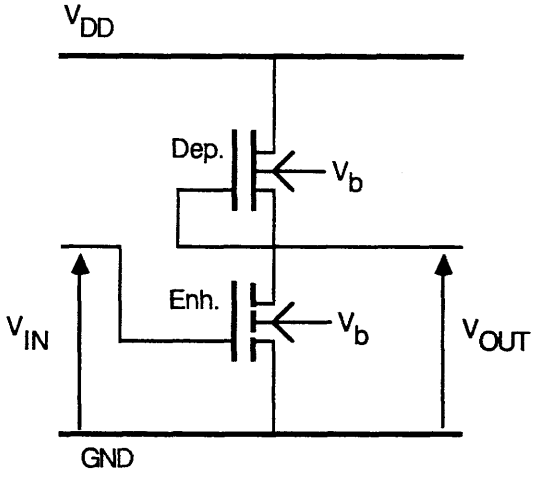


Figure (9-2) Figure showing n-MOS inverter incorporating enhancement and depletion mode devices.

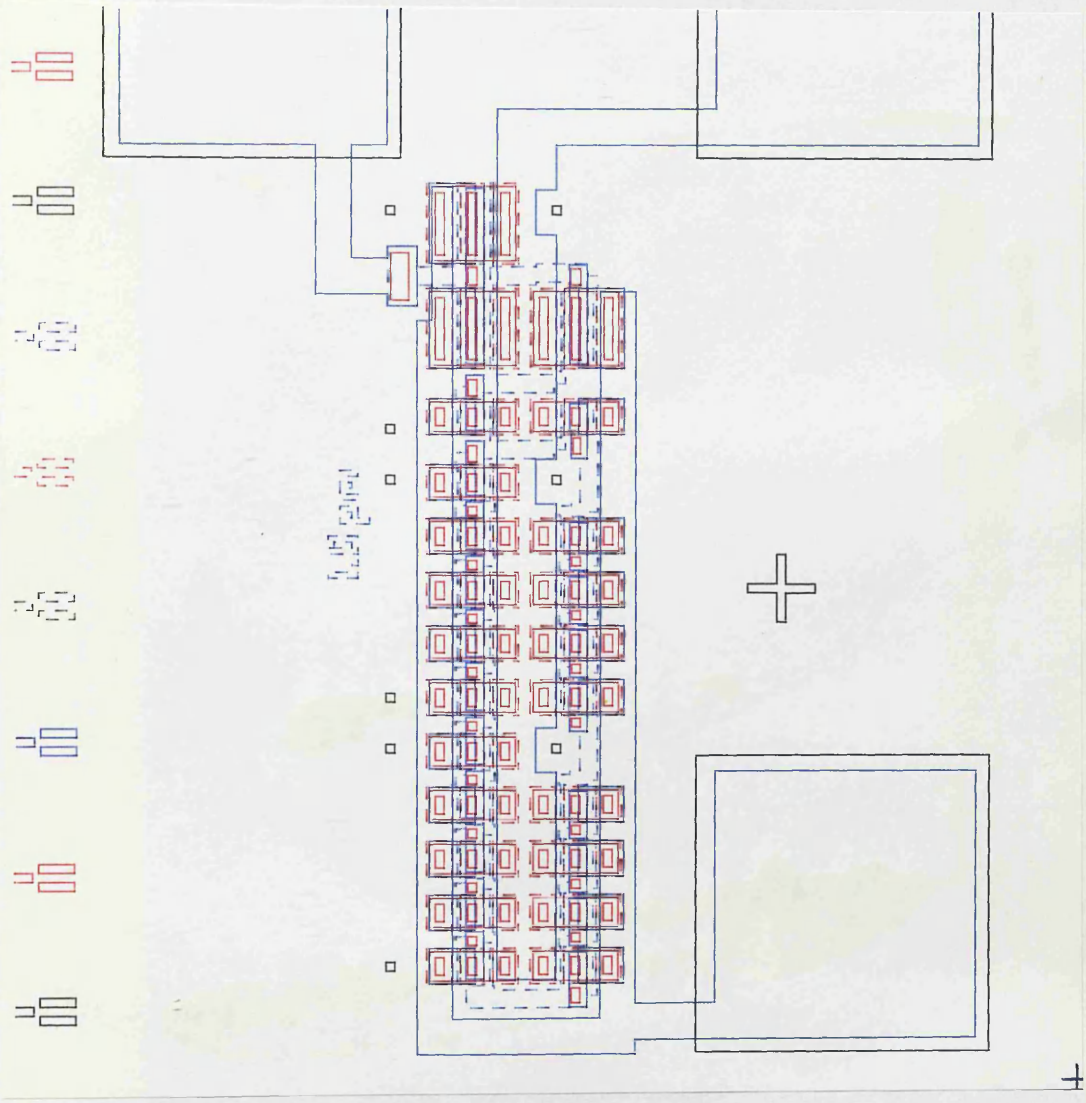


