

MOSFET Characterisation and its Application to

Process Control and VLSI Circuit Design

Anthony Gribben

Doctor of Philosophy

University of Edinburgh

1988



Abstract

As the silicon fabrication industry has rapidly expanded, competition has led to smaller geometry circuits in order to maximise profit and obtain optimum performance. Device operation has to be characterised more rigorously because the tolerances on device operation are reduced and designers are constantly endeavouring to push the limits of the technology. In order to characterise MOSFETs, parameters for the SPICE level 3 model can be extracted. Although SPICE has been around for several years, commercial programs which extract parameters using numerical optimisation have only recently become available. A program, PARAMEX, has been developed to physically extract parameters which accurately simulate device operation. A thorough analysis of parameters for different geometry devices has been carried out and recommendations for simulating devices of different sizes are provided. Of particular interest to designers is the definition of a 'worst case' parameter set and by extracting parameters from numerous sites on a single wafer, a method for determining a 'worst case' set is proposed. Ideally if SPICE parameters are to be central to the design of integrated circuits, it would be useful to link them to specific steps in fabrication. Parameters from wafers fabricated using different processes were correlated with the process steps which had been varied and the effects on both first and second order parameters are described. The subthreshold region is of increasing importance in small geometry circuits. As fabrication processes have evolved, more implants have been made in the channel region with only limited regard to the effect on the subthreshold currents. By thoroughly analysing the subthreshold currents in transistors manufactured with different channel profiles, conclusions about the effect of channel implants on subthreshold operation and the consequences for simple circuits are set out.

Acknowledgements

I would like to thank Professor J.M. Robertson for arranging this project with Motorola Ltd. and SERC, for overseeing the work and for reading the script. I would also like to extend my gratitude to Dr. A.J. Walton for the large amount of supervision he has given me during the research.

I am grateful to the staff of the EMF for fabricating the samples which were used in the experiments.

Finally the financial support from SERC and Motorola Ltd., and the measuring equipment supplied by Hewlett-Packard Ltd were essential in carrying out the project and I greatly appreciate their assistance.

Tony Gribben

February 1988

Declaration

I declare that the contents of this thesis are my own work. Wherever another author has been quoted, this is stated and a reference is provided.

Papers Associated with this Thesis

1. A. Gribben, A.J. Walton, and J.M. Robertson, "Accurate Physical Parameter Extraction for Small Geometry Devices," *Semiconductor International Conference Proceedings*, pp. 186-202, Birmingham, 1986.
2. A. Gribben, A.J. Walton, and J.M. Robertson, "Parametric Testing to Link Design and Fabrication," *IEE Colloquium on Testing and Inspection of Electronic Components and Circuits*, pp. 3/1-3/3, London, 1987.
3. P. Tuohy, A. Gribben, A.J. Walton, and J.M. Robertson, "Realistic Worst-case Parameters for Circuit Simulation," *IEE Proceedings*, vol. 134 Pt. 1, no. 5, pp. 137-140, October 1987.
4. A. Gribben and A.J. Walton, "A Review of Parametric Testing," *Semiconductor International Conference Proceedings*, pp. 39-63, Birmingham, 1987.

Glossary

β	$q kT^{-1}$	V^{-1}
<i>Beta</i>	gain of MOSFET	$A V^{-2}$
C_d	depletion capacitance per unit area	$F m^{-2}$
C_{ox}	oxide capacitance per unit area	$F m^{-2}$
δ	width effect on threshold	
Δ_w	channel width reduction	m
<i>Dep</i>	enhancement or depletion (0/1)	
E_c	conduction band energy	eV
E_f	Fermi energy	eV
E_i	intrinsic Fermi energy	eV
E_v	valence band energy	eV
ϵ	permittivity	$F m^{-1}$
ϵ_{ox}	permittivity of silicon dioxide	$F m^{-1}$
ϵ_{si}	permittivity of silicon	$F m^{-1}$
η	drain feedback coefficient	
E	electric field	$V m^{-1}$
E_x	electric field in x-direction	$V m^{-1}$
E_y	electric field in y-direction	$V m^{-1}$
F_b	geometry dependent body factor	
F_n	narrow channel factor	
F_s	short channel factor	
G	conductance	$A V^{-1}$
γ	substrate bias coefficient	$V^{\frac{1}{2}}$
I_d	drain current	A
J	current density	$A m^{-2}$
k	Boltzmann's constant	$J K^{-1}$
κ	saturation field factor	
L	effective channel length	m
L_B	Debye length	m
L_d	diffusion length	m
L_m	mask channel length	m

Glossary (cont)

L_{\min}	minimum channel length for long channel behaviour	m
μ	carrier mobility	$m^2 V_s^{-1}$
μ_{eff}	effective mobility	$m^2 V_s^{-1}$
μ_o	maximum carrier mobility	$m^2 V_s^{-1}$
μ_s	surface carrier mobility	$m^2 V_s^{-1}$
n	n-carrier concentration	m^{-3}
N_A	no. of acceptors per unit volume	m^{-3}
N_D	no. of donors per unit volume	m^{-3}
N_{fs}	fast surface state density	m^{-2}
n_i	intrinsic carrier concentration	m^{-3}
N_{sub}	substrate doping concentration	m^{-3}
p	p-carrier concentration	m^{-3}
ψ	potential	V
ϕ_b	potential difference between intrinsic and extrinsic Fermi levels	V
ϕ_{ms}	metal-semiconductor work function	V
ψ_s	surface potential	V
ψ_{s0}	surface potential at the source	V
ψ_f	Fermi potential	V
ψ_n	quasi-Fermi level for electrons	V
ψ_p	quasi-Fermi level for holes	V
q	electron charge	C
$Q_b(y)$	depletion charge in channel at y from source	$C m^{-2}$
Q_f	fixed oxide charge	$C m^{-2}$
$Q_i(y)$	inversion charge in channel at y from source	$C m^{-2}$
Q_m	mobile ionic charge	$C m^{-2}$
Q_{ot}	oxide trapped charge	$C m^{-2}$
$Q_t(y)$	total charge in channel at y from source	$C m^{-2}$
ρ	charge density per unit volume	$C m^{-3}$
ρ_t	parameter governing the transition at threshold	
S	subthreshold swing	$mV dec^{-1}$
σ_c	conductivity of channel	$Cm V_s^{-1}$

Glossary (cont)

T	absolute temperature	K
θ	mobility modulation coefficient	V^{-1}
t_{ox}	gate oxide thickness	m
<i>Type</i>	n- or p-channel (1/-1)	
v	velocity	$m\ s^{-1}$
V_b	substrate voltage	V
V_d	drain voltage	V
V_{fb}	flatband voltage	V
V_g	gate voltage	V
V_{geff}	effective gate voltage	V
v_{max}	maximum carrier velocity	$m\ s^{-1}$
V_{to}	threshold voltage ($V_b = 0$)	V
V_{tran}	transition voltage at the onset of saturation	V
w	depletion width	m
W	effective channel width	m
W_m	mask channel width	m
w_d	depletion region around the drain	m
w_s	depletion region around the source	m
X_d	depletion region coefficient	$m\ V^{-\frac{1}{2}}$
x_i	inversion layer width	m
x_j	junction depth	m

Index

Title Page	i
Abstract	ii
Acknowledgements	iii
Publications	iv
Glossary	v
Index	viii
Chapter 1 : Introduction	
1.1 The Development of the MOS Industry	1
1.2 Device Scaling and its Influence on Device Performance	5
1.3 SPICE Parameters to Link Design and Fabrication	10
1.4 The Subthreshold Region	12
Chapter 2 : Device Modelling	
2.1 Introduction to the MOSFET	14
2.2 Fundamental Physics for MOSFET Analysis	16
2.3 Basic MOSFET Model	21
2.4 Advanced MOSFET Model	25
2.5 The SPICE 2 Levels 1 and 3 MOSFET Models	30
2.6 Other MOSFET Models	36
2.7 Numerical Device Modelling : MINIMOS	40
Chapter 3 : MOS Processing	
3.1 The EMF NMOS Process	42
3.2 Lithography	49
3.3 Ion Implantation	52
3.4 Small Geometry Processing	54
3.5 Process Simulation	56

Chapter 4 : Parameter Extraction		
4.1	Different Extraction Philosophies	
4.1.1	Numerical Optimisation or Physical Parameter Extraction	63
4.1.2	Parameter Extraction by Numerical Optimisation : TECAP	64
4.1.3	The Physical Parameter Extraction Program : PARAMEX	71
4.2	SPICE 2 Parameter Extraction	
4.2.1	Threshold Voltage V_{to}	74
4.2.2	Diffusion Length L_d	76
4.2.3	Substrate Bias Coefficient γ	78
4.2.4	Drain Feedback Coefficient η	81
4.2.5	Width Reduction Δ_w	84
4.2.6	Narrow Channel Factor δ	86
4.2.7	Fast State Density N_{fs}	86
4.2.8	Carrier Mobility μ_o and θ	89
4.2.9	Maximum Carrier Velocity v_{max}	90
4.2.10	Saturation Slope Coefficient κ	94
4.3	Simulation of Characteristics	
4.3.1	NMOS Enhancement Results	97
4.3.2	Depletion Device Simulation	106
4.3.3	P-channel Device Simulation	110
4.4	Summary	119
Chapter 5 : The Influence of Device Size, Manufacturing Variations and Process Variations on Parameters		
5.1	Parameters and Device Size	
5.1.1	Introduction	121
5.1.2	Threshold Voltage	121
5.1.3	Substrate Bias Coefficient	125
5.1.4	Mobility Variation with Gate Voltage μ_o and θ	125
5.1.5	Carrier Mobility and Drain Voltage v_{max}	127
5.1.6	Threshold Modulation by Drain Voltage η	132
5.1.7	Saturation Slope Coefficient κ	137
5.1.8	Formulation of Parameter Sets for Different Geometries	139

5.2	Manufacturing Variations and Parameters	
5.2.1	The Importance of Manufacturing Variations	141
5.2.2	Best and Worst Case Parameter Sets	142
5.2.3	Correlation Between Parameters	148
5.2.4	Worst Case Parameters and Circuits	156
5.2.5	Derivation of a Worst Case Parameter Set	173
5.3	Parameters and the Fabrication Process	
5.3.1	Introduction	175
5.3.2	The Experimental Batches	177
5.3.3	Process-Parameter Relationships	180
5.3.4	Conclusions and Future Experiments	187
	Chapter 6 : Subthreshold Operation	
6.1	The Subthreshold Region	190
6.2	Subthreshold Model for Uniformly Doped Devices	191
6.3	Subthreshold Slope	199
6.4	Discussion and Conclusions	217
	Chapter 7 : Conclusions	221
	Bibliography	223
	APPENDICES	
A	SPICE MOSFET Models	229
B	OSIRIS Process Simulation Summary	236
C	PARAMEX Userguide	241
D	Switching Point of CMOS Inverter	260
E	Numerical Differentiation and Inverse Interpolation	262
F	Publications	265

Chapter 1 : Introduction

1.1 The Development of the MOS Industry

There have been numerous significant discoveries in the development of MOS (metal-oxide-semiconductor) transistors. The field effect was first proposed by Heil in Berlin and also by Lilienfeld in New York.^{1,2,3} Shockley and Pearson⁴ carried out experiments demonstrating the modulation of current in a semiconductor and illustrated their theory in an energy band diagram. In 1953 Brown⁵ discovered that a semiconductor surface could be inverted by bringing it into close proximity with a high voltage. This led to use of minority carrier devices and overcame the severe geometrical restrictions to provide an 'off' state in majority carrier devices. Atalla et al⁶ investigated the growth of silicon dioxide as an insulator between the electrode and the gate. They produced silicon dioxide with adequate isolation and with high enough dielectric strength so that the electric field required to electrostatically induce a channel could be obtained at a reasonably low gate voltage. Subsequently Atalla and Kahng fabricated the first MOS transistor in 1960. Improvements were made by Snow et al⁷ who eliminated sodium ions from the oxide layer and by Sarace et al⁸ who developed the self-aligned polysilicon gate which reduced overlap capacitances and wastage of silicon area. A few landmarks in the industry then followed. The first commercial MOS devices were produced in 1962, the first MOS memory was made in 1968 and the first MOS calculator appeared in 1970.

During the 1960's, simple circuit functions were implemented on a single chip. It was realised that integrating more functions on a chip is economically beneficial. The cost of production increases less rapidly than packing density and therefore a net reduction in cost per function results. Speed of operation is improved by larger scale integration and the size of circuit boards is reduced allowing instruments to be more compact. To avoid excessively large chips which result in low yield, device dimensions have to be scaled. This has an economic advantage in that more chips can be put on a wafer and can lead to the extra bonuses of lower power dissipation and faster switching speeds. Wafer sizes have also increased to produce a higher percentage of yielding silicon area. Currently 6 inch wafers represent state of the art for most companies although one IBM plant is being supplied with 10 inch wafers.

Figure 1.1.1 shows how the number of devices per chip has increased since 1960.⁹ By designing more and more complex individual chips, the cost per gate was significantly reduced. Given that there are a certain number of defects on the wafer, as chip size increases the percentage of chips not containing a defect (working die) is reduced (see figure 1.1.2). Therefore the effort to design more functions on a single die has to be coupled with an effort to reduce device dimensions. The reduced device dimensions also lead to improved circuit performance and this is discussed in section 1.2.

Large scale integration, (over 10^3 components per chip), began in the 1970's and progressed into VLSI, (over 10^4 components per chip), in the late 1970's. The greatest number of devices per chip has been attained for memory circuits where the cells are identical. At present, 1 Megabit RAMS are commercially available and 4 and 16 Megabit RAMS are at the pre-production stage. Logic chips are somewhat behind this level of integration.

The fabrication industry has to meet the demand for circuits containing smaller geometry devices. The lithography equipment has to be able to define features of the order of $1\ \mu\text{m}$ or less which have to be etched by anisotropic dry etching. Impurity doses have to be more closely controlled both in quantity and distribution and the redistribution of impurities at high temperature becomes particularly important. The thicknesses of layers grown on the silicon surface have to be more precise and the layer quality must be good e.g. no pinholes in thin oxides and a low concentration of oxide charge. When tolerances similar to those for larger processes are applied to smaller geometry processes, they will produce much larger variations in device characteristics. Typical measurements which are carried out at process parameter check include linewidths, contact resistance and sheet resistance. These are quantities which can be directly related to particular process steps and are used to generally ensure that processing has been successful. The implications of variations in these process parameters on circuit operation are largely unknown and hence only if one of these parameters is a long way out of specification can a non-working circuit be predicted. As the limits of the technology are pushed and the error margins are reduced, it would be extremely useful and probably ultimately essential to have some parameters which quantify device operation. Both process engineers and designers could aim to meet parameter tolerances to ensure functional silicon is produced.

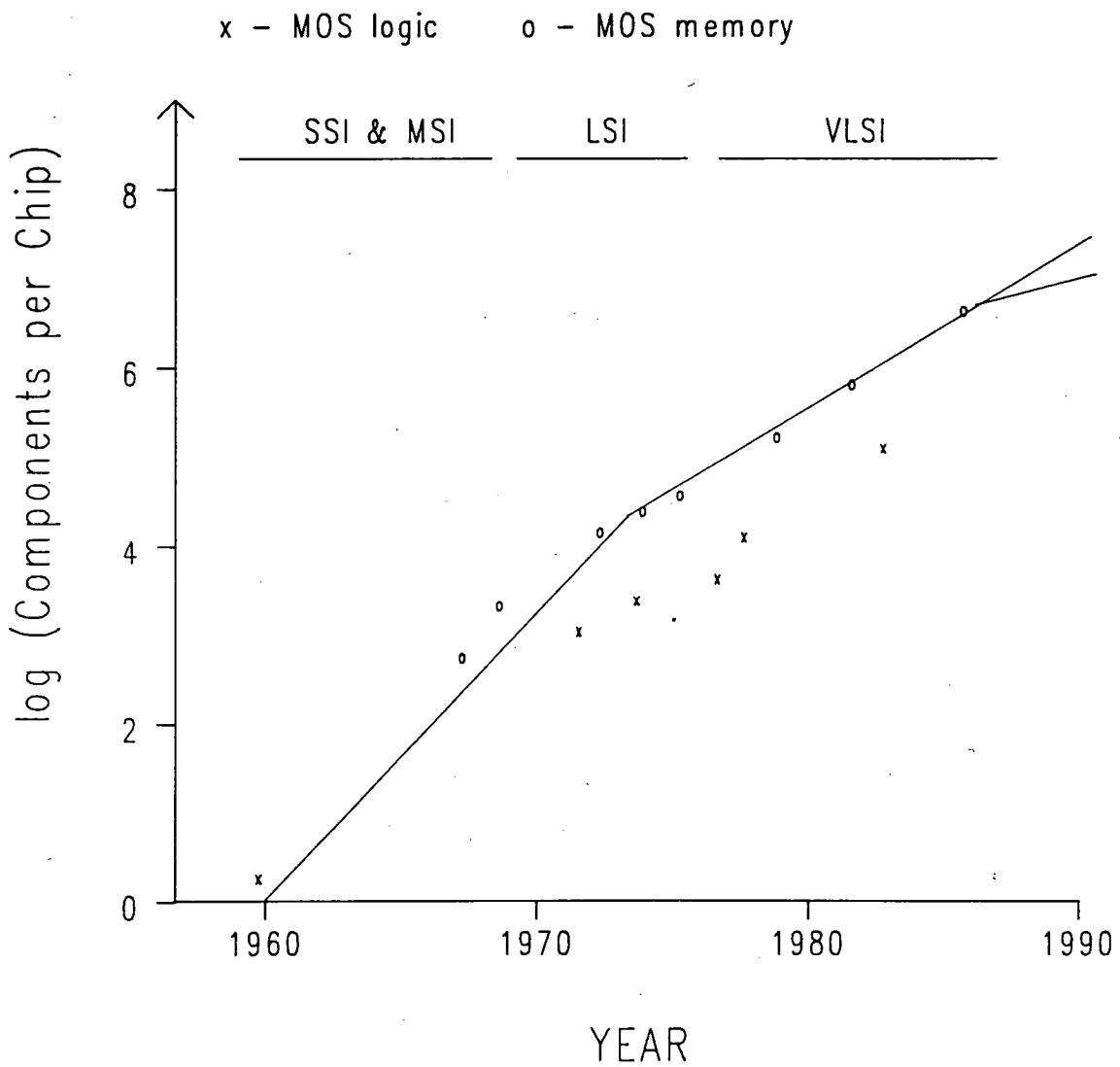
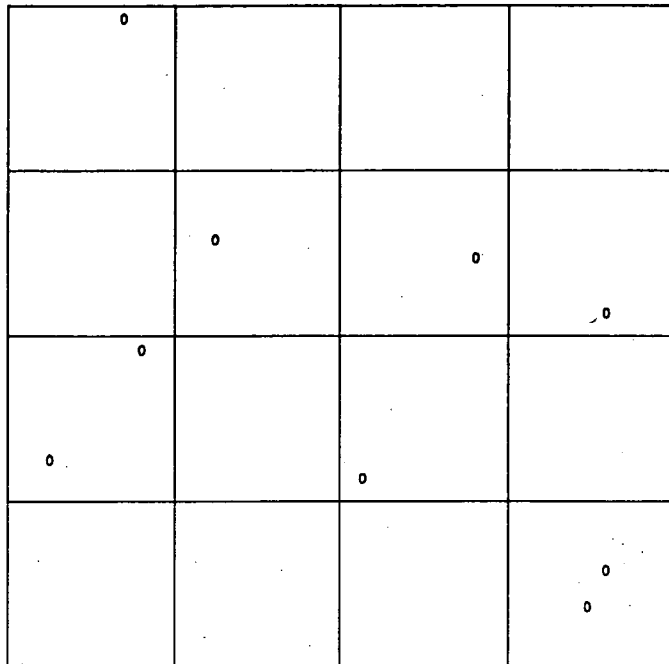
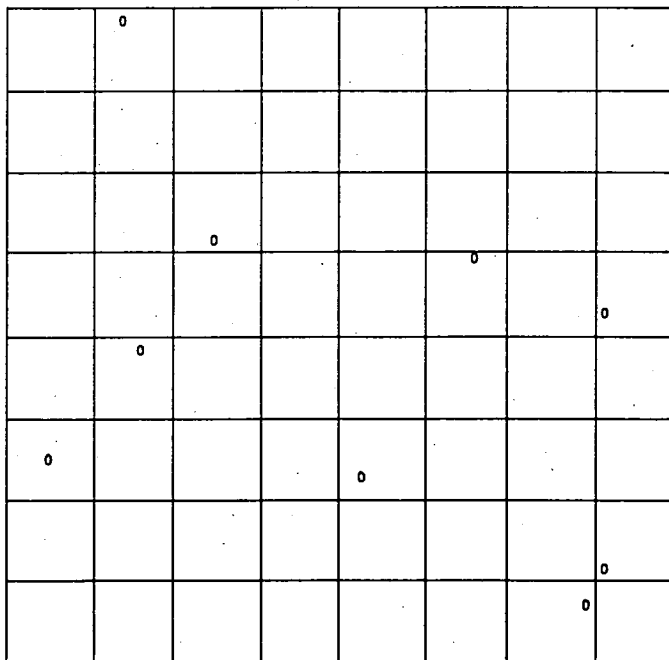


Figure 1.1.1 Exponential Growth in Chip Complexity



Working die = 9

Yield = 56%



Working die = 55

Yield = 86%

Figure 1.1.2 Chip Size and Yield

As chip complexity has increased, the circuit design function has expanded to become a major component of the microelectronics industry and consequently it is now more or less separate from the wafer processing component. In the early days of integrated circuits (SSI and MSI), designers could incorporate redundancy in their circuits in order to allow for very large variations in processing. Faults could be detected easily because of the low level of complexity and two or three iterations of silicon were normal. However with the transition to VLSI and particularly the introduction of ASICs (Application Specific Integrated Circuits), new constraints have had to be imposed. The high complexity means that design faults are more difficult to trace and since silicon area is at a premium only minimal redundancy is permissible. Due to cost and industrial competition, more than one design cycle is intolerable. For designers to design successfully, under these conditions and separate from the fabrication process, it is essential that the semiconductor devices which they use are accurately characterised. The parameters have not only to accurately predict the behaviour of a typical device but the deviations of these parameters over time and geometry will also be required. In order to control these parameters which are precisely linked to circuit operation, the links they have with particular steps in fabrication need to be examined.

The situation outlined in the preceding paragraphs highlights the requirement for thorough device characterisation in order to obtain accurate circuit simulation.

1.2 Device Scaling and its Influence on Device Performance

In section 1.1 it was stated that in order to achieve the desired increase in circuit complexity, device dimensions have to be reduced⁹ so that large chip sizes are avoided. (see figure 1.2.1) Today research is advanced in $0.5\mu\text{m}$ processing and commercial products are functioning using devices with features of just over $1\mu\text{m}$. The bulk of MOS production is at the 1 to $2\mu\text{m}$ level for memories but still at the 3 to $5\mu\text{m}$ level for microprocessors. The reduction of device dimensions, whilst improving device performance, imposes stricter requirements on the process. There is less of a margin for error in mask alignment, impurity doping and layer thickness since the same magnitude of errors in smaller geometry devices will have a much more significant effect on device characteristics.

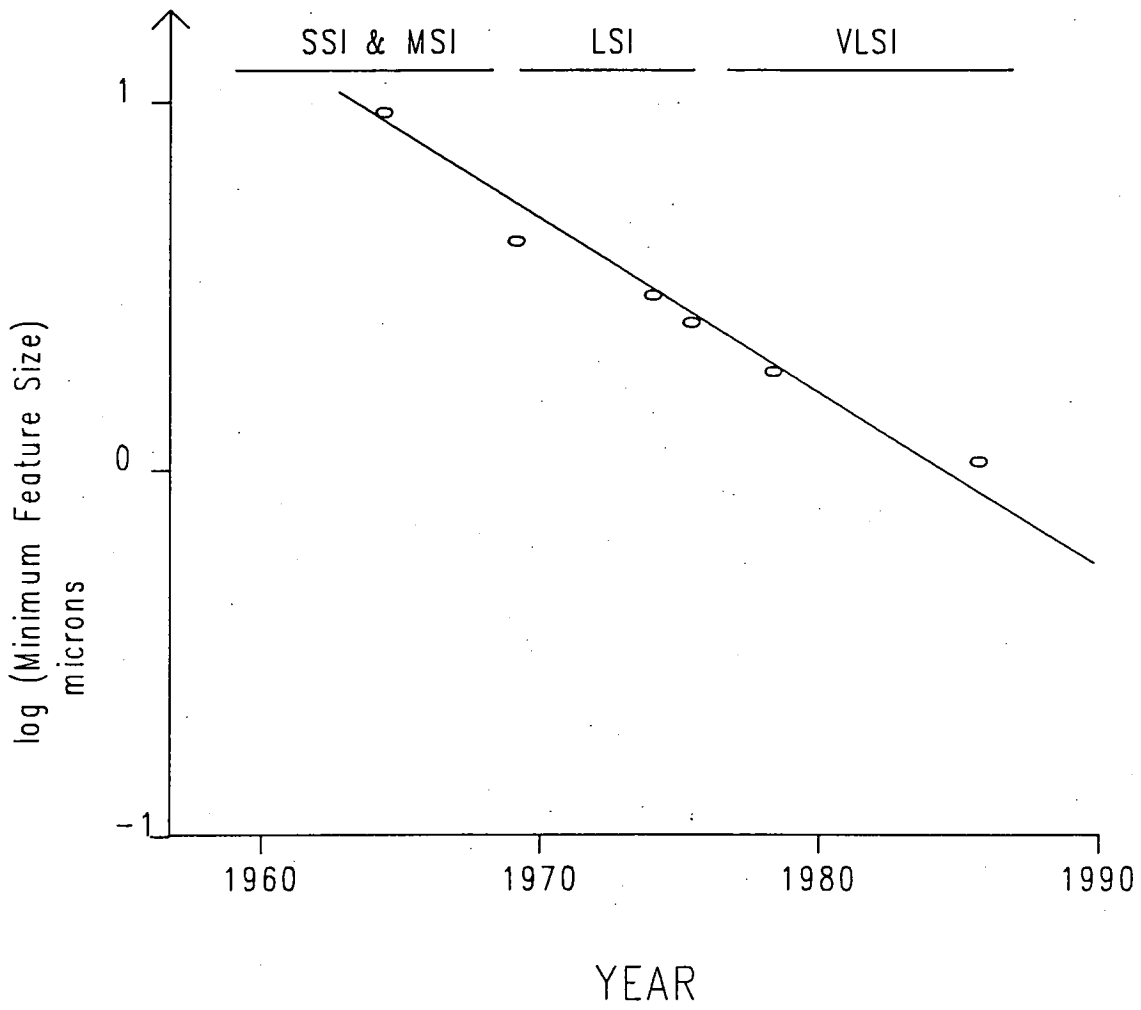


Figure 1.2.1 Gradual Reduction in Minimum Feature Size

Two ideal forms of scaling have been proposed: constant voltage and constant field^{10,11,12,13} and the effects of each of these on device operation are discussed below. The scale factors for each of these methods are set out in table 1.2.1. Constant voltage scaling involves scaling all the dimensions within the circuit but keeping the supply voltage and device threshold voltages constant. This has the advantages of allowing the circuit to interface easily with circuits built at larger geometries and maintains the noise margins around the switching levels of the devices. This method of scaling has several implications on processing. The oxide thickness has been reduced by the scale factor K and therefore a 10 Å variation in oxide thickness is going to hold much greater significance. Pinholes have to be avoided in the thin insulator and the growth process has to be more closely controlled. If the depletion regions around the source and drain meet, the device will be destroyed by a mechanism known as punchthrough. To scale the depletion regions by K , the same factor as the channel length, the doping concentration in the channel must be increased by K^2 . This however would drastically increase threshold voltage. To avoid altering threshold voltage, a high dose of impurity is implanted well below the surface and thereby the depletion regions around the source and drain are reduced without significantly changing threshold. High voltage ion implantation is necessary to achieve this and the quantity and position of the impurity must be carefully controlled to be able to attain the desired device operation. Subsequent high temperature processing steps have to be restricted to avoid redistribution of the impurity. The profile affects several aspects of device operation. The deep implant, although it does not have a substantial effect on threshold, does change the variation in performance with substrate bias. In shorter channels, both saturation current and threshold voltage variation with drain bias are strongly dependent upon the channel impurity concentration. The influence of the deep implant on subthreshold current is largely ignored during process development and this is described in Section 1.4. Device drive current increases and coupled with the reduction in gate capacitance leads to a K^2 increase in switching speed. The two major limitations of constant voltage scaling are the high electric fields and the high power density which result. The electric fields increase by K since dimensions are scaled but the supply voltage is not and the power density goes up by K^3 since the power consumption per gate goes up by K and the gate density by K^2 .

The alternative to constant voltage scaling is constant field scaling. Constant field scaling does not maintain either the noise margins or compatibility with

Table 1.2.1 Scale Factors for Common Scaling Methods

Parameter	Constant Voltage	Constant Field
Feature Size	$\frac{1}{K}$	$\frac{1}{K}$
Gate/Field Oxide Thickness	$\frac{1}{K}$	$\frac{1}{K}$
Gate Capacitance	$\frac{1}{K}$	$\frac{1}{K}$
Junction Depth	$\frac{1}{K}$	$\frac{1}{K}$
Substrate Doping Conc.	K^2	K
Poly Sheet Resistance	K	K
Linewidth	$\frac{1}{K}$	$\frac{1}{K}$
Transistor Gain	K	K
Supply Voltage	1	$\frac{1}{K}$
Enhancement/Depletion V_t	1	$\frac{1}{K}$
Maximum Si/Oxide Field	K	1
Current/Gate	K	$\frac{1}{K}$
Gate Delay	$\frac{1}{K^2}$	$\frac{1}{K}$
Relative RC	1	1
Power/Gate	K	$\frac{1}{K^2}$
Power Density	K^3	1
Delay * Power	$\frac{1}{K}$	$\frac{1}{K^3}$
Gate Density	K^2	K^2
Gates for 1W	$\frac{1}{K}$	1

the unscaled process but does offer much better power consumption than constant voltage scaling. With the scaling of threshold voltage the noise margin around the switching point is reduced and consequently the variation in threshold voltages must be smaller. The ability to define small features and to accurately grow a good quality thin layer of oxide is again important. With scaled supply voltages, the channel implant need only be increased by K . Since threshold voltages are to be reduced, a deep channel implant is usually still necessary. The effects on different areas of device operation due to the profile which were described for constant voltage scaling, also apply for constant field scaling. As a consequence of constant field scaling, the drive current decreases by K and combined with the fact that the gate capacitance and switching points are reduced by K , the speed only increases by K . The big improvement is in power per gate which decreases by K^2 due to the corresponding decreases in device current and supply voltage. Packing density considerations lead to the conclusion that power density remains exactly the same as for the unscaled process.

In reality neither constant voltage scaling nor constant field scaling offer the ideal solution. Constant voltage scaling leads to high electric fields and very high power density and constant field scaling suffers from incompatibility with other geometries and from smaller noise margins. Optimum speed is achieved by constant voltage scaling. The process designer must decide which aspect of performance is most important to him. Physical considerations also influence what is scaled. It is difficult to produce extremely shallow ($0.1 \mu\text{m}$) junctions due to impurity diffusion; thin oxides are prone to pinholes and it is difficult to accurately produce the required thickness when growing a thin oxide layer. Variations from long channel behaviour at short channels such as increasing currents in saturation also complicate circuit design. Sze et al¹⁴ proposed adhering to the empirical formula,

$$L_{\min} = A \left[x_j t_{ox} (w_s + w_d)^2 \right]^{\frac{1}{3}} \quad 1.2.1$$

where A is a proportionality constant, in order to preserve long channel subthreshold behaviour. In practice, most process developers do not opt for constant voltage or constant field scaling but taking into account the facts mentioned above come up with some recipe for scaling. The consideration given to the effects on device behaviour and their consequences with regard to circuit operation are often limited.

If the allowable tolerances of the unscaled process are not changed for the

scaled process, there will be a much larger spread in device characteristics. These may lead to the malfunctioning of the circuit or to the circuit running at a slower speed. As devices are scaled it is important to assess how changes in device operation affect circuits.

1.3 SPICE Parameters to Link Design and Fabrication

Circuit design and wafer fabrication in MOS IC Technology have evolved into two essentially separate entities. As feature sizes are reduced and circuits become more complex, the controls on the process become more stringent. Quite frequently, process control consists of a series of visual inspections during fabrication and the measurement of a set of process parameters on a process control chip at the end of the process. These parameters typically include sheet resistances, oxide thicknesses and device threshold voltages providing a general indication of whether the chips will function. However, the link with circuit operation is obscure. In order to link device operation and hence circuit operation with process control, electrical measurement of physical effects in MOS transistors can be made. These measurements can be manipulated to yield parameters which may be used by a simulator such as SPICE (see below) to simulate circuit operation. A schematic diagram of the function of such a link is shown in figure 1.3.1.¹⁵

With the transition to VLSI, engineers were no longer able to breadboard a prototype in order to test a design and circuit simulation became essential. Before 1970 simulation packages were very slow and expensive to run.¹⁶ The second generation of simulators, developed during the 1970's, includes the most widely used SPICE2 program, where SPICE is an acronym for 'Simulation Program with Integrated Circuit Emphasis'. The mathematical models for various semiconductor devices are made to represent the actual devices by measuring a set of parameters on physical devices.

In the past, SPICE parameters have been extracted on an occasional basis or just estimated with no certainty that they accurately represent device operation. These approximate parameters provided general device characteristics allowing circuit simulation to be carried out. Despite the fact that SPICE has been in existence for many years, it is only recently, in the past two years, that commercial software packages for parameter measurement have become available. These packages include TECAP

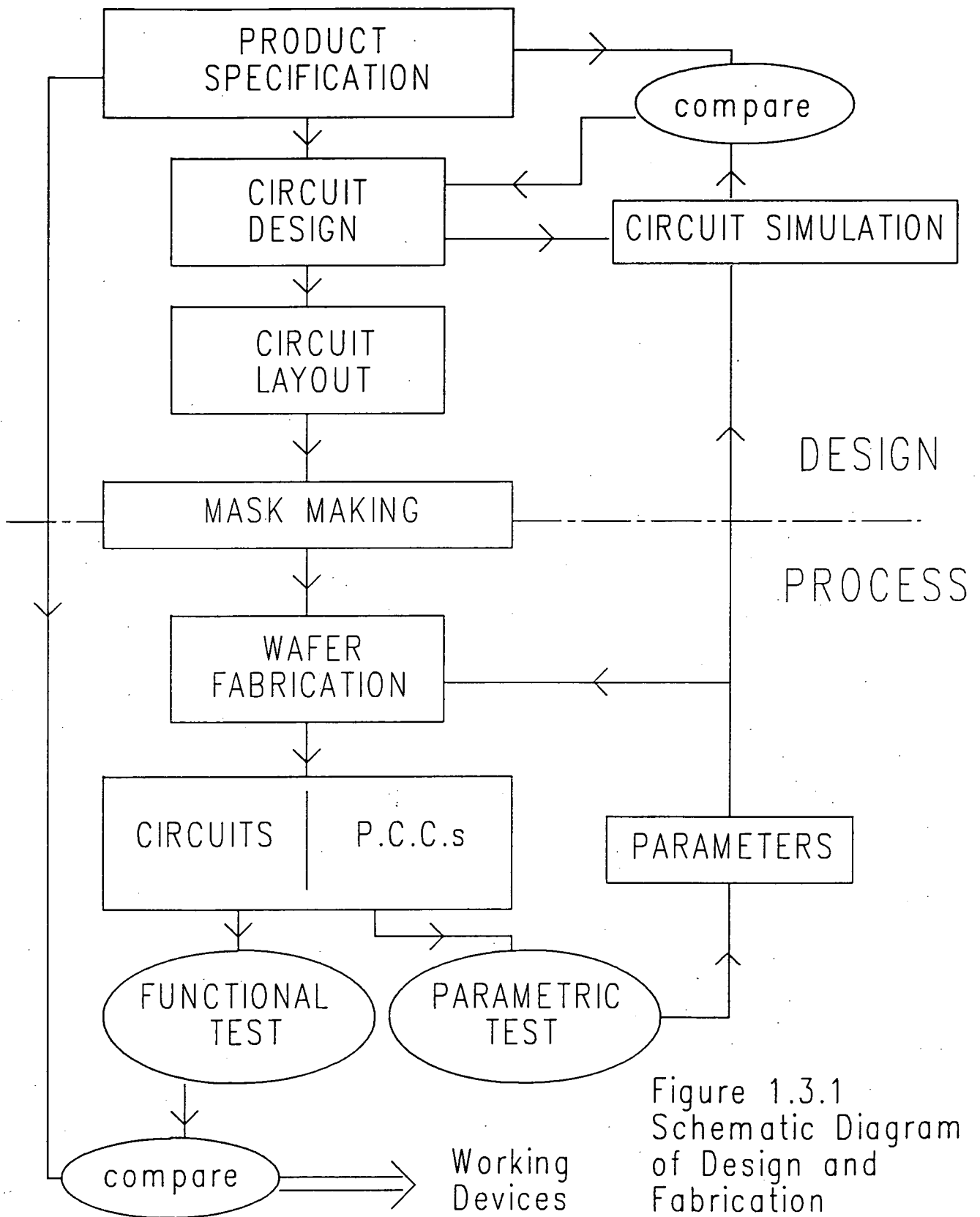


Figure 1.3.1
Schematic Diagram
of Design and
Fabrication

(Transistor Electrical Characterisation and Analysis Program) written by Hewlett-Packard,¹⁷ SUXES written by Stanford University and XEROX and MOSFIT written by MOSAID. The technique behind all these programs is numerical optimisation; the optimum fit between a simulated and a measured transistor characteristic is obtained by altering the values of the parameters. The solution is non-unique and the resulting parameters do not have precise definitions. Referring to figure 1.3.1, the feedback from parametric test will be useful for circuit simulation since accurate representation of the device characteristics can be obtained. However it will be difficult to relate the optimised parameters back to the process. Measuring the electrical effects on devices separately and applying a precise definition to each parameter means that the measured parameters are useful not only for device simulation but also for process control. Relating parameters evaluated in this manner with particular steps in the fabrication process provides an extremely valuable link between chip design and silicon fabrication.

Using parameter-process links, it may be possible to pinpoint a particular process step which is out of specification. Hence, if it is realised that the gate oxide has been grown too thinly and the wafers are still under fabrication, it may be possible to compensate for this by changing a step later in the process. Perhaps a higher implant dose might help. Ultimately, with tight enough process controls and sufficient characterisation, it might be possible to design a circuit to operate with a particular parameter set and automatically generate the process run sheet to produce devices with those parameters. This could be particularly important with ASICs which may need very special performance-process links.

1.4 The Subthreshold Region

Until recently, the subthreshold bias region has been regarded as a cut-off region where the drain to source current is approximately zero. In smaller geometry digital circuits, especially where the supply voltage has also been scaled,¹⁸ the leakage current is of greater significance. The promise of noise and gain improvements in analogue circuits when MOS devices are biased in the subthreshold region¹⁹ has also led to extra interest. When a voltage below the threshold voltage is applied to the gate, the current flowing between the drain and the source is exponentially dependent upon that voltage.²⁰ A typical subthreshold characteristic slope is 100mV/dec. In other words for a 100mV decrease in gate voltage, the drain to source current goes down by a factor

of 10. At small geometries, the capacitive nodes become smaller and can be charged or discharged more easily. If the gate voltage on an 'off' device is 0.2V higher than anticipated then the leakage current could be 100 times higher. It is important that devices turn off quickly below threshold so that the likelihood of a spuriously turned on device is minimised.

As a bias voltage is applied to the substrate,²¹ the depletion region widens and the subthreshold current between drain and source changes. The distribution of the depletion region charge is dependent upon the doping profile in the device. The expansion of the depletion region, on the application of a substrate bias, has the effect of increasing threshold voltage and therefore the magnitude of the subthreshold current is reduced. As the depletion region widens, the subthreshold slope changes as well. The amount by which the slope changes is dependent upon the depletion layer width which in turn is dependent upon the channel profile. Hence there is an important relationship between the subthreshold slope and substrate bias.

In Sections 1.1 and 1.2, the extra impurity implants required when devices are scaled were outlined. Frequently during process development, the effect on subthreshold due to threshold adjust or punchthrough implants are largely ignored. However with the added significance of the subthreshold region, it has become necessary to investigate the effect of these implants on the subthreshold region.

Chapter 2 : Device Modelling

2.1 Introduction to the MOSFET

The MOSFET is the basic component of all MOS integrated circuits. The first requirement of an integrated circuit designer's simulation program is an accurate device model. This chapter reviews the theory behind most device models and describes the advantages and disadvantages of a few models: the SPICE levels 1 and 3 models,²² Brews' charge sheet model,²³ CASMOS²⁴ and Wright's model.²⁵

The physical layout of an MOS transistor is shown in figure 2.1.1. The voltage applied to the gate is used to modulate the current which flows between the drain and source terminals. By definition, the drain is at a higher potential than the source and therefore, without any bias on the gate, the p-n junction between the n^+ drain region and the p substrate is reverse biased and only leakage current is able to flow. When a positive voltage is placed on the gate, holes are repelled from the silicon surface below the insulating oxide and, if the bias is sufficiently high, the surface is inverted and a channel of electrons connects the source and drain.

Essentially there are three regions of operation: subthreshold, linear and saturation. Before the gate voltage is high enough to invert the channel, little current flows and this is the subthreshold region. When an inversion region has been created, current increases linearly with drain voltage and hence this is the linear region. As the drain voltage increases, there comes a point where inversion cannot be sustained at the drain end of the channel. This is the saturation region where drain current is almost constant.

Among the properties which the device exhibits are a very high input impedance because of the insulator between the gate and the channel, and a high 'off' impedance because of the reverse biased p-n junction. The device is also symmetrical; the source and drain can be interchanged without altering device performance.

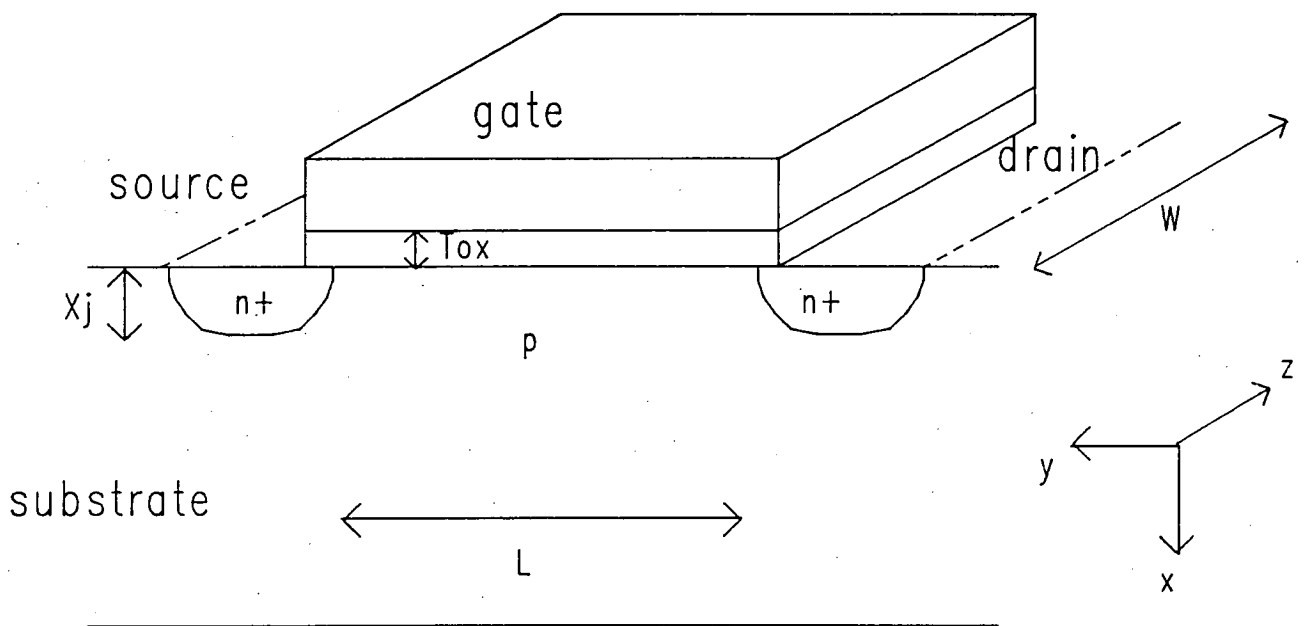


Figure 2.1.1 Structure of a MOSFET

2.2 Fundamental Physics for MOSFET analysis

In order to analyse the operation of the MOS transistor in more detail, it is useful to use energy band diagrams such as in figure 2.2.1.^{26,27} A single electron in a vacuum is defined as having zero potential energy and an electron loses energy as it goes from the conduction band E_c to the valence band E_v . The region between is the forbidden gap for silicon and the average energy of the electrons in the outermost state (the Fermi energy) of undoped silicon E_i is approximately at the midgap. The silicon has been doped with an acceptor, usually boron, which has moved the Fermi level E_f below the intrinsic Fermi level showing that there are more free holes than electrons and the silicon is therefore p-type. The horizontal axis denotes distance into the silicon substrate with the surface on the left. In this case there is no variation of energy as this distance (x) increases. This is a special case called flat-band where the silicon is neutral at all depths because the holes exactly balance the acceptor dopant ions.