Figure (9-3) EU567 mask layout of the 19 stage ring oscillator circuit.

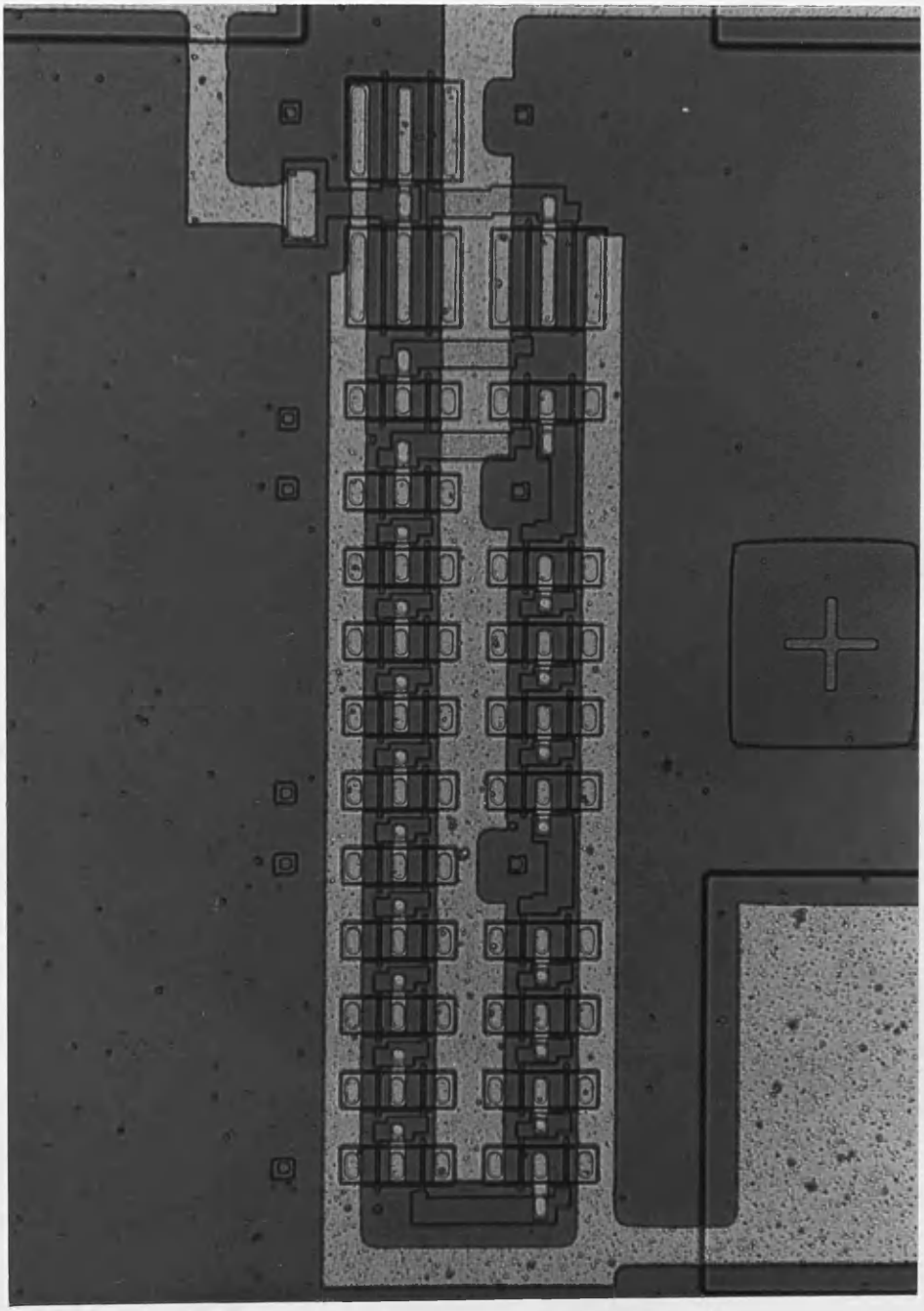


Figure (9-4) Optical micrograph showing a fabricated ring oscillator circuit.