In normal MOS devices, flat-band does not occur at zero gate bias because of the metal-semiconductor work-function difference and because of charge both in the oxide and at the surface. The work-function is the energy required to move an electron from the Fermi level to the vacuum energy level. When the metal or degenerately doped n-type polysilicon gate is placed on the opposite side of the insulating oxide to the p-type substrate, electrons are attracted to the silicon surface and so the p-type semiconductor surface is depleted. This is caused by the difference in Fermi levels, which is exactly equal to the difference in work-functions.

The other factor which leads to the energy bands not being flat when there is no gate bias is that there is charge in the oxide and at the surface. The charge is classified into fixed oxide charge, oxide trapped charge and mobile ionic charge. Fixed oxide charge cannot easily be moved from its position 30 Å or so from the silicon-silicon dioxide interface. Its quantity is dependent upon the oxidation conditions and the silicon orientation and it is generally positive. In chapter 1, it was mentioned that Snow et al⁷ were first to realise that alkali ions such as sodium were mainly responsible for the instability of the oxide insulator. These mobile ionic charges can move back and forth through the oxide with bias giving rise to threshold voltage shifts. Again mobile ionic charge is positive and tends to deplete a p-type semiconductor surface. Oxide traps are usually a result of defects in the silicon dioxide and are normally

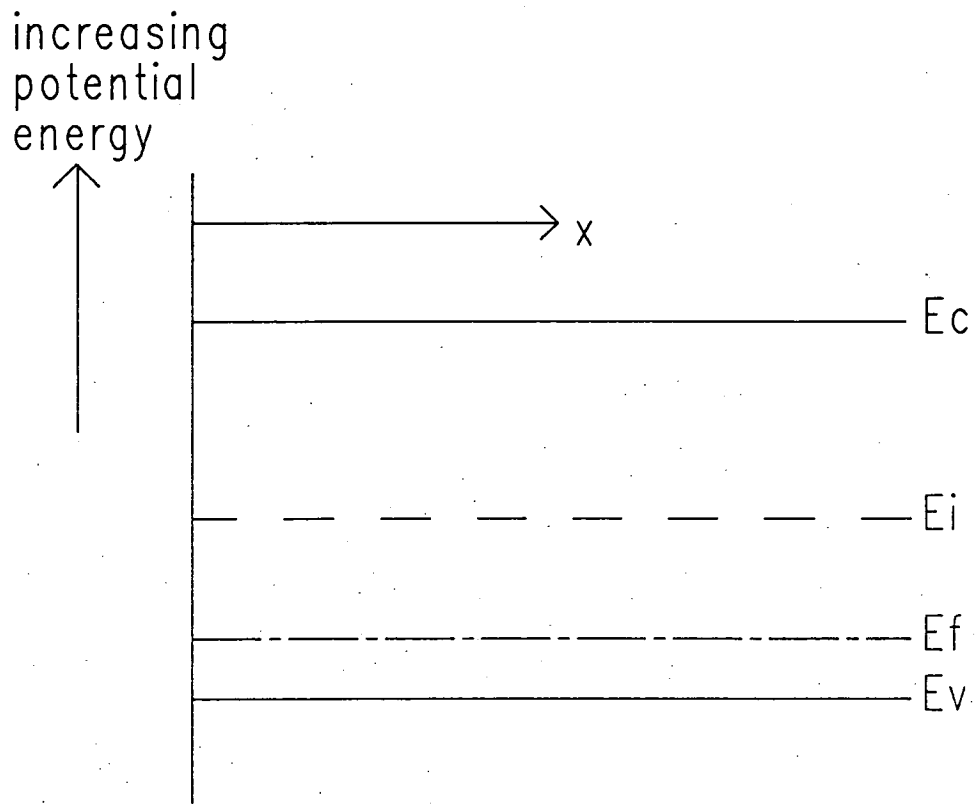


Figure 2.2.1 An example of an energy band diagram

electrically neutral. They are charged by introducing electrons and holes into the oxide.

The combined result of the work-function difference and the oxide charge is a flat-band voltage (the voltage required on the gate to achieve the flat-band situation) given by

$$V_{fb} = \phi_{ms} - \frac{Q_f + Q_m + Q_{ot}}{C_{ox} A_{ox}} \quad 2.2.1$$

where A_{ox} is the area of the oxide capacitor. Flat-band voltage appears in transistor models as an offset in the threshold voltage.

In analysing the MOS transistor, it is easier to look first at the properties of an MOS structure without source and drain.^{28,29} There are three different bias regions of the MOS structure: accumulation, depletion and inversion and the energy band diagrams for each of these conditions are shown in figure 2.2.2. In 2.2.2(a), the applied gate voltage is less than zero with the result that holes are attracted to the surface and accumulate just below the silicon dioxide. This makes the surface more p-type. If the gate voltage is positive then holes are repelled from the surface leaving a state known as depletion just below the silicon-silicon dioxide interface as in figure 2.2.2(b). Increasing this gate voltage further results in electrons becoming more abundant than holes. The surface has been inverted and is now n-type (figure 2.2.2(c)).

One important aspect of the MOS system which has to be expressed analytically is the depletion width. In the analysis below, Poisson's equation is integrated to provide electric field and potential across a depletion layer of width w , assuming that the silicon is uniformly doped with N_{sub} atoms.cm⁻³. An expression for the depletion width, in terms of the doping concentration N_{sub} and the surface potential ψ_s , is found by solving Poisson's equation in one dimension for an MOS structure without either source or drain.^{30,31} The x-direction is defined in figure 2.1.1.

The general form of Poisson's equation in 1-dimension is

$$\frac{d^2\psi}{dx^2} = -\frac{\rho}{\epsilon} \quad 2.2.2$$

where $\rho = -q N_{sub}$ is the charge density per unit volume.

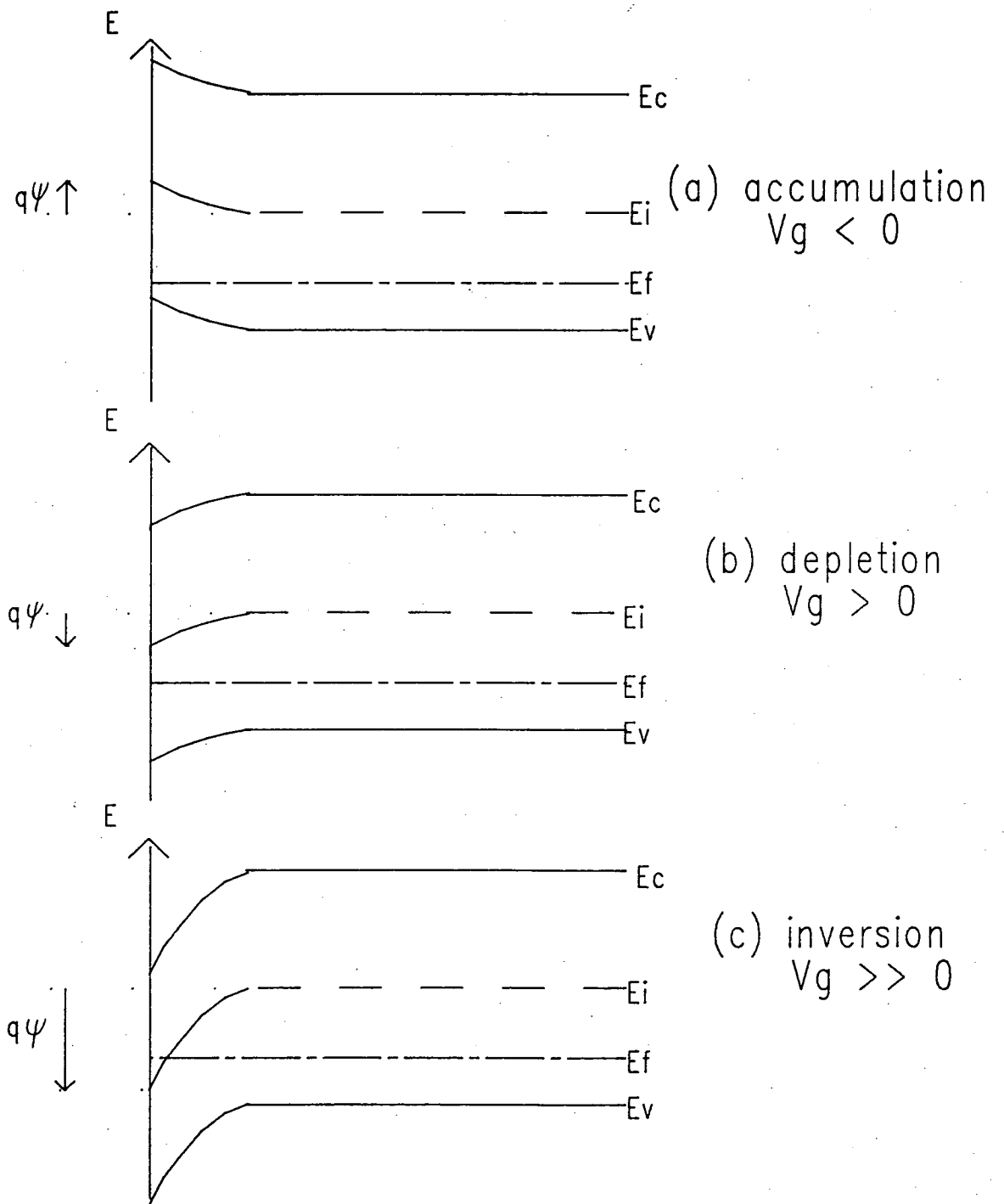


Figure 2.2.2 Three Bias Regions of the MOS Structure

Three boundary conditions will be used:

$$(i) \frac{d\psi}{dx} = 0, x = w$$

$$(ii) \psi = 0, x = w \text{ (by definition)}$$

$$\text{and (iii) } \psi = \psi_s, x = 0.$$

Integrating equation 2.2.2 gives

$$\frac{d\psi}{dx} = \frac{q N_{sub} x}{\epsilon_{si}} + \text{constant} \quad 2.2.3$$

from boundary condition (i)

$$\text{constant} = -\frac{q N_{sub} w}{\epsilon_{si}}$$

Therefore 2.2.3 becomes

$$\frac{d\psi}{dx} = \frac{q N_{sub}}{\epsilon_{si}} (x - w) \quad 2.2.4$$

Integrating 2.2.4 gives

$$\psi(x) = \frac{q N_{sub}}{\epsilon_{si}} \left(-wx + \frac{x^2}{2} + \text{constant} \right) \quad 2.2.5$$

from boundary condition (ii)

$$\text{constant} = \frac{w^2}{2}$$

Therefore 2.2.5 becomes

$$\begin{aligned} \psi(x) &= \frac{q N_{sub}}{\epsilon_{si}} \left(-wx + \frac{x^2}{2} + \frac{w^2}{2} \right) \\ \Leftrightarrow \psi(x) &= \frac{q N_{sub} w^2}{2 \epsilon_{si}} \left(1 - \frac{x}{w} \right)^2 \end{aligned} \quad 2.2.6$$

The definition of a surface potential ψ_s at $x=0$ (condition (iii)) leads to

$$\psi(x) = \psi_s \left(1 - \frac{x}{w} \right)^2 \quad 2.2.7$$

The depletion width can be expressed in terms of the surface potential and the doping concentration.

$$\begin{aligned}\psi_s &= \frac{q N_{sub} w^2}{2 \epsilon_{si}} \\ \Rightarrow w &= 2^{\frac{1}{2}} L_B (\beta \psi_s)^{\frac{1}{2}}\end{aligned}\tag{2.2.8}$$

where the Debye length L_B is defined by

$$L_B = \left(\frac{k T \epsilon_{si}}{q^2 N_{sub}} \right)^{\frac{1}{2}} \quad \text{and} \quad \beta = \frac{q}{kT}$$

This expression for the depletion region under an MOS capacitor is used in the derivation of the MOSFET model in section 2.3. Figure 2.2.3 shows the field and potential variations as calculated from the expressions above.

2.3 Basic MOSFET Model

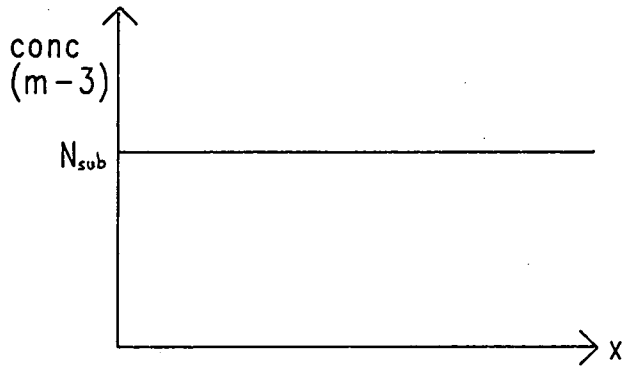
There are two different techniques most often used in deriving a model for the MOSFET. In the derivation below, a model for the transistor is obtained by summing elemental resistances in the channel.^{32,33,34,35} This leads to the simple analytical expressions used in many engineering models. In section 2.4, a second method is described in which Poisson's equation is solved and drift and diffusion components of current are summed to yield an equation for current. This leads to a similar but more complicated mathematical solution.

To a first approximation, the current flowing in the MOSFET before the gate voltage reaches threshold voltage can be considered to be zero. In subsequent parts of this thesis, a more accurate model for this region will be derived but at this stage it is assumed that the current is negligible.

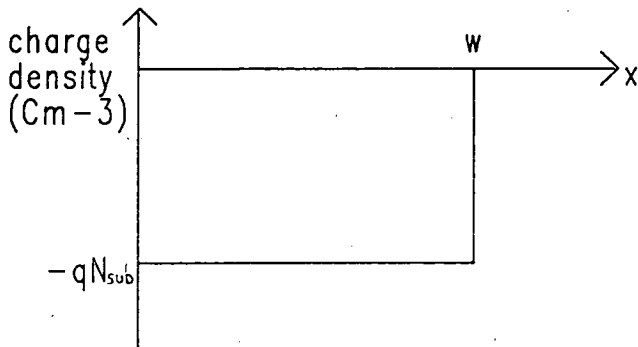
The following analysis derives an expression for current flowing in a MOSFET in the linear region of operation. The x, y and z directions are defined in figure 2.1.1. The following assumptions are made:

- (i) only drift current flows
- (ii) carrier mobility in the inversion layer is constant
- (iii) doping in the channel is uniform

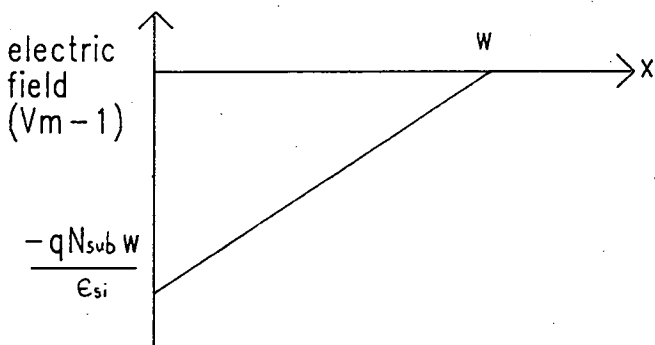
and (iv) the transverse electric field E_x is much greater than the parallel electric field E_y . This is known as the gradual channel approximation.



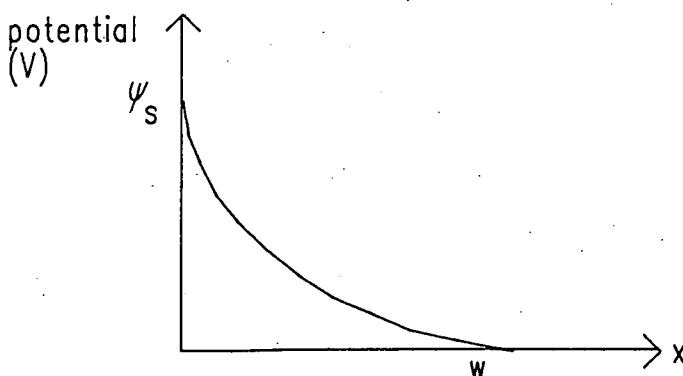
(a) uniform doping



(b) charge distribution
impurity atoms ionised
in depletion layer



(c) field distribution
from eq 2.2.4



(d) potential variation
from eq 2.2.7

Figure 2.2.3 The Depletion Layer

The total charge induced in the semiconductor at a distance y from the source is

$$Q_t(y) = \left(\frac{-V_g + V_{fb} + \psi_s(y)}{t_{ox}} \right) \epsilon_{ox} \quad 2.3.1$$

The inversion layer charge is equal to the total charge minus the depletion layer charge:

$$Q_i(y) = Q_t(y) - Q_b(y)$$

$$\Leftrightarrow Q_i(y) = \left(\frac{-V_g + V_{fb} + \psi_s(y)}{t_{ox}} \right) \epsilon_{ox} - Q_b(y) \quad 2.3.2$$

now

$$Q_b(y) = -q N_{sub} w$$

Using equation 2.2.7

$$Q_b(y) = - \left(2 \epsilon_{si} q N_{sub} (\psi(y) + 2\phi_b - V_b) \right)^{\frac{1}{2}} \quad 2.3.3$$

where $\psi_s(y)$ is expressed as $2\phi_b + \psi(y)$ and $\psi(y)$ is the voltage relative to the source at a distance y from the source. A surface potential of $2\phi_b$ results in a semiconductor surface which has a carrier concentration of opposite type but exactly the same as the unbiased silicon surface. This is considered to be the onset of strong inversion and hence threshold is defined as the point at which the surface potential reaches $2\phi_b$. Then the inversion charge at a distance y from the source is

$$Q_i(y) = \left(\frac{-V_g + V_{fb} + 2\phi_b + \psi(y)}{t_{ox}} \right) \epsilon_{ox} +$$

$$\left(2 q \epsilon_{si} N_{sub} (\psi(y) + 2\phi_b - V_b) \right)^{\frac{1}{2}} \quad 2.3.4$$

The conductivity at any part of the channel at depth x into the silicon is

$$\sigma_c(x) = q \mu n(x) \quad 2.3.5$$

where $n(x)$ is the number of free carriers per unit volume. Accounting for the area of conduction and the channel length yields the channel conductance

$$G = \frac{W}{L} \int_0^{x_i} \sigma_c(x) dx$$

where x_i is the inversion layer width.

$$G = \frac{q W \mu}{L} \int_0^{x_i} n(x) dx = \frac{W \mu |Q_i(y)|}{L} \quad 2.3.6$$

The elemental conductance is

$$dG = \frac{1}{dR} = \frac{W \mu |Q_i(y)|}{dy} \quad 2.3.7$$

From Ohm's law

$$dV = I_d dR = \frac{I_d dy}{W \mu |Q_i(y)|} \quad 2.3.8$$

Assuming μ is independent of voltage, substituting from 2.3.4 and integrating from source to drain

$$W \mu \int_0^{V_d} (V_g - V_{fb} - 2\phi_b - \psi(y)) C_{ox} -$$

$$(2 q \epsilon_{si} N_{sub})^{\frac{1}{2}} (\psi(y) + 2\phi_b - V_b)^{\frac{1}{2}} d\psi(y) = \int_0^L I_d dy$$

$$I_d = \mu C_{ox} \frac{W}{L} \left\{ \left[V_g - V_{fb} - 2\phi_b - \frac{V_d}{2} \right] V_d - \right.$$

$$\left. \mu C_{ox} \frac{W}{L} \left\{ \frac{2}{3} \frac{(2 q \epsilon_{si} N_{sub})^{\frac{1}{2}}}{C_{ox}} \left[(V_d + 2\phi_b - V_b)^{\frac{3}{2}} - (2\phi_b - V_b)^{\frac{3}{2}} \right] \right\} \right\} \quad 2.3.9$$

By expanding the term below from equation 2.3.9 using the Maclaurin expansion then

$$(V_d + 2\phi_b - V_b)^{\frac{3}{2}} - (2\phi_b - V_b)^{\frac{3}{2}} \approx \frac{3}{2} V_d (2\phi_b - V_b)^{\frac{1}{2}}$$

and so

$$I_d = \mu C_{ox} \frac{W}{L} \left\{ V_g - V_{fb} - 2\phi_b - \frac{V_d}{2} - \frac{(2 q \epsilon_{si} N_{sub})^{\frac{1}{2}}}{C_{ox}} (2\phi_b - V_b)^{\frac{1}{2}} \right\} V_d$$

$$I_d = \mu C_{ox} \frac{W}{L} \left\{ V_g - V_{th} - \frac{V_d}{2} \right\} V_d \quad 2.3.10$$

where the threshold voltage, V_{th} is

$$V_{th} = V_{fb} + 2\phi_b + \frac{(2 q \epsilon_{si} N_{sub})^{\frac{1}{2}}}{C_{ox}} (2\phi_b - V_b)^{\frac{1}{2}} \quad 2.3.11$$

This expression predicts the drain to source current in an MOS transistor in the linear region of operation. When the drain voltage becomes sufficiently high, V_{dsat} , the inversion charge at the drain end of the channel becomes zero.

$$Q_i(L) = 0 = \left[-V_g + V_{fb} + 2\phi_b + V_{dsat} \right] C_{ox} + (2q \epsilon_{si} N_{sub})^{\frac{1}{2}} \left[V_{dsat} + 2\phi_b - V_b \right]^{\frac{1}{2}}$$

$$\Rightarrow V_{dsat} = V_g - V_{fb} - 2\phi_b + \frac{q \epsilon_{si} N_{sub}}{C_{ox}^2} \left[1 - \left(1 + \frac{C_{ox}^2 (2V_g - 2V_{fb} - 2V_b)}{q \epsilon_{si} N_{sub}} \right)^{\frac{1}{2}} \right] \quad 2.3.12$$

The saturation current I_{dsat} is obtained by using the V_{dsat} found above in the expression for current 2.3.10.

2.4 Advanced MOSFET Model

The model formulated in 2.3 is useful for engineering, especially CAD, since it results in simple equations which can be evaluated quickly. A more accurate solution of the physical equations is provided below where, in particular, diffusion current is included. Diffusion current was first recognised as being important in MOSFETs by Pao and Sah in 1966³⁶ and the derivation of their double integral formula is explained below.^{37,28,38} Perhaps in the future, with the search for more accurate models for VLSI devices and with the development of more powerful computers, models which include diffusion current and involve a more rigorous mathematical derivation may be used for simulation.

First of all the Poisson equation is solved and Gauss' law applied to relate surface potential to the gate voltage. To begin with the substrate is assumed to be nondegenerate so that Maxwell-Boltzmann statistics are used to express the charge in the channel as a function of potential. The potential arises due to the voltage applied to the gate causing band bending ψ and due to the substrate bias V_b .

$$p = n_i \exp \left[\frac{q}{k T} (-\psi + V_b + \psi_f) \right] \quad 2.4.1a$$

$$n = n_i \exp \left[\frac{q}{k T} (-\psi_f + \psi - V_b) \right] \quad 2.4.1b$$

These equations are only valid in equilibrium. In the MOSFET there is current flow and so the quasi-Fermi levels ψ_p and ψ_n must be used. In an n-channel device, current flow is attributed to electrons so the quasi-Fermi level for holes is just

$$\psi_p = \psi_f$$

whereas for electrons

$$\psi_n = \psi_f - V_b \quad \text{at the source}$$

$$\text{and } \psi_n = \psi_f - V_b + V_d \quad \text{at the drain}$$

Revising equation 2.4.1 gives

$$p = n_i \exp \left[\frac{q}{k T} (-\psi + V_b + \psi_f) \right] \quad 2.4.2a$$

$$n = n_i \exp \left[\frac{q}{k T} (-\psi_n + \psi - V_b) \right] \quad 2.4.2b$$

These charge densities can now be used in the Poisson equation:

$$\text{div} (\epsilon_{si} \nabla(\psi)) = -\rho \quad 2.4.3$$

where the charge per unit volume ρ is

$$\rho = q (N_D - N_A + p - n) \quad 2.4.4$$

where N_D are ionised donors (positively charged) and N_A are ionised acceptors (negatively charged). In the bulk, without any applied potential, the net charge in the semiconductor must be zero.

$$N_D - N_A + p - n = 0$$

$$\Rightarrow N_D - N_A = n_i \exp(-\beta\psi_f) - n_i \exp(\beta\psi_f)$$

Hence

$$\begin{aligned} \rho = q n_i \left\{ \exp \left[\beta\psi_f - \beta\psi + \beta V_b \right] - \exp \left[\beta\psi - \beta V_b - \beta\psi_n \right] \right\} + \\ q n_i \left\{ \exp(-\beta\psi_f) - \exp(\beta\psi_f) \right\} \end{aligned} \quad 2.4.5$$

The Poisson equation becomes

$$\begin{aligned} \text{div} (\nabla\psi) = -\frac{q n_i}{\epsilon_{si}} \left\{ \exp \left[\beta\psi_f - \beta\psi + \beta V_b \right] - \exp \left[\beta\psi - \beta V_b - \beta\psi_n \right] \right\} - \\ \frac{q n_i}{\epsilon_{si}} \left\{ \exp(-\beta\psi_f) - \exp(\beta\psi_f) \right\} \end{aligned} \quad 2.4.6$$

It is assumed that $\frac{\partial^2\psi}{\partial z^2} = 0$ and also the gradual channel approximation applies :

$\frac{\partial^2 \psi}{\partial x^2} \gg \frac{\partial^2 \psi}{\partial y^2}$. Hence the Poisson equation can be reduced to 1-dimension

$$\frac{d^2 \psi}{dx^2} = -\frac{q n_i}{\epsilon_{si}} \left\{ \exp \left[\beta \psi_f - \beta \psi + \beta V_b \right] - \exp \left[\beta \psi - \beta V_b - \beta \psi_n \right] \right\} - \frac{q n_i}{\epsilon_{si}} \left\{ \exp(-\beta \psi_f) - \exp(\beta \psi_f) \right\} \quad 2.4.7$$

To integrate this, the mathematical identity below is used.

$$\frac{1}{2} \frac{d}{dx} \left(\frac{d\psi}{dx} \right)^2 = \frac{d\psi}{dx} \left(\frac{d^2 \psi}{dx^2} \right)$$

$$\frac{1}{2} \frac{d}{dx} \left(\frac{d\psi}{dx} \right)^2 = -\frac{q n_i}{\epsilon_{si}} \left\{ \exp(\beta \psi_f - \beta \psi + \beta V_b) - \exp(\beta \psi - \beta V_b - \beta \psi_n) \right\} \frac{d\psi}{dx} - \frac{q n_i}{\epsilon_{si}} \left\{ \exp(-\beta \psi_f) - \exp(\beta \psi_f) \right\} \frac{d\psi}{dx}$$

Integrate from below the depletion region where $\frac{d\psi}{dx}=0$, $\psi=V_b$ and $x=\infty$ to an arbitrary point

$$\begin{aligned} \left[\left(\frac{d\psi}{dx} \right)^2 \right]_{\infty}^x &= \left[-\frac{2 q n_i}{\epsilon_{si}} \left\{ -\frac{1}{\beta} \exp \left[\beta \psi_f - \beta \psi + \beta V_b \right] - \frac{1}{\beta} \exp \left[\beta \psi - \beta V_b - \beta \psi_n \right] \right\} \right]_{V_b}^{\psi} \\ &\quad - \left[\frac{2 q n_i}{\epsilon_{si}} \left\{ \psi \exp(-\beta \psi_f) - \psi \exp(\beta \psi_f) \right\} \right]_{V_b}^{\psi} \\ \Rightarrow - \left[\left(\frac{d\psi}{dx} \right)^2 \right]_x &= \frac{2 q n_i}{\beta \epsilon_{si}} \left\{ \exp \left[\beta \psi_f - \beta \psi + \beta V_b \right] + \exp \left[\beta \psi - \beta V_b - \beta \psi_n \right] \right\} + \\ &\quad \frac{2 q n_i}{\beta \epsilon_{si}} \left\{ \beta \psi \exp(\beta \psi_f) - \beta \psi \exp(-\beta \psi_f) - \exp(\beta \psi_f) - \exp(-\beta \psi_n) \right\} + \\ &\quad \frac{2 q n_i}{\beta \epsilon_{si}} \left\{ -\beta V_b \exp(\beta \psi_f) + \beta V_b \exp(-\beta \psi_f) \right\} \\ \left(\frac{d\psi}{dx} \right)_x &= - \left(\frac{2 q n_i}{\beta \epsilon_{si}} \right)^{\frac{1}{2}} \left\{ \exp \left[\beta \psi_f - \beta \psi + \beta V_b \right] + \exp \left[\beta \psi - \beta V_b - \beta \psi_n \right] + \right. \\ &\quad \left. (\beta \psi - \beta V_b) \left[\exp(\beta \psi_f) - \exp(-\beta \psi_f) \right] - \exp(\beta \psi_f) + \exp(-\beta \psi_n) \right\}^{\frac{1}{2}} \end{aligned}$$

$$\left(\frac{d\psi}{dx} \right)_x = - \left(\frac{2 q n_i}{\beta \epsilon_{si}} \right)^{\frac{1}{2}} F(\psi, \psi_n) \quad 2.4.8$$

$F(\psi, \psi_n)$ is used to represent the string of exponentials in the equation above 2.4.8 and the negative square root has been chosen since the potential is higher at the surface than in the bulk of the silicon. At the surface

$$\left(\frac{d\psi}{dx} \right)_0 = - \left(\frac{2 q n_i}{\beta \epsilon_{si}} \right)^{\frac{1}{2}} F(\psi_s, \psi_n)$$

Using Gauss' Law, assuming that there is no appreciable charge in the oxide, the electric field densities can be equated at the silicon-silicon dioxide interface

$$\epsilon_{ox} \left[\frac{-V_g + \psi_s}{t_{ox}} \right] = - \left(\frac{2 q n_i \epsilon_{si}}{\beta} \right)^{\frac{1}{2}} F(\psi_s, \psi_n) \quad 2.4.9$$

$$\Leftrightarrow \psi_s = V_g - \frac{1}{C_{ox}} \left(\frac{2 q n_i \epsilon_{si}}{\beta} \right)^{\frac{1}{2}} F(\psi_s, \psi_n) \quad 2.4.10$$

As stated previously, the quasi-Fermi levels at the source and drain are known so the surface potentials and carrier densities corresponding to a particular gate voltage can be found at these points in the channel. The next step is to sum drift and diffusion components of current density.

$$J = j_{drift} + j_{diff}$$

$$J = -q \mu n \frac{\partial \psi}{\partial y} + q D \frac{\partial n}{\partial y} \quad 2.4.11$$

This can be simplified by using the Einstein relationship and substituting for $\frac{\partial n}{\partial y}$ using

$$n = n_i \exp(\beta\psi - \beta\psi_n)$$

$$\frac{\partial n}{\partial y} = \beta n_i \exp(\beta\psi - \beta\psi_n) \left[\frac{\partial \psi}{\partial y} - \frac{\partial \psi_n}{\partial y} \right]$$

$$\frac{\partial n}{\partial y} = \beta n \left[\frac{\partial \psi}{\partial y} - \frac{\partial \psi_n}{\partial y} \right]$$

This yields

$$J = -q \mu n \frac{\partial \psi_n}{\partial y} \quad 2.4.12$$

Equation 2.4.12 can be integrated to give current

$$I = \frac{W}{L} \int_0^L \int_{x_i}^0 q \mu n(x) \frac{\partial \psi_n}{\partial y} dx dy$$

Assuming that the quasi-Fermi level is independent of x then

$$I = \frac{W q \mu}{L} \int_0^L \frac{\partial \psi_n(y)}{\partial y} \int_{x_i(y)}^0 n(x,y) dx dy$$

$$I = \frac{W q \mu}{L} \int_{\psi_f - V_b}^{\psi_f - V_b + V_d} \int_{x_i(y)}^0 n(x,y) dx d\psi_n \quad 2.4.13$$

The last requirement is to express n(x,y) in terms of ψ and ψ_n .

$$\int_{x_i(y)}^0 n(x,y) dx = \int_{\psi_f + V_b}^{\psi_s} n(\psi, \psi_n) \frac{1}{\frac{d\psi}{dx}} d\psi$$

From the solution of Poisson, equation 2.4.8

$$\frac{d\psi}{dx} = - \left(\frac{2 q n_i}{\beta \epsilon_{si}} \right)^{\frac{1}{2}} F(\psi, \psi_n)$$

$$\int_{\psi_f + V_b}^{\psi_s} n(\psi, \psi_n) \frac{1}{\frac{d\psi}{dx}} d\psi = \int_{\psi_f + V_b}^{\psi_s} \frac{n_i \exp(\beta\psi - \beta\psi_n - \beta V_b)}{\left(\frac{2 q n_i}{\beta \epsilon_{si}} \right)^{\frac{1}{2}} F(\psi, \psi_n)} d\psi$$

where $\psi_f + V_b$ is the potential at the bottom of the inversion layer and ψ_s is the potential at the surface. Finally

$$I = \frac{W q \mu}{L} \left(\frac{\beta \epsilon_{si}}{2 q n_i} \right)^{\frac{1}{2}} n_i \int_{\psi_f - V_b}^{\psi_f - V_b + V_d} \int_{\psi_f + V_b}^{\psi_s} \frac{\exp(\beta\psi - \beta\psi_n - \beta V_b)}{F(\psi, \psi_n)} d\psi d\psi_n \quad 2.4.14$$

This equation which is the Pao-Sah double integral formula³⁶ applies to all regions of MOSFET operation but is rather complicated for use as an engineering model for chip design. Most engineering models are based on the basic model described in 2.3. In the description of various models which follows, the Brews' charge sheet model²³ is based on this more precise analytical approach.

2.5 The SPICE 2 Levels 1 and 3 MOSFET models

SPICE became the industry standard for integrated circuit simulation during the 1970's. Many users have now incorporated their own improvements or turned to more recently developed competitors but SPICE remains the industry standard against which all others are compared. In APPENDIX A, the equations for the SPICE Levels 1 and 3 MOSFET models are listed.^{22,39} Both models arise from the Basic MOSFET theory developed in 2.3. The level 1 model is very simple and is used in circuit simulation to provide a fast but less accurate prediction of circuit operation. The level 3 model includes several empirical factors to cope with small geometry operation.

The level 1 model is described first. To a first approximation, the current flowing below threshold is essentially zero and for this complexity of model, it is assumed to be zero. The expression for threshold voltage in equation 2.3.11 is simplified to

$$V_{th} = V_{to} + \gamma |V_b|^{1/2} \quad 2.5.1$$

where

$$V_{to} = V_{fb} + 2\phi_b + \gamma |2\phi_b|^{1/2}$$

$$\gamma = \frac{(2q \epsilon_{si} N_{sub})^{1/2}}{C_{ox}}$$

and

$$2\phi_b = \frac{2kT}{q} \ln \left(\frac{N_{sub}}{n_i} \right)$$

Both V_{to} and γ are electrically measured parameters. The standard expression for current in the linear region is used:

$$I_d = \mu_{eff} C_{ox} \frac{W}{L} \left[V_g - V_{th} - \frac{V_d}{2} \right] V_d \quad 2.5.2$$

where $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$. Several parameters are needed to evaluate this expression. The oxide thickness, t_{ox} provides the oxide capacitance per unit area C_{ox} . The effective length L and effective width W are calculated from

$$W = W_m - 2 \Delta_w \quad 2.5.3$$

$$L = L_m - 2 L_d \quad 2.5.4$$

where Δ_w and L_d are electrically measured reductions from the mask width W_m and mask length L_m , respectively. The carrier mobility μ_{eff} is found using the empirical relationship

$$\mu_{eff} = \frac{\mu_o}{1 + \theta(V_g - V_{th})} \quad 2.5.5$$

where μ_o is the maximum carrier mobility found when $V_g = V_{th}$ and θ is a constant relating mobility and gate voltage.

Finally, the saturation region has to be modelled. To do this, the derivative of the current in the linear region is

$$\frac{dI_d}{dV_d} = \mu_{eff} C_{ox} \frac{W}{L} [V_g - V_{th} - V_d]$$

The point of saturation is obtained where $\frac{dI_d}{dV_d} = 0$.

$$V_{dsat} = V_g - V_{th} \quad 2.5.6$$

For simplicity, the current is assumed to be a constant in saturation. Using the V_{dsat} found above in the linear region expression (equation 2.5.2) gives

$$I_d = \mu_{eff} C_{ox} \frac{W}{L} \frac{(V_g - V_{th})^2}{2} \quad 2.5.7$$

This model, despite its simplicity, can produce fairly accurate simulations for device lengths of 10 μm and above and even at smaller dimensions will provide a fast general indication of how the device/circuit will operate under a particular bias condition. For greater accuracy the level 3 model uses empirical terms to model small geometry effects and this model is described below.

The technique for modelling threshold voltage is to add terms accounting for various different effects to the level 1 expression.

$$V_{th} = V_{fb} + 2\phi_b - \sigma V_d + \gamma F_s (2\phi_b - V_b)^{\frac{1}{2}} + F_n (2\phi_b - V_b) \quad 2.5.8$$

The band-bending necessary for channel inversion is the sum of the flatband voltage V_{fb} and $2\phi_b$ which is equal to twice the difference between the fermi level and the intrinsic fermi level of the unbiased device. The depletion of the silicon below the surface is defined as $\gamma F_s (2\phi_b - V_b)^{\frac{1}{2}}$ where F_s is an empirical factor modelling the effect on this

term for a short channel device. Similarly, F_n measures the effect of a narrow channel on threshold. For a higher drain voltage, the device turns on more quickly and this is modelled by the coefficient σ .

F_n , F_s and σ are all geometry dependent and are derived from the basic parameters δ , L_d and η respectively. The geometrical variation of these effects is allowed for as follows:

$$F_n = \frac{2 \pi \delta \epsilon_{si}}{4 W C_{ox}} \quad 2.5.9$$

which is proportional to t_{ox} and inversely proportional to W .

$$F_s = 1 - \frac{x_j}{L} \left\{ \left(\frac{W_c}{x_j} + \frac{L_d}{x_j} \right) \left[1 - \left(\frac{W_{ps}}{x_j + W_{ps}} \right)^2 \right]^{\frac{1}{2}} - \frac{L_d}{x_j} \right\} \quad 2.5.10$$

where

$$\frac{W_c}{x_j} = D0 + D1 \frac{W_{ps}}{x_j} + D2 \left[\frac{W_{ps}}{x_j} \right]^2 \quad 2.5.11$$

and

$$W_{ps} = X_d (2\phi_b - V_b)^{\frac{1}{2}} \quad 2.5.12$$

The empirical constants $D0$, $D1$ and $D2$ have values of 0.063135, 0.801329 and -0.0111077 respectively. The second term in the equation for F_s is reduced as the length of the device increases. Thus, F_s increases as the length increases and consequently the voltage needed to deplete the region below the surface in the threshold equation is increased.

$$\sigma = \frac{\eta C0}{C_{ox} L^3} \quad 2.5.13$$

The constant $C0$ has the value $8.15E -22 \text{ Fm}^{-3}$ according to Vladimirescu and Liu.²² σ is inversely dependent upon L^3 with the result that its effect quickly decreases as L is increased.

The two quantities related to the bulk, V_b and $2\phi_b$, are treated differently when $V_b > 0$ and hence $2\phi_b - V_b$ is possibly negative. In that case, in all the above equations, the back bias term $2\phi_b - V_b$ is replaced by

$$\frac{2\phi_b}{\left[1 + \frac{V_b}{4\phi_b}\right]^2}$$

Since there is no precise transition from subthreshold, a voltage V_{on} is defined just above threshold at which the exponential dependence of drain current upon gate voltage ceases.

$$V_{on} = V_{th} + \frac{N k T}{q} \quad 2.5.14$$

where

$$N = 1 + \frac{C_s}{C_{ox}} + \frac{C_d}{C_{ox}} \quad 2.5.15$$

$$\frac{C_d}{C_{ox}} = \frac{Q_b}{2 C_{ox} (2\phi_b - V_b)} \quad 2.5.16$$

and

$$\frac{C_s}{C_{ox}} = \frac{q N_{fs}}{C_{ox}} \quad 2.5.17$$

When gate voltage is both above V_{on} and above the drain voltage plus threshold so that the whole channel is in a state of inversion the current through the device is linearly dependent upon the gate voltage. The expression governing current which is based upon the theory in 2.3, is

$$I_d = \mu_{eff} C_{ox} \frac{W}{L} V_{dxx} \left[V_g - V_{th} - \frac{1+F_b}{2} V_{dxx} \right] \quad 2.5.18$$

where C_{ox} is the gate oxide capacitance per unit area as in level 1 and W and L are the effective width and length of the device respectively. These are calculated in the same way as for level 1 (see equations 2.5.3 and 2.5.4).

The effective mobility μ_{eff} is calculated from a maximum low field mobility μ_o just as for level 1, except that in level 3, it is dependent upon drain bias as well. There are two relationships used by the model.

$$\mu_s = \frac{\mu_o}{1 + \theta(V_{gsx} - V_{th})} \quad 2.5.19$$

where

$$V_{gsx} = \max (V_g, V_{on}) \quad 2.5.20$$

and

$$\mu_{eff} = \frac{\mu_s}{1 + \frac{V_{dsx} \mu_s}{L v_{max}}} \quad 2.5.21$$

These equations govern the variation of mobility of carriers in all regions although in subthreshold the effect of the gate voltage will be zero. The effect of the parallel field, due to the drain voltage, is contained in the second equation where v_{max} is the parameter. Physically, according to Baum and Beneking⁴⁰, carriers reach their maximum (saturation) velocity before the drain potential is high enough to pinch off the channel.

The device exhibits an exponential dependence of current on gate voltage for gate voltages lower than V_{on} . The equation from which current is calculated is

$$I_d = \mu_{eff} C_{ox} \frac{W}{L} V_{dsx} \left[V_{on} - V_{th} - \frac{1+F_b}{2} V_{dsx} \right] \exp \left[\frac{q (V_g - V_{on})}{N k T} \right] \quad 2.5.22$$

where

$$V_{dsx} = \min (V_d, V_{dsat}) \quad 2.5.23$$

and

$$F_b = \frac{\gamma F_s}{4 (2\phi_b - V_b)^{\frac{1}{2}}} + F_n \quad 2.5.24$$

The slope of the subthreshold characteristic is primarily dependent on the parameter N_{fs} from which N is calculated. Continuity of current is obtained at V_{on} where the exponential term becomes $\exp(0)=1$ and so the current equation is the same as that for the linear or saturation region.