Load devices	Active devices	
	L_g=0.23μm W_g= 10μm	L_g=0.39μm W_g= 10μm
L_g=0.20μm W_g= 10μm	RO1	RO3
L_g=0.20μm W_g= 5μm	RO2	RO4

Table (9-1) Table showing the gate dimensions for the fabricated ring oscillators.

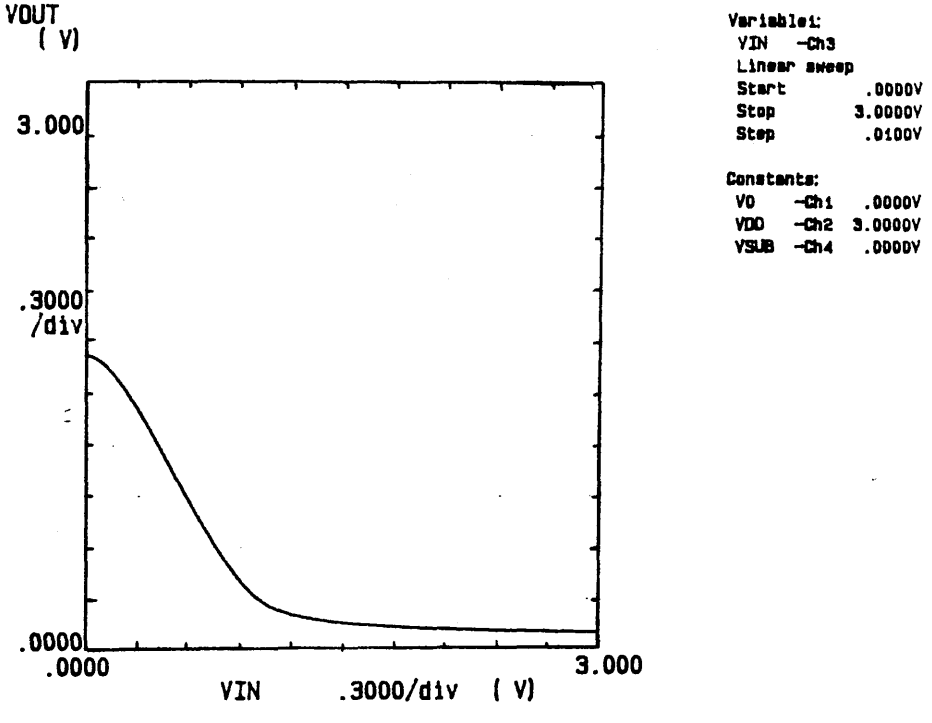


Figure (9-5) Voltage transfer characteristic for a typical inverter as used for the RO2 ring oscillator design.

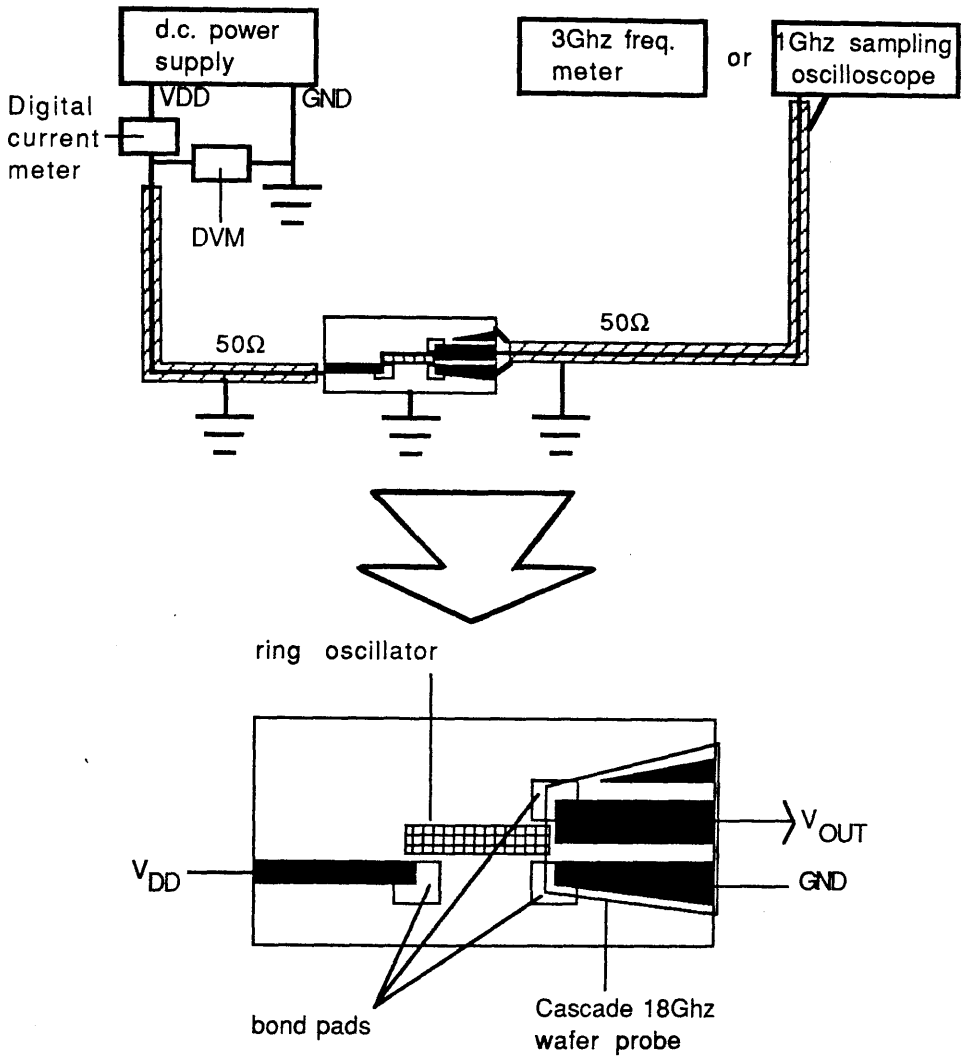


Figure (9-6) Equipment configuration used for testing the ring oscillators.

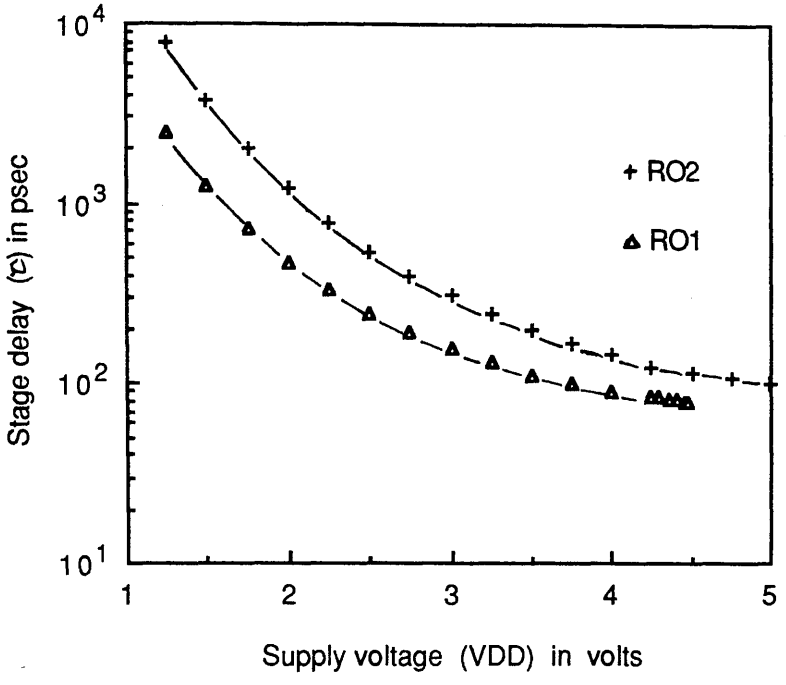


Figure (9-7) Plot of stage delay (τ) versus positive supply voltage (V_{DD}) for the RO1 and RO2 ring oscillators.