When the drain voltage is increased so that it is less than a threshold below the gate bias, the drain end of the channel ceases to be in inversion. Conduction continues in this area since the high parallel electric field forces carriers across into the channel. The analysis below was carried out by Grove and Frohman-Bentchowsky.⁴¹ The first order approximation for the voltage V_{dsat} at which this mechanism begins is

$$V_{dsat} = \frac{V_{gsx} - V_{th}}{1 + F_b} \quad 2.5.25$$

If carrier velocity saturation is being taken into account then

$$I_d = Q(\text{drain}) v_{\max} \quad 2.5.26$$

where the charge per metre at the drain end of the channel

$$Q(\text{drain}) = W C_{ox} (V_{gxx} - V_{th} - (1+F_b)V_{dsat}) \quad 2.5.27$$

Substituting in the above

$$\frac{W}{L} \mu_s C_{ox} \left[V_{gxx} - V_{th} - \frac{1+F_b}{2} V_{dsat} \right] V_{dsat} = W C_{ox} (V_{gxx} - V_{th} - (1+F_b)V_{dsat}) v_{\max} \quad 2.5.28$$

Rearranging

$$V_{dsat} = \frac{V_{gxx} - V_{th}}{1 + F_b} + \frac{v_{\max} L}{\mu_s} - \left[\left[\frac{V_{gxx} - V_{th}}{1 + F_b} \right]^2 + \left[\frac{v_{\max} L}{\mu_s} \right]^2 \right]^{\frac{1}{2}} \quad 2.5.29$$

At the point of saturation, the saturation current is defined where $V_d = V_{dsat}$. Any increase in current beyond that point is due to the lengthening of the pinched off region in the channel thus reducing the channel length. Differentiating the saturation drain current gives the saturation drain conductance, G_{dsat} . From this the maximum transverse electric field at which pinch-off begins can be found.

$$G_{dsat} = \frac{I_{dsat} V_{dsx}}{V_{dsx} + \frac{L v_{\max}}{\mu_s}} \quad 2.5.30$$

$$E_{\max} = \frac{I_{dsat}}{G_{dsat} L} \quad 2.5.31$$

The length reduction when v_{\max} is not defined is

$$L_{del} = [\kappa X_d^2 (V_d - V_{dsat})]^{\frac{1}{2}} \quad 2.5.32$$

otherwise

$$L_{del} = \left[\left[\frac{X_d^2 E_{\max}}{2} \right]^2 + \kappa X_d^2 (V_d - V_{dsat}) \right]^{\frac{1}{2}} - \frac{X_d^2 E_{\max}}{2} \quad 2.5.33$$

where κ is an empirical factor used to increase the slope of the characteristic in saturation. As the depletion region extends further from the drain, it does not increase with drain voltage at the same rate. The above relationship does not account for this

and hence for high drain voltage, if the change in length becomes too great, a negative channel length is the result. To avoid this and model the device more realistically, if the channel length reduction is greater than half the device length then the formula

$$L_{del} = L - \frac{L^2}{4 L_{del}} \quad 2.5.34$$

is used. The current in saturation is then given by

$$I_d = \frac{I_{dsat}}{1 - \frac{L_{del}}{L}} \quad 2.5.35$$

This summarises the SPICE Level 3 model which is still the most widely used model that is suitable for simulating small geometry devices.

2.6 Other MOSFET Models

Three other models are discussed here: the Brews' charge sheet model²³, the CASMOS model²⁴ and the improvements to the SPICE level 3 model proposed by Wright.²⁵ The Brews' model is a mathematical model derived in the same way as the Pao-Sah double integration formula (equation 2.4.14) and includes diffusion current. CASMOS is intended to be computationally efficient with a minimal number of device parameters and is specifically engineering orientated. Wright bases his model improvements on new explanations for physical effects in particular for carrier velocity saturation and subthreshold current. Both output conductance and transconductance become continuous.

The Brews' charge sheet model is formulated in a similar fashion to the Pao-Sah formula except that only acceptor ions and electrons are considered when expressing the charge density. The resulting expression for surface potential, derived from Poisson and Gauss, is

$$\psi_s = V_g - \frac{1}{C_{ox}} \left(\frac{2 q n_i \epsilon_{si}}{\beta} \right)^{\frac{1}{2}} \left\{ \exp \left[\beta \psi_s - \beta \psi_n - \beta V_b \right] + \beta \psi_s \exp(\beta \psi_f) - \beta V_b \exp(\beta \psi_f) \right\}^{\frac{1}{2}} \quad 2.6.1$$

Comparing this with equation 2.4.10 shows that this is the same equation except that several terms have been neglected. The starting point for deriving an expression for

current is equation 2.4.12 and current is reduced to an arithmetic expression which applies to all regions of operation.

$$\begin{aligned}
 I_d = & \frac{\mu W}{\beta L} \left\{ C_{ox} (1 + \beta V_g) (\psi_s - \psi_{s0}) - \frac{\beta}{2} C_{ox} (\psi_s^2 - \psi_{s0}^2) \right\} - \\
 & \frac{\mu W}{\beta L} q N_{sub} L_B 2^{\frac{1}{2}} \frac{2}{3} \left\{ (\beta \psi_s - 1)^{\frac{3}{2}} - (\beta \psi_{s0} - 1)^{\frac{3}{2}} \right\} + \\
 & \frac{\mu W}{\beta L} q N_{sub} L_B 2^{\frac{1}{2}} \left\{ (\beta \psi_s - 1)^{\frac{1}{2}} - (\beta \psi_{s0} - 1)^{\frac{1}{2}} \right\}
 \end{aligned} \tag{2.6.2}$$

The derivative of this expression for current is continuous. Continuous derivatives are very useful in circuit simulation. Most circuit simulators try to find a d.c. solution for all nodes in the circuit. This is done using an iterative method involving the first derivatives and a discontinuity in these derivatives can disrupt the process causing it to become very slow or fail. Two further properties of the model are that the channel conductance approaches zero asymptotically and the transconductance saturates asymptotically. The combination of these facts means that the device gain, which is the product of the transconductance and the output resistance, remains finite as gate and drain voltages increase. This is realistic since practical devices do not have infinite gain.

The developers of CASMOS, Oakley and Hocking from Plessey, felt that the SPICE models had two major drawbacks. Firstly the factors included to cope with different geometry devices were unsatisfactory and secondly the complex nature of the level 3 model made obtaining parameters difficult. There were three objectives CASMOS had to fulfil:

(i) the model should be accurate for devices of between 1 μm and 50 μm operating under voltages of between 0 and 5V with a substrate bias between 0 and -2.5V.

(ii) the model should maintain a reasonable degree of simplicity so that parameters can be determined easily and unambiguously.

and (iii) the model should have a physical basis so that the electrical parameters could provide a link with the process.

The added bonus of a fairly simple model is that it is computationally efficient.

The model has the same basic form as SPICE level 1. The subthreshold

current is equal to zero and this is the main weakness of the CSMOS model. In particular for small geometry circuits, where supply voltages and threshold voltages may be lower, and for analogue circuits, subthreshold leakage currents have an important role. The d.c. device model has seven parameters. For the basic equations, the device gain, the threshold voltage, the body effect coefficient and the carrier mobility coefficient are required. Three second order effects are modelled empirically to improve the accuracy of the model. These are carrier velocity saturation, channel shortening in saturation and static feedback. The band bending necessary for inversion, $2\phi_b$, is included in the body effect in the same way as it is for SPICE level 3.

The geometry dependences of the seven parameters governing the effects mentioned above are modelled empirically. Firstly, the reductions of the mask widths and lengths during processing are found. This can be done by comparing the gains of devices of different sizes. The gain is altered according to the aspect ratio of the particular device to be simulated. The body effect coefficient and mobility modulation coefficient are dependent upon both length and width whereas carrier velocity saturation, channel shortening and static feedback are only length dependent. Finally a short channel factor is introduced to cope with the sharper than expected reduction in threshold voltage for very short channels.

This is an outline of the CSMOS model. The philosophy of maintaining a physical basis for the model while using empirical factors to account for second order effects allows the model to maintain its accuracy over a wide range of operating voltages. Further empirical factors enable the model to cope accurately with a wide range of geometries. Nevertheless the model remains in essence very simple making it computationally efficient and allowing the parameters to be found easily. The biggest shortfall of the model is the lack of any sort of expression covering the subthreshold region as mentioned above. Just as for SPICE, the conductance is discontinuous from region to region.

Wright overcomes the two problems with SPICE, highlighted by Oakley and Hocking, by using numerical optimisation for parameter extraction and a preprocessor to cope with different geometries.³⁹ Wright presents new explanations of physical phenomena and revises the model equations to reflect these. By redefining the voltages at which the transitions from one region to another are made, continuity of

conductance and transconductance are obtained.

The basic model equations no longer contain the complex, analytical, geometrical factors: F_s and F_n that are in SPICE. These complicated the model and had to be evaluated very precisely in order to enhance the accuracy of the simulation.

The mobility modulation by gate voltage begins at flatband voltage rather than threshold voltage.

$$\mu_s = \frac{\mu_o}{1 + \theta (V_g - V_{fb})} \quad 2.6.3$$

V_{fb} is a fixed quantity and so derivatives are easier to calculate for this new relationship.

The relationship between drain voltage and mobility applies more to holes than electrons since it is derived from the empirical equations between hole velocity v , and electric field E .

$$v = \frac{\mu_s}{1 + \frac{\mu_s E}{v_{\max} L}} \quad 2.6.4$$

Electrons, being much more mobile, reach their saturation velocity more rapidly and their corresponding relationship is

$$v = \left[\frac{\mu_s}{1 + \left(\frac{\mu_s E}{v_{\max} L} \right)^2} \right]^{\frac{1}{2}} \quad 2.6.5$$

This would make the current equation excessively complicated and therefore for electrons, a mobility multiplier which is drain voltage dependent is used.

In saturation, Wright questions the philosophy of an increase in the drain depletion region effectively shortening the channel and leading to the increase in current in saturation. It is stated that both charge density and carrier velocity remain unchanged and hence the current cannot change by this mechanism. Instead, the saturation drain voltage is reduced as the depletion region moves along the channel; there is a higher charge density and more current.

Although charge density and carrier velocity do not change in the gate controlled part of the channel, this region is shortened. In the depletion region, carriers experience a very high electric field and correspondingly travel at high velocities. Hence the average carrier velocity over the whole channel is increased and a higher current results. This is the more traditional view which is not disproved by Wright. Continuity of the derivative from the linear to the saturation region is achieved by defining a transition voltage V_{tran} at a point where the derivatives are equal.

$$V_{tran} = (1 - \rho_t) V_{dsat} \quad 2.6.6$$

The parameter ρ_t is used to adjust V_{tran} . A similar method is used at the step from the subthreshold to the linear region where V_{geff} is defined.

In the subthreshold region, it is predicted that current is source barrier limited rather than diffusion limited for short channel devices. The exponential dependence is maintained with the introduction of V_{geff} as mentioned but the fitting parameter N_{fs} from SPICE is replaced by s where

$$n = s \left(1 + \frac{C_d}{C_{ox}} \right) \quad 2.6.7$$

since N_{fs} is not physically realistic.

The new model overcomes the discontinuities in the derivatives and some of the unrealistic physical explanations contained in the SPICE model. It is still a complex model however and so parameter extraction is complicated, often requiring numerical optimisation. The model will be inefficient with regard to CPU time in comparison with the largely empirical CSMOS.

2.7 Numerical Device Modelling : MINIMOS

The advantage of an analytical MOSFET model, where device operation is described by several arithmetic expressions, is that the overall device operation can be calculated quickly and efficiently. In certain cases, particularly for very short channels, the inaccuracies in the analytical models become large. The assumptions made in the derivation of the model, when solving the fundamental, physical equations, become invalid. The result is that empirical factors, accounting for second order effects, have

to be introduced in the final equations in order to preserve model accuracy. This is usually reasonably successful down to channel lengths of $1\mu\text{m}$ but complicates the model and increases the number of parameters. Using numerical techniques to solve the fundamental physical equations, avoids some of the approximations needed to achieve an analytical solution. This method can easily be extended to two or three dimensions to provide very accurate simulation of even very small devices ($<1\mu\text{m}$). It can also show the internal state of the device: electric fields, carrier concentrations and potential. The major disadvantage of a numerical device simulator is that a great amount of computation is required for each bias condition.

The writers of MINIMOS⁴² had two purposes for their program. Firstly, they aimed to make the program as flexible as possible by allowing the user to specify what is input, what is calculated and what is output. Secondly they wished to make as many simplifications as possible to the original, physical equations in order to retain accuracy but reduce the amount of computation. The program uses finite differences to solve Poisson's equation, the continuity equation and the expression for current density. The current density is the sum of the drift and diffusion components. Six assumptions are made to achieve the two purposes mentioned above:

- (i) homogeneity of permittivity
- (ii) completely ionised impurity
- (iii) no bandgap narrowing
- (iv) no generation or recombination
- (v) only minority carriers contribute to current

and (vi) homogeneous temperature distribution.

Empirical relationships for p and n mobilities are implemented where the mobilities are functions of electric field, temperature, position in the channel and intrinsic and extrinsic carrier concentrations.

Numerical simulation overcomes some of the assumptions required to obtain an analytical solution for device operation. It is much slower to compute than an analytical result but the array of points at which carrier concentration, electric field and potential are calculated within the device can provide a detailed insight into how the device is operating. This can be used to obtain threshold voltage, inversion and depletion layer widths and the pinch-off point.

Chapter 3 : MOS Processing

3.1 The EMF NMOS Process

The wafer fabrication process consists of over 50 separate steps which combine to produce integrated circuits containing thousands of MOSFETs. These steps include cleaning, growing layers on the wafer, etching material off the wafer and doping the silicon with an impurity. Some steps have a direct effect on the electrical characteristics of the devices being manufactured, while others only define physical features. The complete EMF NMOS process is described to show how each step leads to the formation of the finished devices.

A complete summary of the the major process steps in the EMF $6\mu m$ NMOS process is provided in table 3.1.1.^{43,44} No cleaning stages or resist strips are included, but in the description below, important resist strips and cleans are mentioned. More details on lithography and ion implantation are given in sections 3.2 and 3.3.

The starting material is a 3-inch diameter, 14-20 Ωcm , $\langle 100 \rangle$ orientation, p-type silicon wafer. The $\langle 100 \rangle$ crystal orientation results in the lowest incidence of surface states at the Si-SiO₂ interface. These surface states have energy levels in the bandgap for silicon, they degrade device transconductance and can also lead to oxide breakdown. The first objective is to produce thick areas of recessed silicon dioxide, called field oxide, to provide isolation between devices on the wafer. This is done by a technique called LOCOS.⁴⁵ A silicon nitride layer is lithographically patterned and etched to define the areas in which the thick field oxide is to be grown. Boron is implanted and then 1.3 μm of silicon dioxide is grown in the exposed areas. The boron ensures that there is no significant conduction under the thick oxide when normal operating voltages are applied to the layers above. At the edges of the field region, there is a gradual increase in the oxide thickness which results in the formation of a bird's beak shape which protrudes into the active region (see figure 3.1.1). This, coupled with the diffusion of the field implant, leads to the effective device widths being less than the mask device widths. This accounts for the Δ_w parameter in the SPICE models (see section 2.4).

Table 3.1.1 Major Process Steps

Step No.	Step Type	Details
1	SUBSTRATE	14-20 Ωcm <100> p-type 3-inch diam.
2	OXIDATION	Buffer Oxide of 500 A
3	NITRIDE DEPOSITION	500 A Thick
4	OXIDATION	Surface Masking Oxide
5	LITHOGRAPHY	Field Oxide Definition
6	FIELD IMPLANT	B_{11}^+ at 130keV dose $2E13 \text{ atoms cm}^{-2}$
7	OXIDE ETCH	Remove layer for field oxide growth.
8	NITRIDE PLASMA ETCH	Remove layer for field oxide growth
9	FIELD OXIDATION	1.3 μm Thick
10	OXIDE ETCH	Remove Masking Layers
11	NITRIDE PLASMA ETCH	Remove Masking Layers
12	OXIDE ETCH	Remove Masking Layers
13	LITHOGRAPHY	Depletion Implant Mask
14	DEPLETION IMPLANT	As_{75}^{++} at 90keV dose $1.5E12 \text{ atoms cm}^{-2}$
15	GATE OXIDATION	800 A Thick
16	ENHANCEMENT IMPLANT	B_{11}^+ at 40keV dose $4.0E11 \text{ atoms cm}^{-2}$
17	ANNEAL	950 C Repair Lattice & Activate Implant
18	LITHOGRAPHY	Buried Contact Mask
19	OXIDE ETCH	Clear Diffusion Area for Contacts
20	POLY DEPOSITION	0.35 μm thick

Table 3.1.1 Major Process Steps(cont)

Step No.	Step Type	Details
21	POLY OXIDATION	Provides an etch mask
22	LITHOGRAPHY	Poly patterning
23	OXIDE ETCH	Remove oxide from polysilicon
24	POLY PLASMA ETCH	Pattern polysilicon
25	OXIDE ETCH	Remove gate and poly oxide
26	PHOSPHOROUS DEPOSITION	Junction diffusion & Poly doping
27	PHOSPHOROUS DEGLAZE	Remove phosphosilicate glass
28	POLY OXIDATION	Aid resist adhesion
29	REFLOW PYRO DEPOSITION	7500 A Thick
30	FIRST REFLOW	Smooth over sharp poly edges
31	LITHOGRAPHY	Opening contact holes in pyro
32	PYRO ETCH	Etch contact holes
33	SECOND REFLOW	Smooth edges at 1050 C
34	ALUMINIUM EVAPORATION	1.2 μ m thick
35	LITHOGRAPHY	Define metal pattern
36	ALUMINIUM ETCH	Pattern metal
37	SINTER	435 C Anneal
38	PYRO DEPOSITION	7500 A Thick
39	LITHOGRAPHY	Define contact pads
40	PYRO ETCH	Open bonding pads

The substrate is lightly p doped and so in order to produce the depletion devices with the correct threshold voltages, a donor impurity, in this case Arsenic, has to be introduced into the areas in which the depletion devices are to be formed. Arsenic is implanted in the areas defined by the second lithography stage. The high temperature oxidation which follows, repairs the silicon lattice which was damaged by the implant and causes the arsenic to ionise and enter the lattice. The result is that the silicon in the depletion channel region becomes n-type as required.

The surface of the wafer is now thoroughly cleaned in preparation for the growth of the gate oxide. The quality of the interface and gate oxide is critically important for the performance of the devices. Organic and inorganic material must be removed as well as any significant thickness of native oxide, (the quality of which is liable to be poor) so that there is nothing to prohibit the growth of a precise thickness of oxide. The gain of the device is dependent upon the oxide thickness. HCl is used during oxidation to remove the troublesome sodium ions which form mobile charge in the oxide.

The substrate was specifically chosen with a light doping in order to minimise currents into the substrate. The gate material, which has not yet been deposited, is n^+ doped polycrystalline silicon (poly). This will tend to invert the p-type silicon surface because of the metal(poly)-semiconductor work-function. The bandgap of silicon is 1.1V and so the work-function difference between a heavily doped n-type poly gate and a p-type substrate is of the order of -1.0V.²⁸ Assuming that the cleaning procedure before gate oxidation restricts the oxide and surface state charge to insignificant numbers, then the flat-band voltage is approximately -1.0V. The low substrate concentration and flat-band voltage of -1.0V means that the devices are depletion or at least that the threshold voltages are very low. Boron is implanted into the surface of the wafer to increase the threshold voltages and it enters all areas (including the depletion channel regions) except the field areas where it doesn't have sufficient energy to penetrate the thick oxide. Annealing is used to repair the crystal lattice damaged by the implanted ions and to activate the impurity. The position and quantity of this implant is a critical factor in determining the threshold voltages of the resulting devices (see Section 3.5). Most device models assume that there is uniform concentration throughout the substrate. The introduction of the channel implant has meant that this is no longer the case although, after the subsequent high temperature

Figure 3.1.1 Field Isolation

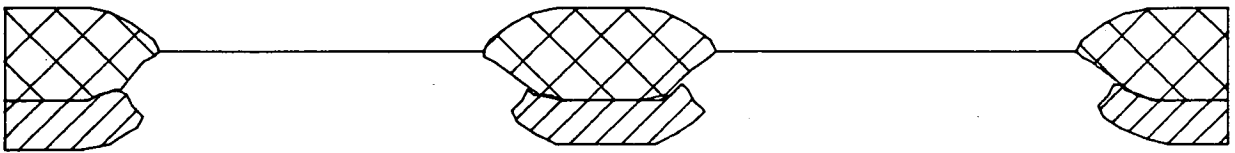
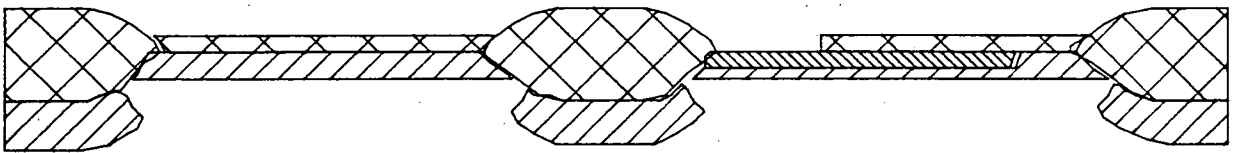





Figure 3.1.2 Channel Implants and Buried Contact



-  Oxide
-  Boron Implant
-  Arsenic Implant

steps, the high diffusivity of boron results in a fairly constant impurity concentration throughout the inversion layer and depletion region.

The next layer to be deposited on the wafer is polysilicon for transistor gates and interconnect. In preparation for the deposition, the third mask defines the areas for buried contacts where poly makes contact with diffusion. The oxide is etched (see figure 3.1.2) and a $0.35 \mu m$ layer of poly applied using low pressure chemical vapour deposition (LPCVD). The upper surface of the poly is oxidised to provide an adherent surface for the photoresist and to assist in masking during the poly etch. Isotropic etching leads to undercutting of the gate and hence a reduction from the mask channel length. This contributes to the channel length reduction L_d . Here the poly is plasma etched and so the contribution to L_d is very small. The oxide is stripped from the surface of the remaining poly.

Phosphorous is diffused into the wafer to increase the conductivity of the poly and to form the source and drain regions in the substrate. The interconnect becomes sufficiently conductive so that there are no excessive RC time delays between gates. Phosphosilicate glass, which forms on the surface of the poly during diffusion, is etched (see figure 3.1.3) and the top $0.2 \mu m$ of the poly is oxidised. The oxidation provides a good adherent surface for the $0.75 \mu m$ of pyrolytic oxide (pyro) which is now deposited and also provides an extra safeguard against pinholes in the insulating pyro. This layer isolates the polysilicon from the metal interconnect above.

The next step in the process is a reflow at a high temperature in order to smooth the pyro over the sharp edges in the poly and to drive-in the impurities in the source and drain regions. In particular, the high temperature reflow stages lead to substantial redistribution of the impurities. The areas which were heavily doped to form the source and drain extend into the substrate. This causes an increase in the junction depth and junction capacitance and also the sideways diffusion further reduces the effective length of the channel. This is the other component of the length shortening parameter L_d . The field implant can spread into the channel region during reflow, resulting in an increase in Δ_w on top of that caused by the bird's beak in the field oxide.

A further mask defines holes to be etched in the pyro to enable the metal to

Figure 3.1.3 After Phosphorous Diffusion

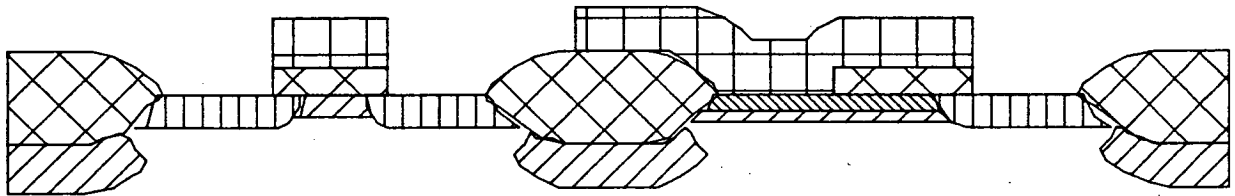
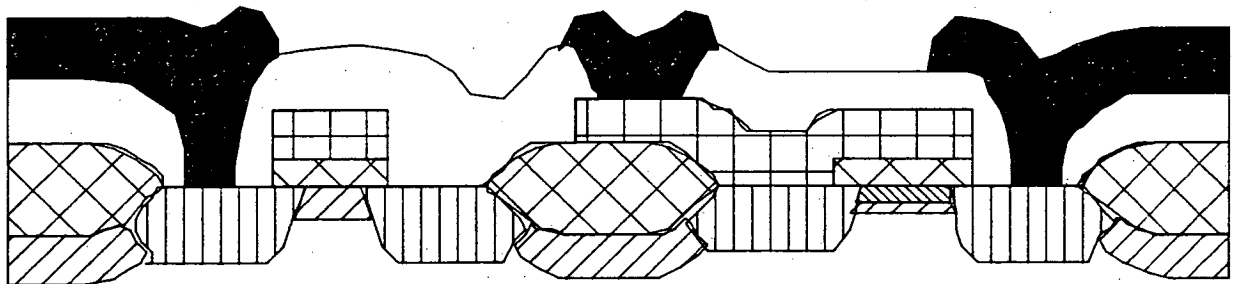









Figure 3.1.4 After Etching Metal



- | | | | |
|---|-----------------------|---|-------------|
|  | Oxide |  | Polysilicon |
|  | Boron Implant |  | Aluminium |
|  | Arsenic Implant |  | Pyro |
|  | Phosphorous Diffusion | | |

make contact with the poly and diffusion areas below. After etching the contact holes, a second reflow takes place to smooth the edges of the pyro around the contact holes allowing the metal to flow in more easily. A layer of aluminium $1.2 \mu\text{m}$ thick is evaporated onto the surface of the wafer. The aluminium is patterned and etched and then a low temperature anneal (termed a sinter) is carried out to ensure that there is a good ohmic contact between the aluminium and the silicon (see figure 3.1.4).

The final stage is the deposition of 7500 Å of pyro to seal the circuit and protect the aluminium interconnect. One last masking step is used to open windows in this pyro to allow access to contact pads and test points.

3.2 Lithography

Lithography^{46,47} is central to the whole fabrication process which is a series of depositing or growing layers on wafer surface; patterning those layers and removing the unwanted material (figure 3.2.1). Lithography is the process whereby patterns are transferred from a predefined mask onto the wafer surface. These patterns, at different points in the fabrication process, define areas of thick oxide isolation, interconnect patterns and contact holes. Layout engineers allow large margins for error (several μm) between different mask levels so that poor registration does not usually lead to chips that don't work. As optical lithography equipment improves, registration is improved and consequently, these margins for error can be reduced, thereby area of silicon is used more efficiently.

Lithography is usually carried out immediately after a layer deposition step. Examples of such layers are silicon dioxide which is normally grown on the silicon surface and aluminium which is evaporated onto the wafer. The photoresist is applied to the surface and to ensure that it adheres adequately, the surface must be dry. This can be achieved by either soft baking the wafers prior to the application of the photoresist or by using an adhesion agent such as hexamethyl-disylazine (hmde) which hydrolyses any surface moisture. The wafer is held firmly in place while a small amount of photoresist is put on the surface. The wafer is spun at several thousand r.p.m. for about 30 seconds so that the photoresist is evenly distributed across the wafer. The resulting thickness is dependent upon the proportion of solids in the resist and the spin speed. An even thickness of photoresist and good adhesion over the entire

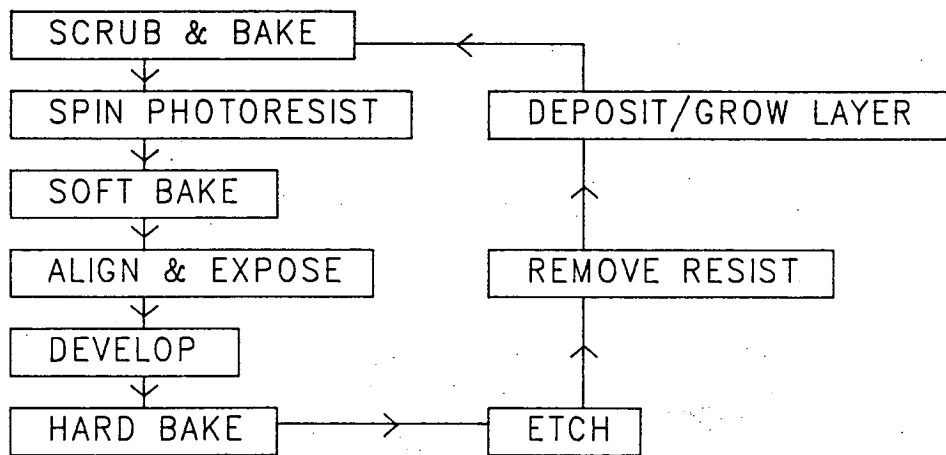


Figure 3.2.1 Outline of a typical Lithographical Sequence

wafer is essential so that the same energy will be needed to expose all points on the wafer and unexposed areas are not washed off during development. This avoids nonuniformities in device gates, areas in the depletion channels which do not receive the depletion implant or enhancement device channels which do and breaks in the interconnect. A soft bake at around 100 C (for positive resist) evaporates some of the solvents within the resist and so assists in adhesion. The wafer is now ready for exposure. The mask is aligned to the pattern already on the wafer, assuming it is not the first layer, and ultra-violet light is directed through the mask onto the wafer. For NMOS, the most important lithography areas from the point of view of the MOSFETs are the definition of the areas to receive the depletion implant and the definition of the shape of the gate in polysilicon. A certain intensity of illumination is maintained until enough energy ($30\text{-}35 \text{ mJ cm}^{-2}$ for positive resist) has entered the photoresist to allow it all to be removed from the exposed areas during development. The photoresist is developed and the wafer is washed and dried. The final step in the lithographical sequence is a hard bake (100-140 C) to increase the adhesion of remaining resist in order to help it withstand some of the etching processes which follow.

There are three pieces of equipment used for optical pattern transfer: the contact printer, the proximity printer and the projection printer.⁴⁷ The contact printer holds the mask and the wafer about $10 \mu\text{m}$ apart while they are aligned. The wafer and mask are then brought into contact and minor adjustments made to the alignment. The whole wafer is exposed and features down to $1 \mu\text{m}$ can be resolved using $1 \mu\text{m}$ of positive photoresist. This resolution is good enough for producing VLSI features but the major drawback with the contact printer is dirt which damages the mask when the wafer and mask are brought together. The resulting defect will be printed in all subsequent exposures. To avoid this problem, proximity printers do not bring the mask and wafer into contact but keep them between 10 and $25 \mu\text{m}$ apart. This saves damaging the mask, but because of diffraction, the practical resolution limit is about $3 \mu\text{m}$. Proximity printers are therefore unsuitable for defining VLSI features. The third option is projection printing where the mask is many centimetres away from the wafer. The mask can be actual chip size but more commonly is five or ten times the final chip size and the image is optically reduced onto the wafer. A single chip or small group of chips is exposed at a time, allowing very high resolution ($0.7 \mu\text{m}$) and accurate alignment across even a 6 inch diameter wafer. This fulfils the requirements for VLSI fabrication. Because each chip is aligned and exposed individually,



throughput is relatively low. Exposing a complete wafer takes several minutes compared with 20 or 30 seconds for contact and proximity printers. Projection lithography machines usually called wafer steppers are the most widely used means of patterning larger geometry wafers and smaller geometry circuits.

3.3 Ion Implantation

In the manufacture of integrated circuits, ion implanters are used to dope the silicon substrate. Implantation^{48,49} can be used for several purposes: to adjust device threshold voltages, to create source and drain areas and to help isolate active regions from one another. In the early days of wafer fabrication, doping was achieved by diffusing the impurity into the wafer at a high temperature. This resulted in the introduction of a high impurity dose which penetrated well into the substrate. Using ion implantation, where the doping atoms are charged and accelerated into the silicon, allows the dose to be controlled both in quantity and distribution. There is little lateral spreading of the high purity dopant and the process takes place at a low temperature. Ion implanters are expensive and, including the time for loading and unloading, the time taken to implant is significant, so throughput is relatively low. In VLSI, however, implantation is important for doping the wafer with a precise dose of impurity for threshold voltage adjustment and for forming shallow source and drain junctions.

An outline of the basic construction of an ion implanter is shown in figure 3.3.1. To avoid scattering of the ions in the ion beam, the whole system, from ion source through to the wafer is evacuated. Ions emerge from the ion source and those which are required are directed into the acceleration tube by the analysing magnet. Selecting ions of a particular mass and charge results in a high purity beam and ions having a precise velocity when emerging from the accelerator. This velocity determines the depth to which the ions penetrate the substrate. Therefore the impurity distribution can be precisely controlled as is required for VLSI. When the ions leave the acceleration tube, the beam is focussed and gated to remove neutrals. The ion beam is passed between horizontal and vertical plates, to which sawtooth voltages are applied to scan the beam and provide uniform implantation across the wafer.

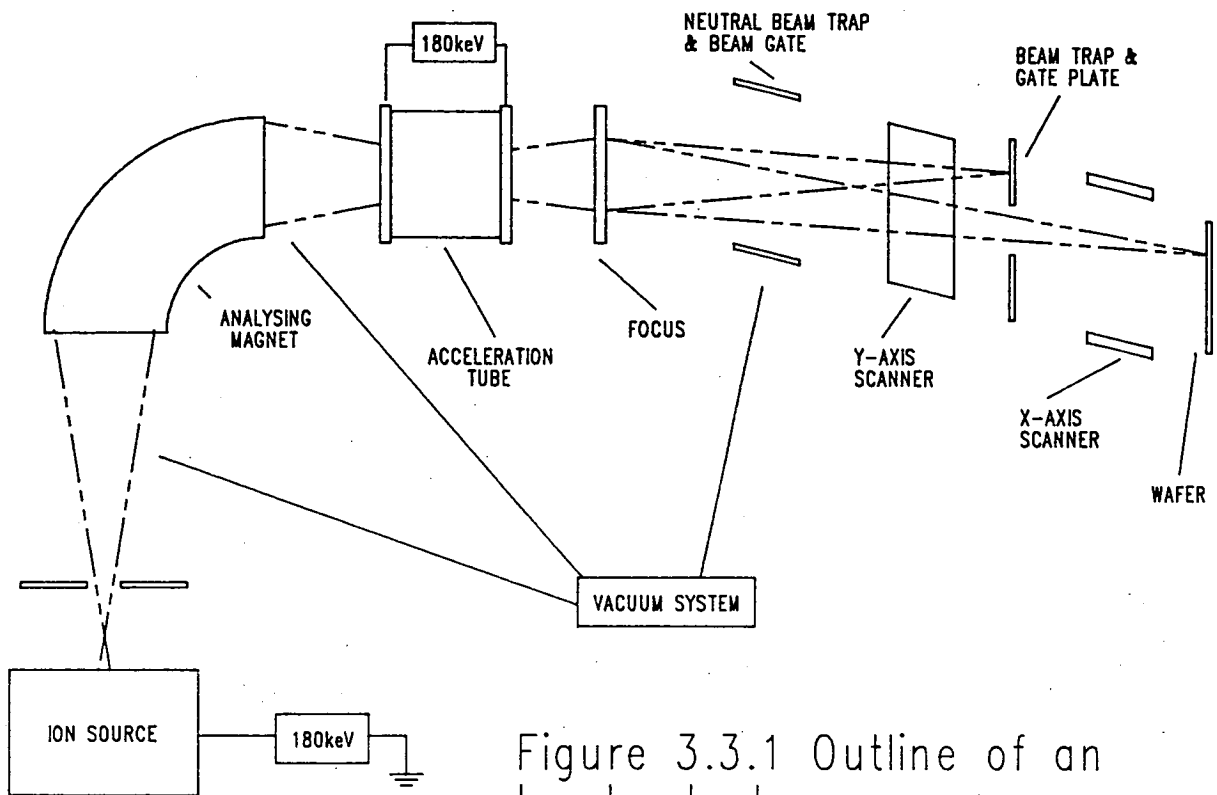


Figure 3.3.1 Outline of an Ion Implanter

3.4 Small Geometry Processing

Three areas in which the process has been changed for the fabrication of small geometry devices are the deep channel implant, SElective Polysilicon Oxidation (SEPOX) for device isolation and the thorough cleaning procedure used before high temperature processing.

When voltages are applied to the source and drain terminals, depletion regions are created around the junctions. If the voltages are high enough, the depletion regions meet and this leads to the destruction of the device by a mechanism called punchthrough. With shorter devices, the voltages needed to cause punchthrough are lower and may be close to the normal supply voltage. To ensure that punchthrough can only occur at potentials well above the operating voltage, a higher impurity dose must be implanted in the channel. By implanting a heavy dose below the surface of the silicon, breakdown can be avoided without significantly changing the threshold voltage. However, other aspects of device operation are affected. The substrate bias coefficient γ will increase due to the increase in substrate concentration in the region through which the depletion width extends. The carrier mobility parameters μ_0 and θ will be altered as will η , the shift in threshold voltage with drain bias. There will be a reduction in η because the depletion region around the drain is a smaller proportion of the channel. These are parameters which can be monitored for process control to decide whether the implant has been carried out successfully.

It has already been pointed out that increasing packing density is a means of increasing financial profit (Section 1.2). One area in which packing densities can be improved is in the isolation between active areas on the silicon wafer. The most widely used method for isolating devices is LOCOS⁴⁵ but this results in a fairly large bird's beak ($0.8\mu m$) which is the parameter Δ_w in the SPICE model. A processing method which results in a reduced bird's beak is SElective Polysilicon OXidation (SEPOX).⁵⁰ The authors claim that both stacking faults and dislocations are also reduced by using SEPOX. In SEPOX, polysilicon, which has been deposited on top of the buffer oxide, is oxidised to produce the thick field oxide. The masking nitride is etched using Reactive Ion Etching (RIE) in order to produce the fine lines which are not achievable using isotropic wet etch techniques. The nitride and polysilicon are removed from the active areas after field oxidation, leaving a small amount of polysilicon under the walls

of the thick oxide. This is converted to oxide during the growth of the thin oxide layer and the resulting bird's beak is only about $0.15\mu m$. As well as reducing the channel width reduction Δ_w , other parameters will probably be affected. If the number of stacking faults and dislocations are reduced then carrier mobilities will be improved.

The final aspect of small geometry processing to be discussed in this section is wafer cleaning. Device stability and leakage currents are both dependent upon contamination of the wafer. Reverse leakage currents are due to metal ions in the junctions and metal ions in the oxide can help invert the semiconductor surface. Early in the development of MOS technology, it was determined that most impurities are introduced during the high temperature steps in particular during gate oxidation⁷ and that good cleaning before high temperature processing can avoid significant contamination. Burkman⁵¹ in a review of cleaning procedures, categorised the materials which need to be removed as organic residue, inorganic ions and inorganic atoms. In the VLSI NMOS process used in the EMF, the RCA clean has been adopted.⁵² First of all, the wafer is placed in a solution of ammonia, hydrogen peroxide and water in proportions 1:1:5 for 10 minutes at 80 C to remove organic films and particles. After a rinse in deionised water, a dip in 10% hydrogen fluoride for 15 seconds removes the native oxide. A further rinse is then performed before another 10 minute soak at 80 C; this time in hydrochloric acid, hydrogen peroxide and water in proportions 1:1:5. Inorganic atoms and ions, in particular the notorious sodium ions, are removed during this period. Finally, after another rinse in deionised water, the wafers are washed and spun dry. This thorough cleaning, removing first organic contaminants, then the native oxide and then metallic impurities, results in a low concentration of metallic ions in the oxide, a low concentration of fixed oxide charges and a low concentration of surface states. Therefore the term including these charges in the flatband voltage is reduced and both threshold voltage and transconductance are stable. For small geometry devices, electrons in the depletion region around the drain can have sufficient energy to enter the oxide (hot carriers). If there are intermediate states (energy levels) available at the interface then this process can take place more readily and threshold voltage and transconductance change.

3.5 Process Simulation

Prediction of device structures by simulating wafer processing steps, such as ion implantation, layer deposition, etching and diffusion, is a useful tool for analysing the link between processing and device operation. Simulation programs were developed because wafer fabrication is a costly and time consuming process and so developing a process by trial and error is not a viable option in the dynamic microelectronics industry. Computer simulation allows the major process variables to be evaluated quickly and cheaply. This takes a matter of a few hours at most and helps to determine which processes are worthy of being used to fabricate devices. One program which is commonly used for process simulation is the Stanford University PProcess Engineering Models program (SUPREM).^{53,54} Three versions have been released of which SUPREM II is available on the EMF VAX computer. It provides the one dimensional structure of the silicon area undergoing the specified process. For smaller geometry devices, the lateral diffusion of impurities at the edges of the source and drain regions and at the edge of the field oxide becomes more important. Therefore two-dimensional process simulation becomes necessary. The OSIRIS program,⁵⁵ developed by the Laboratoire des Composants a Semiconductors in Grenoble in France in 1985, is a two-dimensional simulation program available at the EMF.

For the EMF NMOS process, SUPREM has been used to predict the impurity profiles of the enhancement device at several stages of the fabrication⁵⁶ and of the depletion device on completion of processing. The output of OSIRIS showing the two-dimensional channel and junction profile of the VLSI NMOS process in the EMF has also been found and one-dimensional profiles of the channel and of the junction have been found using SUPREM to back up the OSIRIS results. The OSIRIS run files are in APPENDIX B and the results are presented and described below.