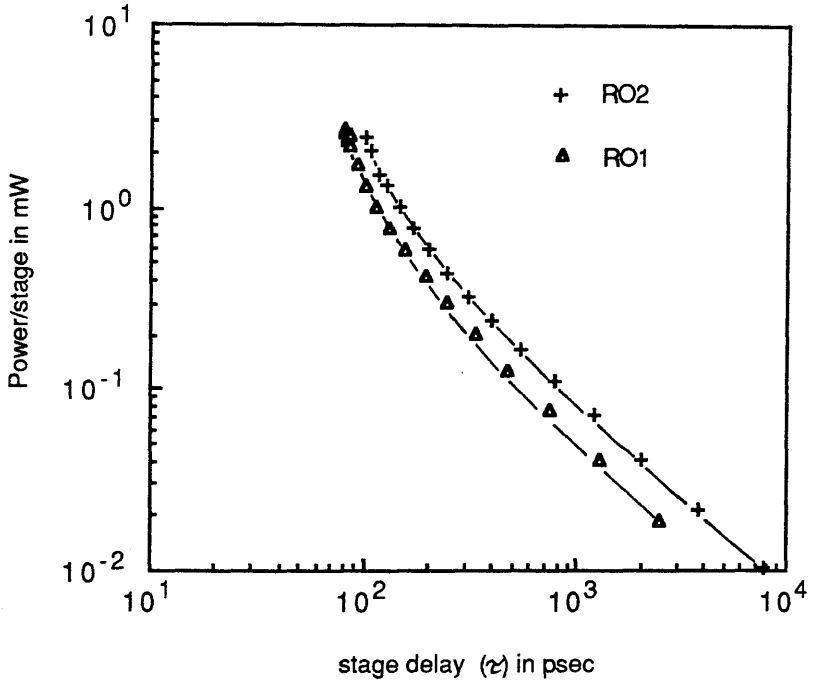


Figure (9-8) Plots of power consumption per stage as a function of stage delay (τ) for the RO1 and RO2 ring oscillators.

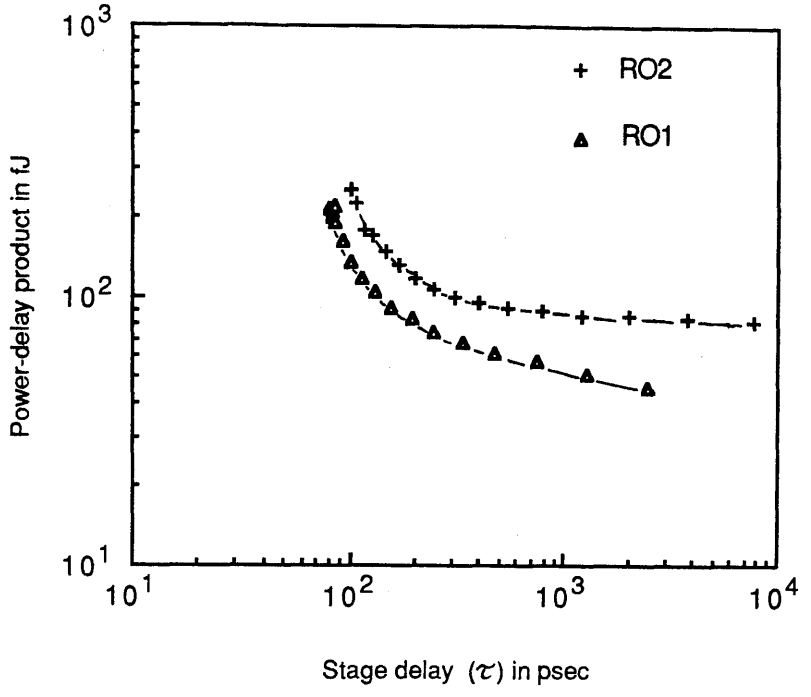


Figure (9-9) Plots of power-delay product as a function of stage delay (τ) for the RO1 and RO2 ring oscillators.