For all the SUPREM simulations, the substrate was set up as being $\langle 100 \rangle$ orientated with a Boron concentration of $8 \times 10^{14} \text{ atoms cm}^{-3}$. The concentrations are calculated to a depth of $2 \mu\text{m}$ at intervals of $0.01 \mu\text{m}$. For the EMF NMOS enhancement device, the SUPREM prediction of the total concentration (which is all Boron) immediately after the threshold adjust implant is given in figure 3.5.1. The maximum concentration is $3.57 \times 10^{16} \text{ atoms cm}^{-3}$ at a depth of $0.06 \mu\text{m}$. At this point, before any subsequent high temperature steps which would cause the Boron to

Log Conc. (cm-3)

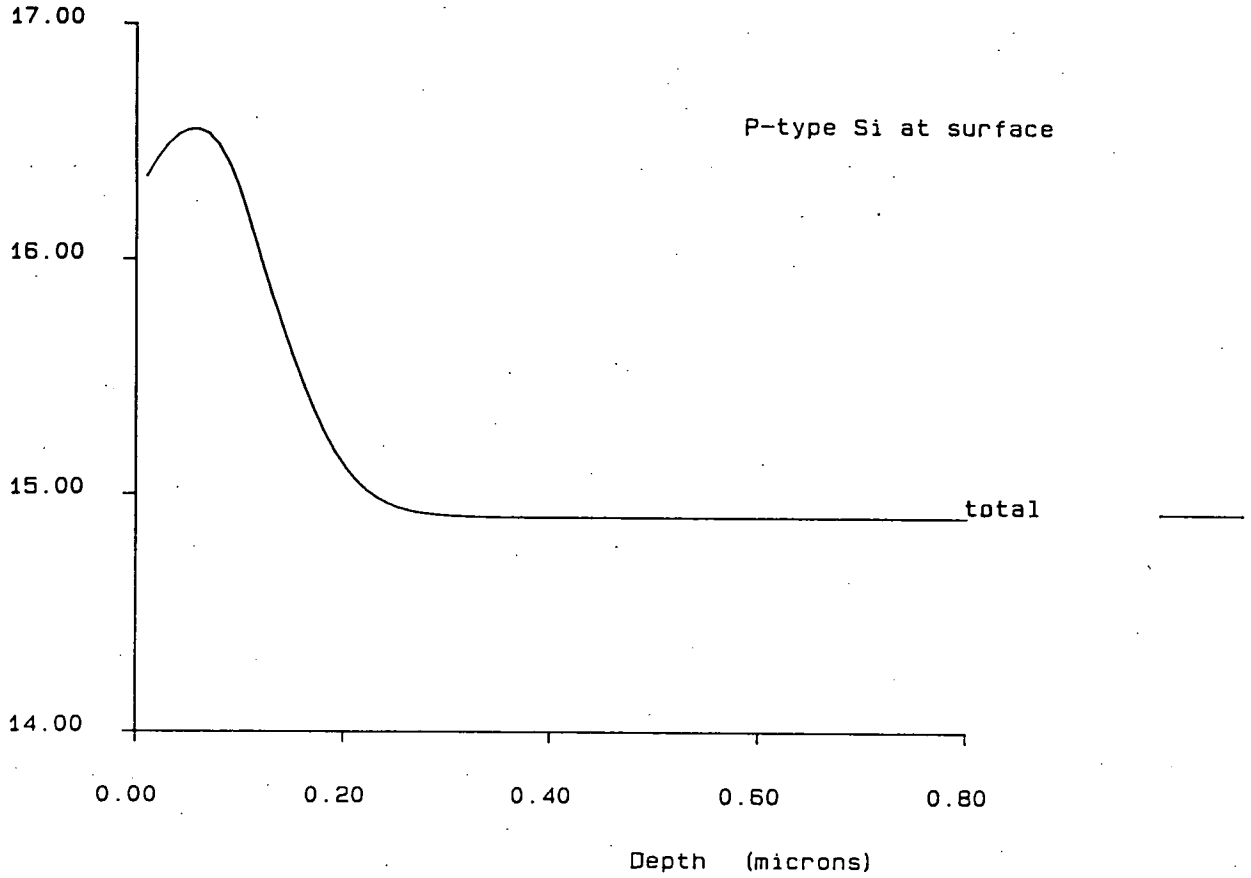
SUPREM DOPING PROFILE

12: 32: 34

70Oct86

After 4E11 40keV Boron Implant

Figure 3.5.1



Log Conc. (cm-3)

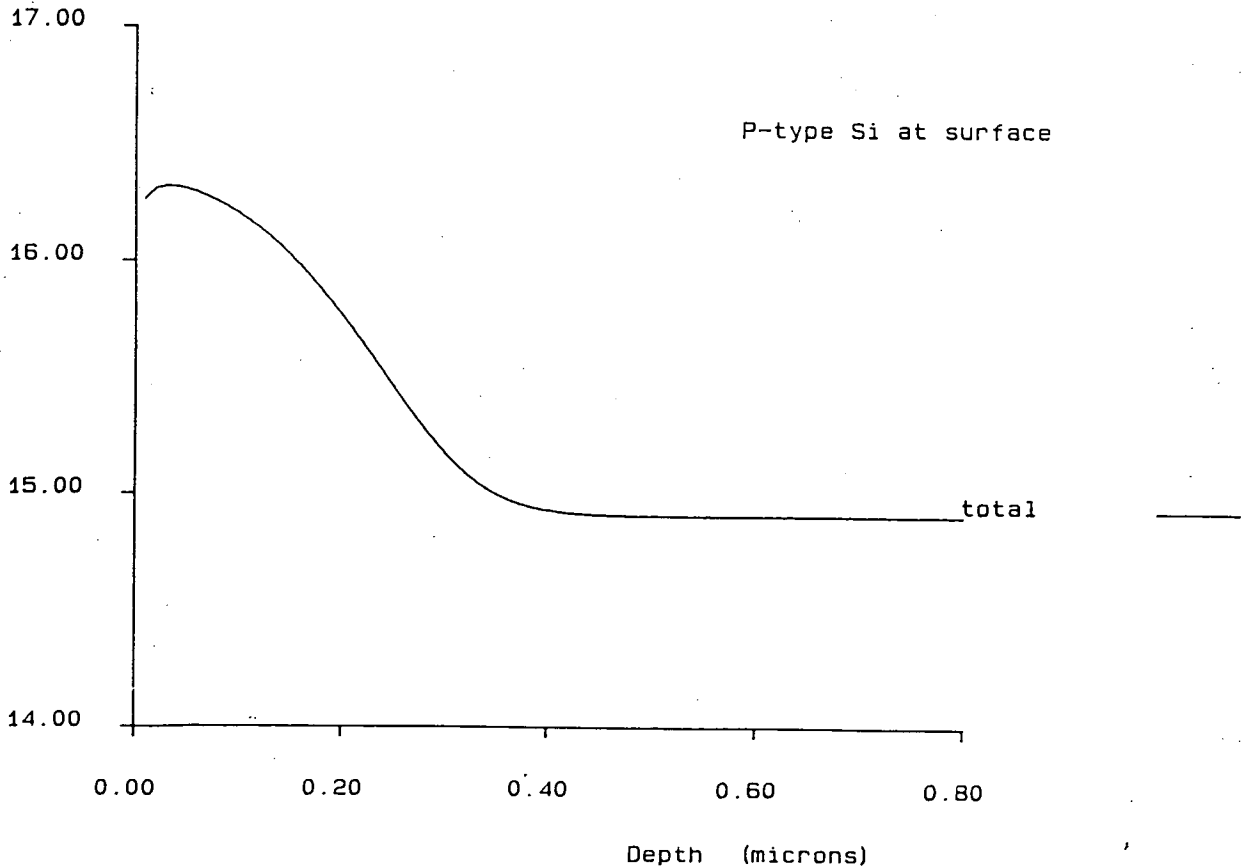
SUPREM DOPING PROFILE

13: 40: 10

70Oct86

Before Pyro Reflow at 1050 C

Figure 3.5.2



diffuse have taken place, there is a sharp peak in Boron concentration within the top 0.2 μm of the substrate. Figure 3.5.2 shows how the Boron has diffused during some of the high temperature steps, but does not take into account the two 20 minute pyrolytic oxide reflows at 1050 C. There is still a peak concentration just below the silicon surface. As a result of the very high temperature reflows, the impurity becomes a maximum at the surface at $1.45 \times 10^{16} \text{ atoms cm}^{-3}$ (see figure 3.5.3) and gradually decreases to the substrate concentration at a depth of around 0.6 μm . The profile shows that the impurity concentration varies and is not a constant as assumed by most MOSFET models. If the surface concentration is $1.45 \times 10^{16} \text{ atoms cm}^{-3}$, and the flatband voltage is assumed to be -1.0V (see Section 3.1) then using

$$V_{to} = V_{fb} + 2\phi_b + \frac{(2q \epsilon_{si} N_{sub})^{\frac{1}{2}}}{C_{ox}} (2\phi_b)^{\frac{1}{2}} \quad 3.5.1$$

and

$$2\phi_b = \frac{2kT}{q} \ln \left(\frac{N_{sub}}{n_i} \right) \quad 3.5.2$$

results in $V_{to} = 1.16V$. For a concentration of $5 \times 10^{15} \text{ atoms cm}^{-3}$, which would be closer to the average concentration across the depletion region, $V_{to} = 0.47V$.

The depletion devices have the added complication of two different impurities in the channel region. SUPREM allows different impurity types to be output separately or a combined net concentration can be provided as in figure 3.5.4. This is the final channel profile of the NMOS depletion device and it can be seen that the silicon is n-type down to about 0.3 μm with a peak concentration of $7.70 \times 10^{16} \text{ atoms cm}^{-3}$ at a depth of 0.14 μm .

The OSIRIS program was used to calculate the two-dimensional profile of the source/drain junction and half of the channel of a 1.5 μm device fabricated using the EMF VLSI NMOS process (figure 3.5.5). The x-direction covers 1.25 μm of which the left 0.5 μm is a region of junction and the other 0.75 μm is under 250 A of gate oxide representing part of the device channel. The concentrations are calculated to a depth of 1 μm using a grid of 80 points in the x-direction and 40 points in the y-direction. The output of OSIRIS, which shows individual impurity concentrations, requires some explanation. If a vertical section is taken at the right of the simulated section, it is found that the surface concentration increases as contours 5b, 4b and 3b

Log Conc. (cm⁻³)

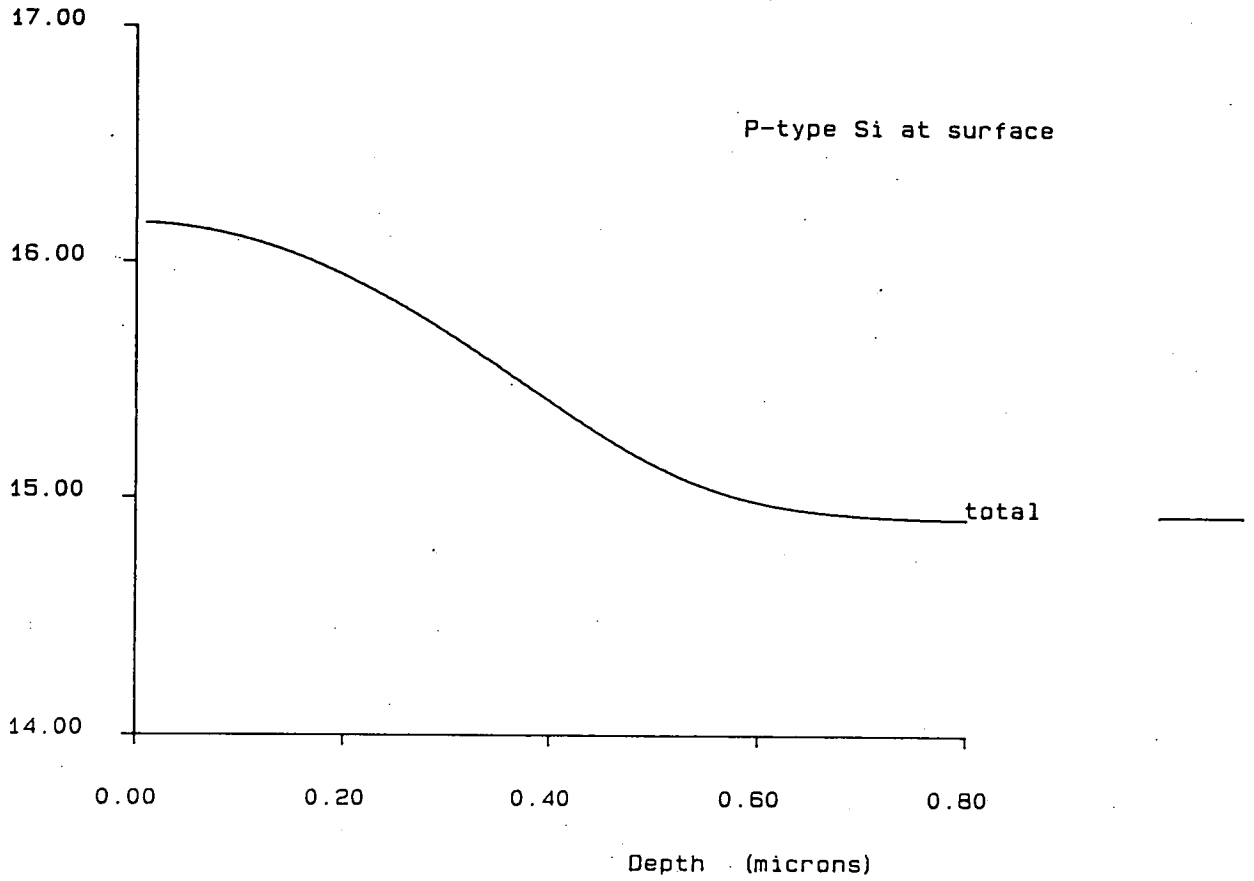
SUPREM DOPING PROFILE

14:02:24

70ct86

Final Channel Profile for Enhancement Device

Figure 3.5.3



Log Conc. (cm⁻³)

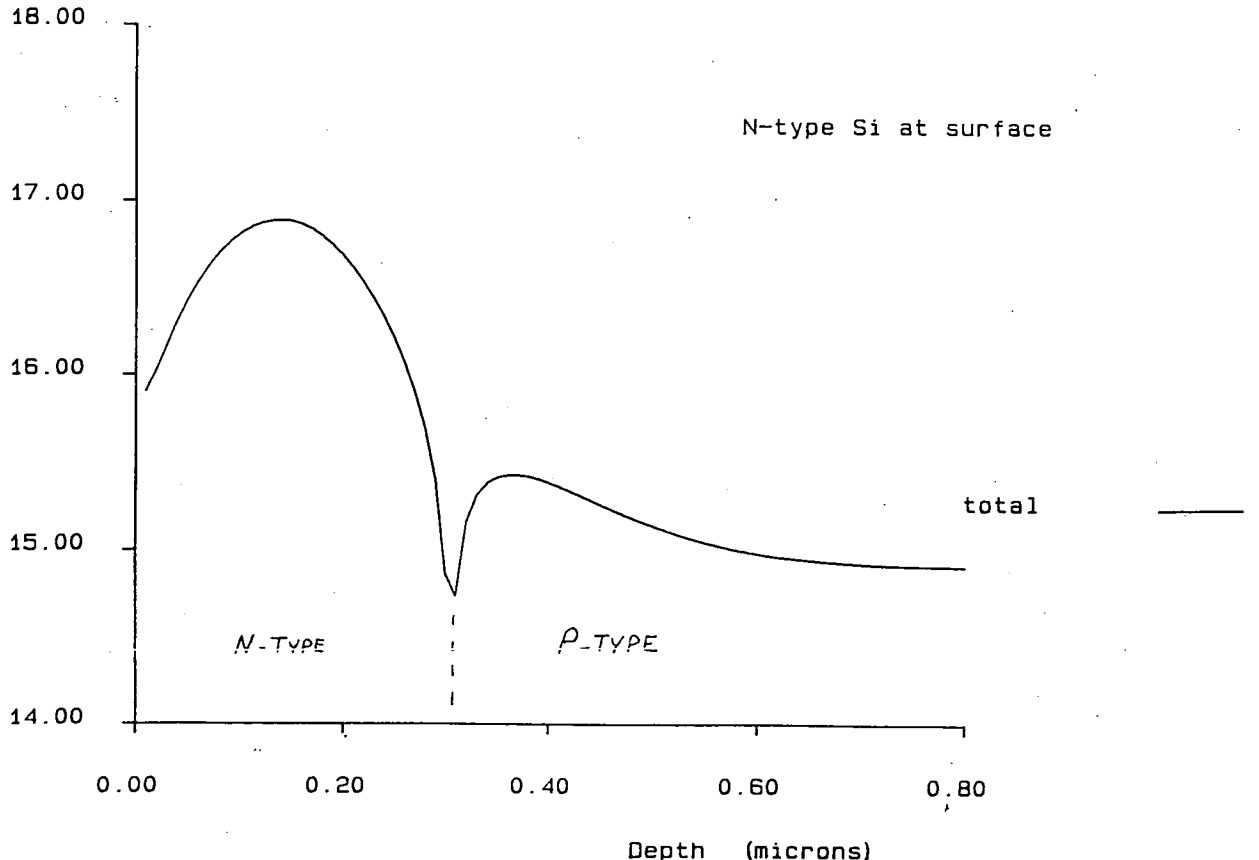
SUPREM DOPING PROFILE

14:24:57

70ct86

Final Channel Profile for Depletion Device

Figure 3.5.4



VLSI 2-D Channel Profile

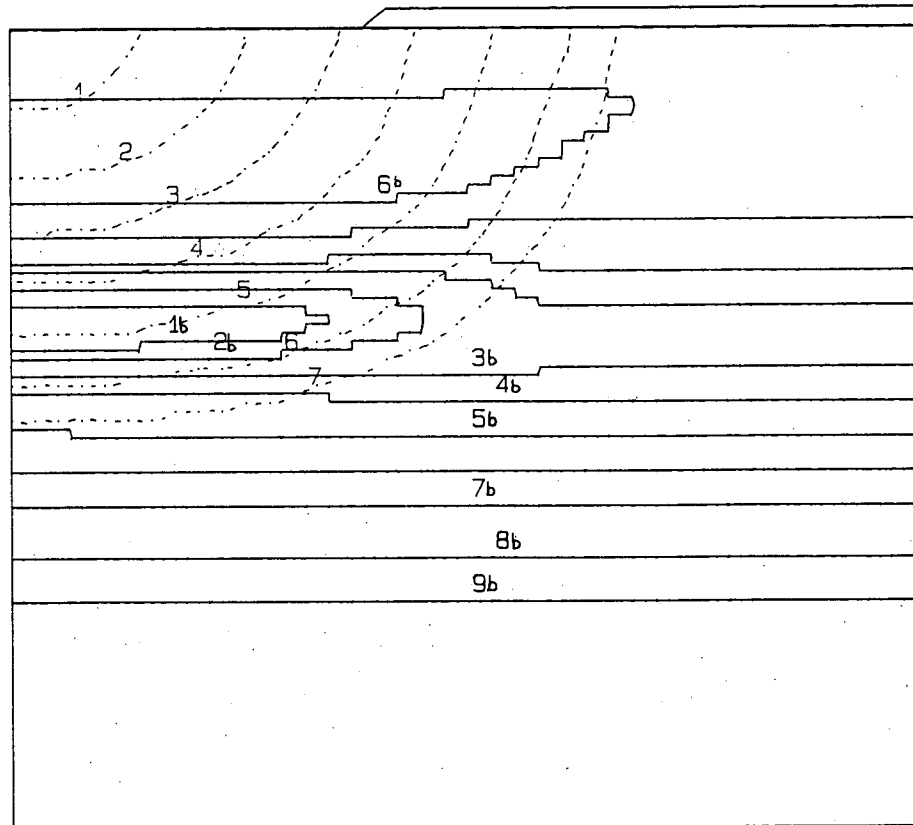
Figure 3.5.5

/ Stage : 17

1 centimeter is :

Horizontal : 735 angstroms.

Vertical : 676 angstroms.



- BORON

1b	0.63E 17
2b	0.60E 17
3b	0.55E 17
4b	0.50E 17
5b	0.40E 17
6b	0.30E 17
7b	0.20E 17
8b	0.10E 17
9b	0.50E 16

... ARSENIC

1	0.30E 21
2	0.25E 21
3	0.20E 21
4	0.15E 21
5	0.10E 21
6	0.50E 20
7	0.10E 20

SUBSTRATE : 0.70E 15

are crossed to a peak concentration of between 5.5×10^{16} atoms cm^{-3} and 6.0×10^{16} atoms cm^{-3} at a depth of $0.4 \mu m$. The boron concentration then tails off into the substrate. On the left side of the profile is a junction region and a vertical section taken in this region shows a very high concentration of Arsenic (above 3×10^{20} atoms cm^{-3}) which decreases very sharply beyond a depth of $0.4 \mu m$ to form a shallow junction. Although the Boron has been implanted in the same way as for the channel, the influence of the arsenic has meant that it has diffused differently. There is a lower surface concentration and a peak concentration of just over 6.3×10^{16} atoms cm^{-3} at a depth of $0.4 \mu m$. The lateral diffusion of the arsenic shows that the L_d parameter is going to be approximately $0.3 \mu m$. The results are backed up by two one-dimensional SUPREM profiles of the junction and channel regions (figures 3.5.6 and 3.5.7).

These simulation tools are an integral part of process development. They help to predict threshold voltages, actual device lengths and junction depths. In addition, they also provide a valuable link between device fabrication and operation and allow complex interactions between many variables to be presented in a simple and direct way.

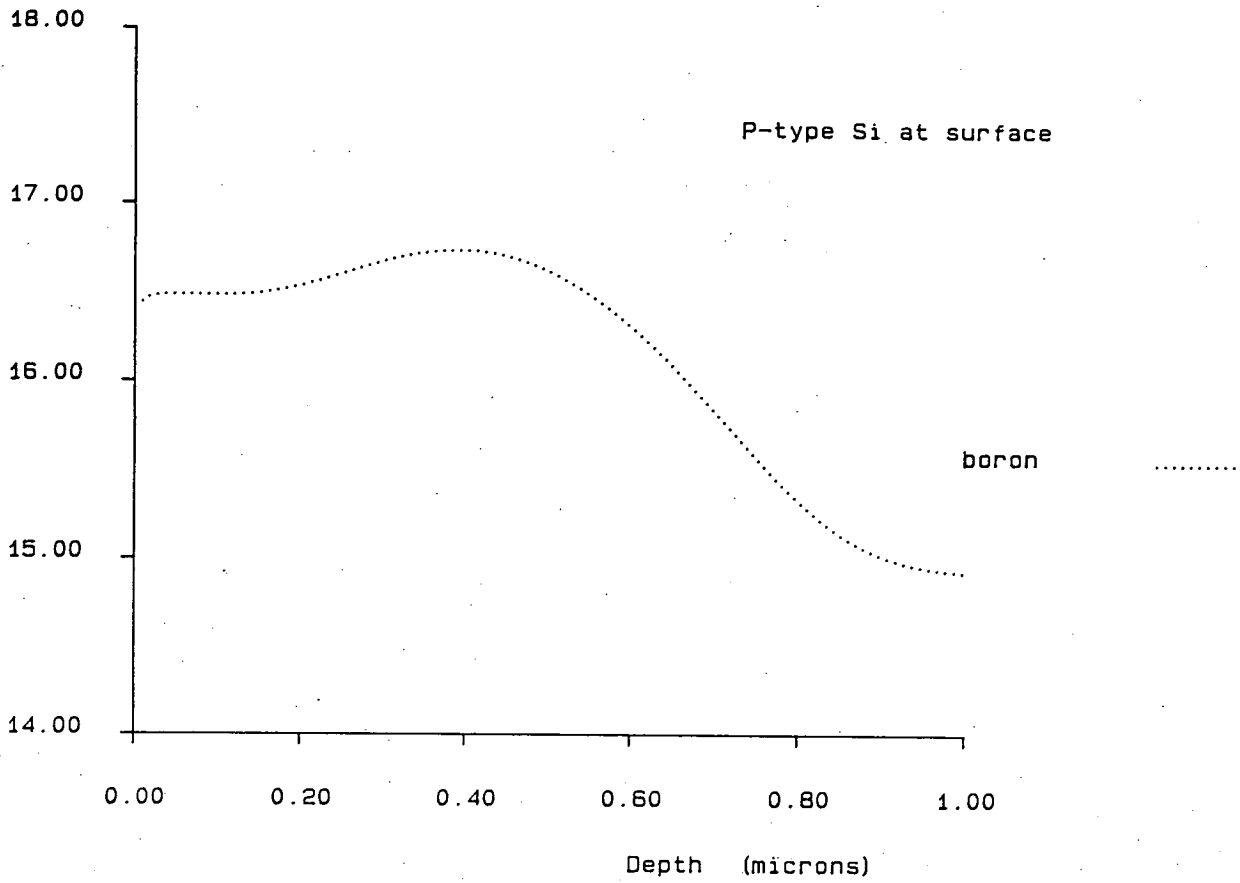
Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