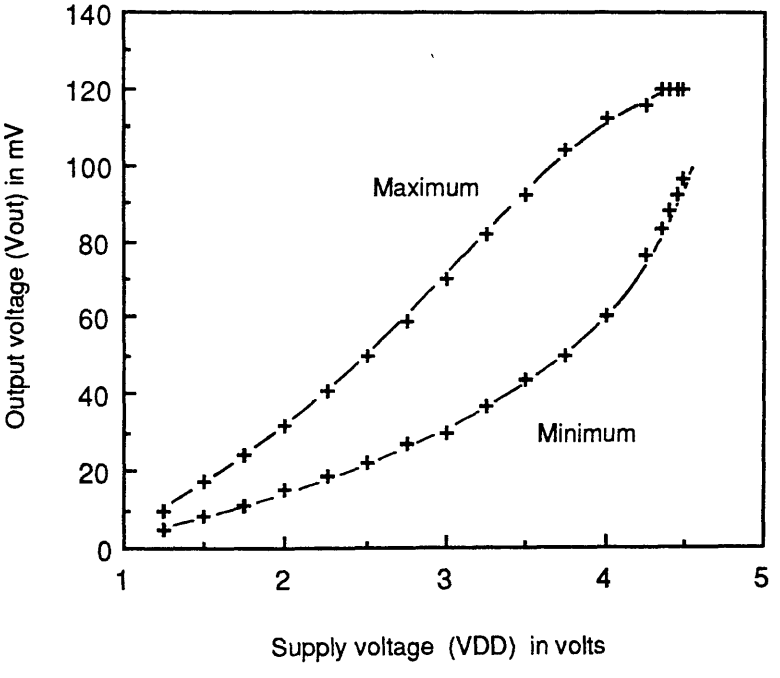


Figure (9-10) Plots of maximum and minimum output voltage as a function of supply voltage (V_{DD}) for the RO1 ring oscillator.

Chapter 10

Conclusion

This research study has demonstrated the fabrication of n–channel silicon MOSFETs with 0.5 to 0.1 micron polysilicon gates. The gates were patterned using high resolution electron beam lithography and dry etching techniques. A wide range of devices have been fabricated and their electrical performance evaluated.

HRN, a negative electron beam resist, has been characterised and demonstrated to have a resolution of 0.1 microns. The resist exhibits excellent properties as a dry etch mask and has been used in conjunction with a chlorine plasma etching process to pattern 0.1 micron polysilicon gate electrodes.

The gate patterning process has been incorporated into a full MOSFET fabrication sequence and functioning devices have been obtained. Devices were implemented with channel doping levels in the range from $3x10^{17}$ to $1.2x10^{18}$ atoms/cm^3 and gate oxide thicknesses down to 150 angstroms. Shallow source/drain drift regions were formed by self–aligned ion implantation and Heatpulse rapid thermal annealing.

Electrical measurements indicated that performance of the devices is limited by the effects of carrier velocity saturation and parasitic source/drain resistance. In addition, smaller devices exhibited a degraded performance due to the onset of short channel effects. These short channel effects are well understood from studies at larger dimensions and lead to a threshold voltage dependence upon both gate length (L_g) and applied drain voltage (V_d). The device output conductance (g_d) is also degraded.

The study has demonstrated the suppression of short channel effects in physically small devices when channel doping levels are increased. Devices with a gate length of 0.25 microns, a channel doping level of approximately $6x10^{17}$ atoms/cm^3 and a gate oxide thickness of 150 angstroms exhibited excellent long channel behaviour. A transconductance of 80 mS/mm was measured with a corresponding open circuit voltage gain of 13. With gate oxide thickness reduced to approximately 50 angstroms these devices appear highly promising for

a future VLSI technology with 0.25 micron design rules. The reduced oxide thickness will decrease threshold voltage and also improve transconductance and voltage gain.

Devices with a gate length of 0.11 microns, a channel doping level of $1.2x10^{18}$ atoms/cm^3 and a gate oxide thickness of 150 angstroms have been implemented. These devices exhibited a transconductance of 70mS/mm, but due to short channel effects, the voltage gain was low.

Lateral encroachment of the drift regions under the gate ($\triangle L$) was calculated from electrical measurements to be approximately 0.06 microns for these devices, indicating minimal dopant redistribution during the Heatpulse anneal step.

The performance of the devices compares favourably with other short gate MOSFETs reported recently. Binnie[10.1] and Howard et al.[10.2] have reported devices with minimum gate lengths of 0.1 and 0.07 microns. In each case lift-off was used to pattern aluminium gate electrodes. Fitchner et al.[10.3], Kobayashi et al.[10.4] and Lee et al.[10.5] have reported devices with dry etched polysilicon gates and minimum gate lengths in the range 0.25-0.2 microns.

A second dry etching process has been developed at Glasgow for patterning 0.1 micron polysilicon gate electrodes. The new process employs silicon tetra-chloride reactive ion etching and has been used to fabricate a further set of sub micron MOSFETs.

The study has progressed from the fabrication of discrete devices to the point at which functioning circuits have been demonstrated. Ring oscillator circuits were chosen as an effective method for establishing the high speed performance which may be achieved through device scaling. These circuits have been implemented using a new mask set and a slightly modified fabrication process. Preliminary results have been reported for ring oscillators with active device gate lengths of 0.23 microns. These circuits have achieved minimum stage delays of only 80 psec and power-delay products of 210 fJ. It is anticipated that a considerably improved performance will be obtained by optimising the load devices and reducing the parasitic substrate capacitances.

Several directions for future research can be identified.

Firstly, the limited scaling approach using electron beam and optical lithography may be used further to investigate the effects of

reduced source/drain junction depths, thinner gate oxides and silicide contacting structures as required to further optimise device performance.

Secondly, a fully scaled experimental n–MOS fabrication process with 0.25 micron design rules can be attempted. This will require patterning at all levels by electron beam lithography in order to scale active areas, device isolation widths, contact windows and interconnect linewidths. The resulting 0.25 micron circuits will permit full performance benefits through scaling to be identified.

Thirdly, the smallest devices which have been fabricated may be used to investigate velocity over–shoot effects as reported by Chou et al.[10.6] for a device with a 0.075 micron channel length.

Finally, the high resolution lithography and polysilicon dry etching techniques which have been developed for this study, may be applied to the fabrication of novel silicon devices such as the recently proposed Thermal Emission Transistor (TET)[10.7]. Fabrication of this device will require both high resolution lithography and silicon M.B.E.[10.8].

References

10.1 C.E. Binnie, 'The fabrication of small geometry MOSFETs using electron beam lithography, 'Ph.D. Thesis, University of Glasgow, Feb 1985.

10.2 R.E. Howard, L.D. Jackel, R.G. Swartz, P. Grabbe, V.D. Archer, R.W. Epworth, E.L. Hu, D.M. Tennant and A.M. Voshchenkov, 'Buried channel MOSFETs with gate lengths from 0.25um to 700Å,' IEEE Electron Device Lett., vol EDL–3, p332, 1982.

10.3 W. Fichtner, R.K. Watts, D.B. Fraser, R.L. Johnson and S.M. Sze, '0.15 um Channel–length MOSFETs, fabricated using e–beam lithography,' IEEE Electron Device Lett., vol EDL–3, p412, 1982.

10.4 T. Kobayashi, Seiji Horiguchi and K. Kiuchi, 'Deep–Submicron MOSFET Characteristics with 5 nm Gate Oxide,' IEEE Tech. Dig. Electron Device Meet., San Francisco, p414, 1984.

10.5 H–S Lee and L.C. Puzio, 'The electrical properties of subquarter–micrometer gate–length MOSFET's,' IEEE Electron Device Lett., vol EDL–7, 1986.

10.6 S.Y. Chou, D.A. Antoniadis and H.I. Smith, 'Observation of Electron Velocity Overshoot in sub–100–nm–channel MOSFET's in Silicon, ' IEEE Electron Device Letters, vol EDL–6, p665, 1985.

10.7 R.F. Kazarinov and Serge Luryi, 'Majority Carrier Transistor Based on Voltage–Controlled Thermionic Emission,' Appl. Phys. A., vol 28, p151, 1982.

10.8 J.C. Bean, 'Silicon MBE,' J. Vac. Sci. Technol, vol 18, p769, 1981.

Appendix – The EU567 mask set design (x400)

GLASGOW
UNIVERSITY
LIBRARY


1.0 μm BIAS ⇨