20:06:23 15Oct88

Final Channel Profile

Figure 3.5.6



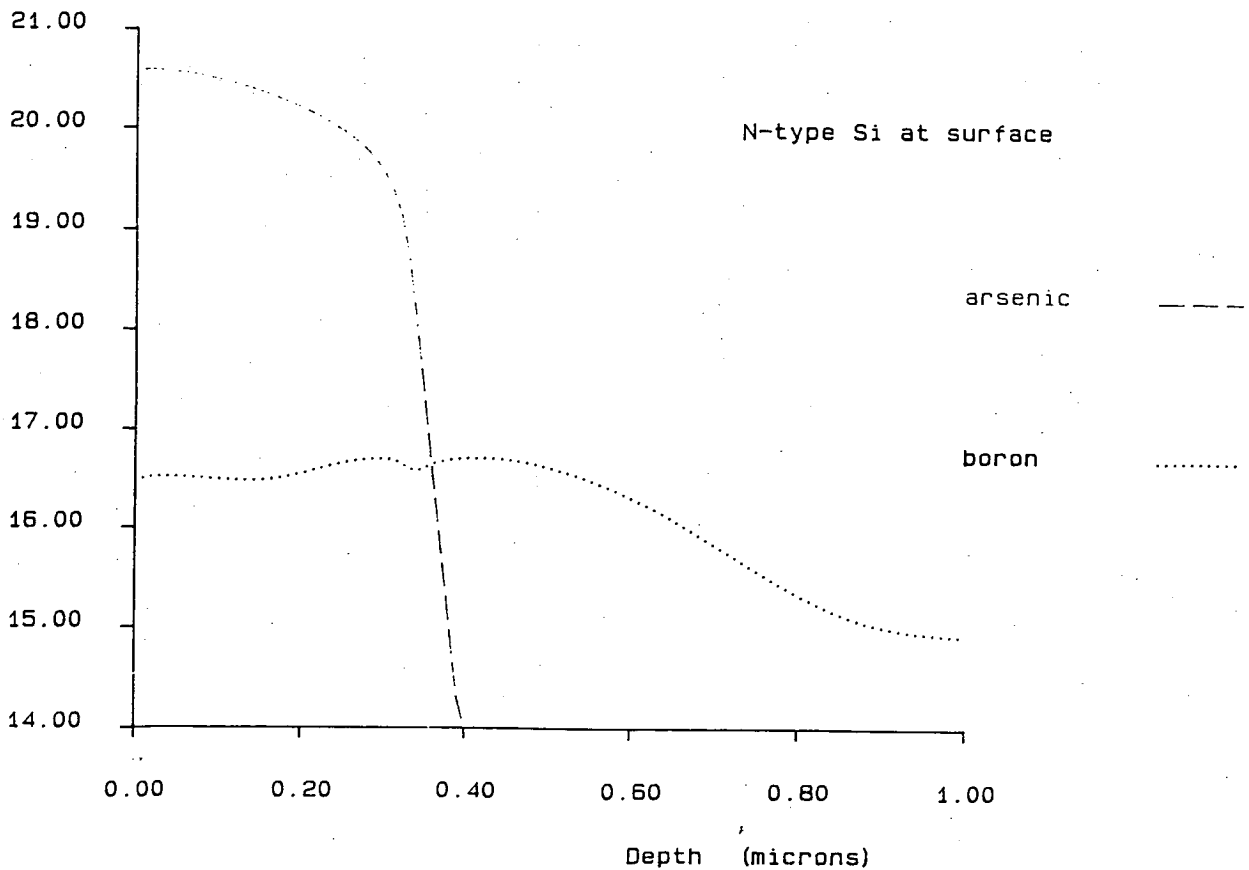
Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

20:16:40 15Oct86

Final Junction Profile

Figure 3.5.7



Chapter 4 : Parameter Extraction

4.1 Different Extraction Philosophies

4.1.1 Numerical Optimisation or Physical Parameter Extraction?

All device models used in circuit simulation packages require certain input parameters in order to link the mathematical equations to particular devices. These are physical quantities such as oxide thickness and junction depth and values quantifying the electrical operation of the devices e.g. threshold voltage, substrate bias coefficient and mobility degradation coefficient. For a reasonably complex model capable of accurately simulating small geometry devices, the extraction process is complicated and two essentially different approaches have been used.¹⁵ Traditionally, parameters have been extracted one by one from particular device characteristics. The other approach is to use numerical optimisation where a simulated characteristic is made to fit as closely as possible to a measured characteristic by systematically adjusting the values of the parameters.

Various people have obtained parameters using the traditional one by one method. The Berkley memo "The Simulation of MOS using SPICE 2"²² gives some guidelines on how to go about this. Other authors have written about the extraction of particular parameters e.g. Takacs et al.⁵⁷ and Moll et al.⁵⁸ However, as device geometries are reduced and the models are extended to include short and narrow channel factors, this method becomes difficult and a detailed, thorough description of a complete extraction process is not readily available. Wright³⁹ does describe a complete extraction process but then the parameters obtained are fed into a statistical computer package to obtain a good fit. Parameters extracted physically by a particular technique from a particular characteristic have a precise definition and so not only combine to simulate device operation but can also be used for process control. Usually, these extraction techniques only apply to one particular model. However, most models have some similar parameters (e.g. threshold voltage and mobility modulation coefficient) and so adapting these techniques is often possible without substantial effort.

As the device models became more complicated, numerical optimisation was developed for parameter extraction. Commercial parameter extraction software

packages, which have only recently become available, all use numerical optimisation. Measured values are input and after a Jacobian iteration or similar process, the parameters are derived. Some programs optimise all the parameters for all bias voltages at once e.g. SUXES produced by Stanford University while others, e.g. TECAP¹⁷ (Transistor Electrical Characterisation and Analysis Program) written by Hewlett-Packard are capable of optimising particular parameters in specific regions of operation separately. Other packages include MOSFIT⁵⁹ marketed by MOSAID and SIMPAR⁶⁰ developed at the ESAT laboratories in Belgium. The meaning of the parameters is not taken into account during the extraction and so since different parameters have similar effects on the characteristics, different sets of parameters can result from the same measurements. Hence, although the parameters obtained from numerical optimisation can be used to accurately simulate device operation, they cannot be used for process control. A major advantage of numerical optimisation is that it can easily be used for other models or even other types of semiconductor device.

4.1.2 Parameter Extraction by Numerical Optimisation:TECAP

The TECAP program written by Hewlett-Packard uses numerical optimisation to extract parameters.⁶¹ The program runs on an HP series 200 computer in the HP Pascal Operating Environment. Two megabytes of RAM are required and a disc drive, a plotter and a printer allow information to be read and stored and results to be displayed. An HP4145A Semiconductor Parameter Analyser is used to make d.c. measurements and capacitance measurements can be performed using an HP4280A 1MHz capacitance meter. Other measuring instruments, a full-Kelvin relay matrix and various automatic wafer probers are all supported by the software. All the computer peripherals and measuring instruments are controlled via the standard, parallel HP interface bus. A block diagram of a suitable hardware configuration which is in use at the EMF is shown in figure 4.1.2.1.

The software is specifically designed to be user friendly. Selections are made from a main menu and this leads to a submenu from which a specific task is chosen. The options in the main menu are listed in figure 4.1.2.2. Basically the program consists of four stages: Setup, Measure, Extract and Simulate. The Setup stage includes choosing a model, defining the device type and size, setting up the voltage ranges for the measurement and making connections in the switching matrix. In most

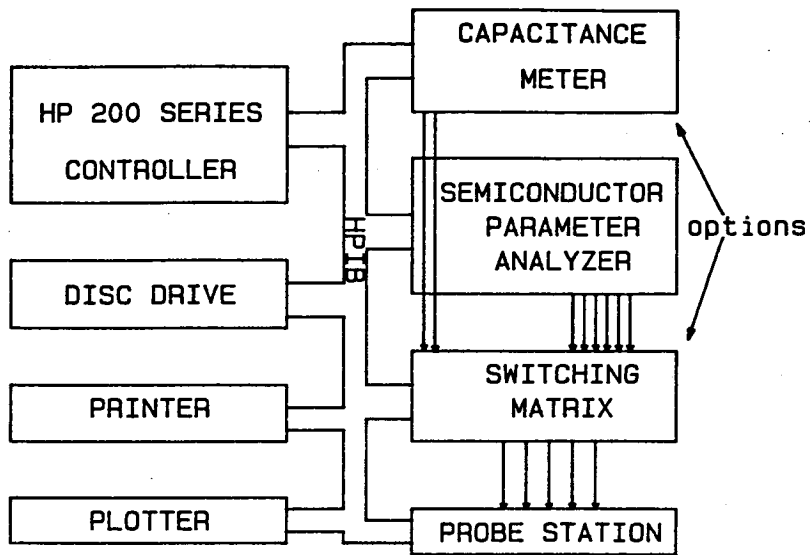


Figure 4.1.2.1 Block Diagram of
Hardware Configuration for TECAP

- | | |
|----------------|-------------------|
| D) Device Data | O) Output Control |
| C) Connections | P) Plot Control |
| U) Use Setup | F) Filer |
| M) Measure | B) Build Setup |
| E) Extract | I) Input Sequence |
| S) Simulate | Q) Use Sequence |
| A) Auxiliary | |

Figure 4.1.2.2 Main Menu Options in TECAP

cases, this is done by altering entries in a table.

After the measurements have been made, the parameters to be numerically optimised from this particular characteristic are chosen. A least-squares fit, using the Levenberg-Marquardt algorithm, is carried out and the parameter values which give the best possible agreement between measured and simulated characteristics are found. However, the user must have a good knowledge of the transistor model in order to know which parameters to extract from which characteristics, what range of currents to use and what are realistic limits to put on parameters. These limits avoid the optimiser homing in on an unrealistic solution.

An example of linear region extraction is illustrated in figures 4.1.2.3 and 4.1.2.4. The curve was measured when $V_d = .1V$. Figure 4.1.2.3 shows the comparison between the measured characteristic and one simulated with default initial parameters. In this case, the currents are of the same order but that is pure coincidence. After optimising V_{to} , μ_o , and θ , a good fit is obtained as shown in figure 4.1.2.4. If the limits on a parameter are too tight, or if the range of currents over which the extraction takes place is wrong (for instance some measurement points in subthreshold) then a result like figure 4.1.2.5 might be obtained. In this case, the mobility μ_o was unusually low and hence outside the limits set. In order to minimise the error between measured and simulated characteristics, very high values are calculated for V_{to} and θ .

There are two reasons why these parameters, although leading to an accurate simulation of device operation, cannot be considered to be physically realistic values. Firstly the maximum mobility μ_o is dependent upon the actual aspect ratio of the device, which is in turn dependent upon the default values of the length and width parameters L_d and Δ_w . Secondly the user has to decide which range of currents to use for the extraction by visually examining the measured characteristic. Measured values which are from the subthreshold region will distort the parameters but ideally the lower end of the linear region should be included.

In practice, it proved difficult trying to derive L_d and Δ_w . Very frequently, the outcome of trying to numerically optimise L_d or Δ_w is an unreasonably high value. However TECAP does yield a set of parameters which accurately simulate the device as is demonstrated in figure 4.1.2.6. This is for a $5\mu\text{m} \times 5\mu\text{m}$ NMOS device with gate

voltages of 1,2,3,4 and 5V.

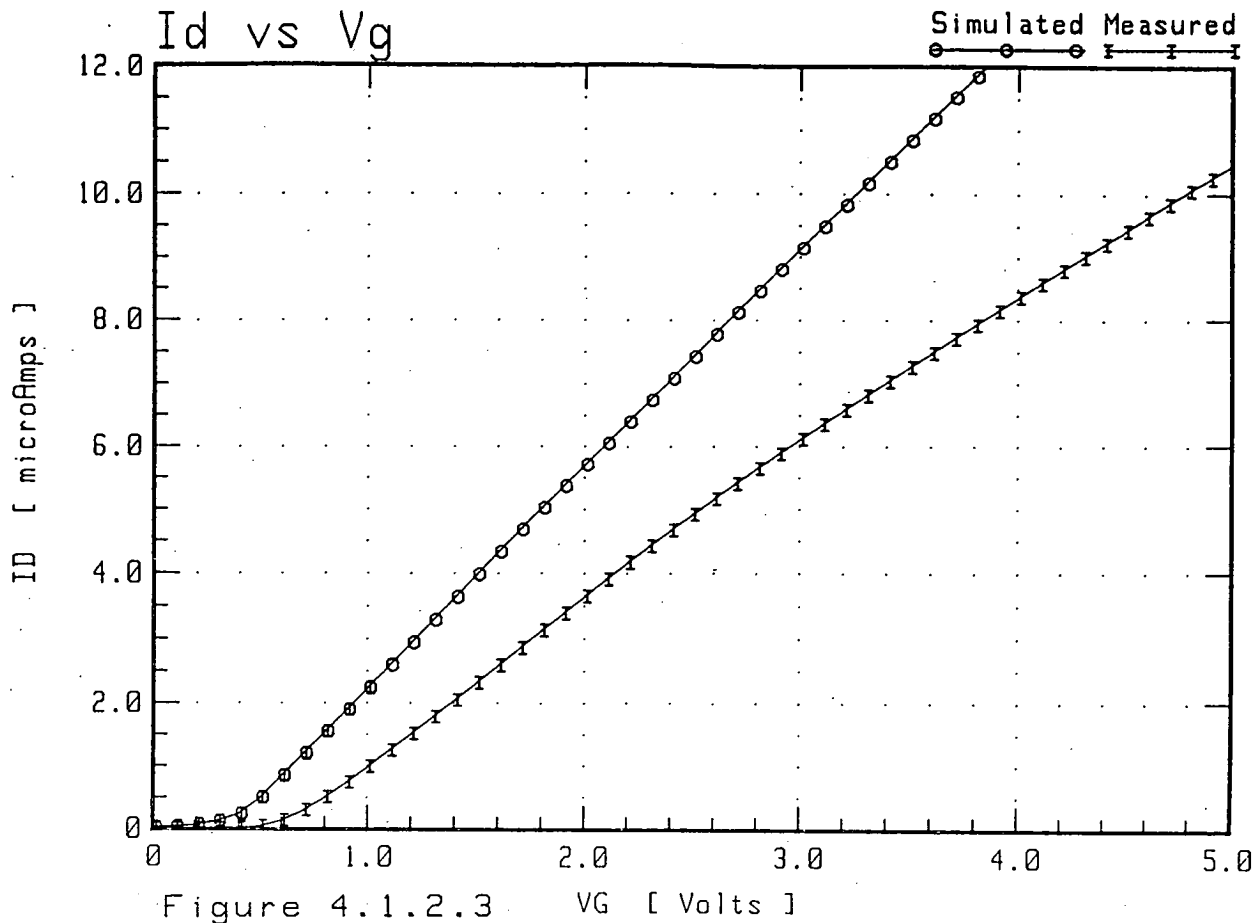


Figure 4.1.2.3 VG [Volts]

$V_B = 0.000$ V

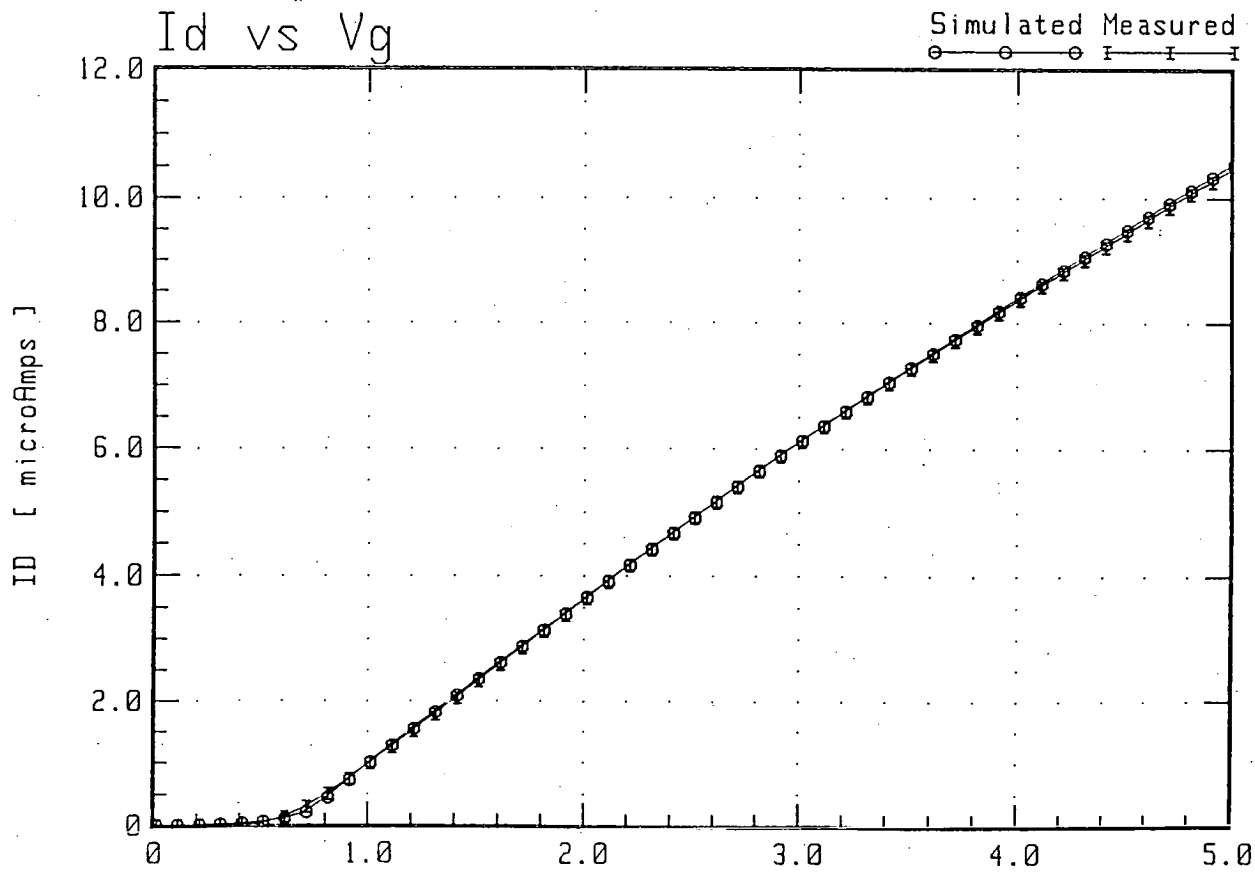


Figure 4.1.2.4 VG [Volts]

$V_B = 0.000$ V

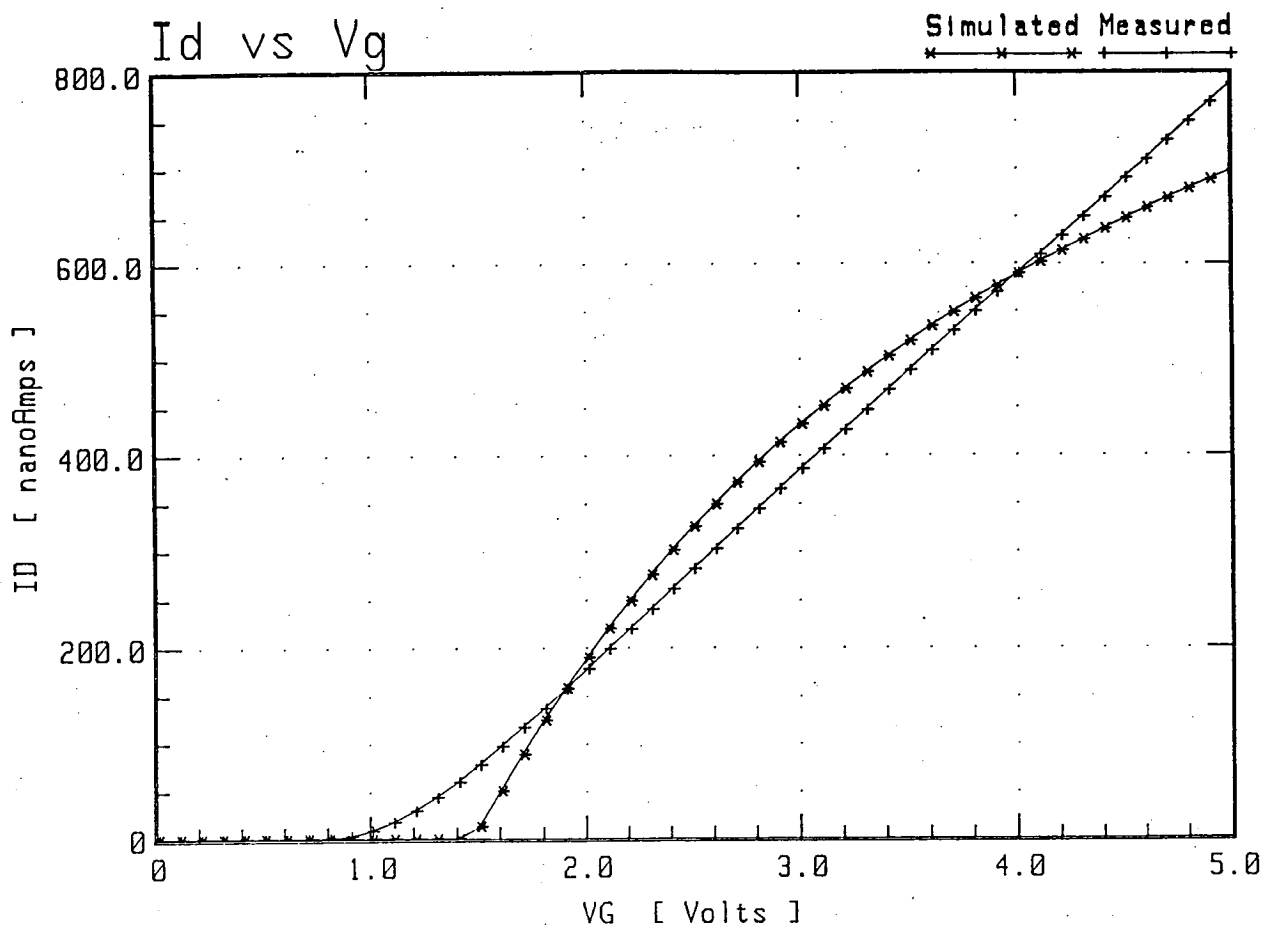
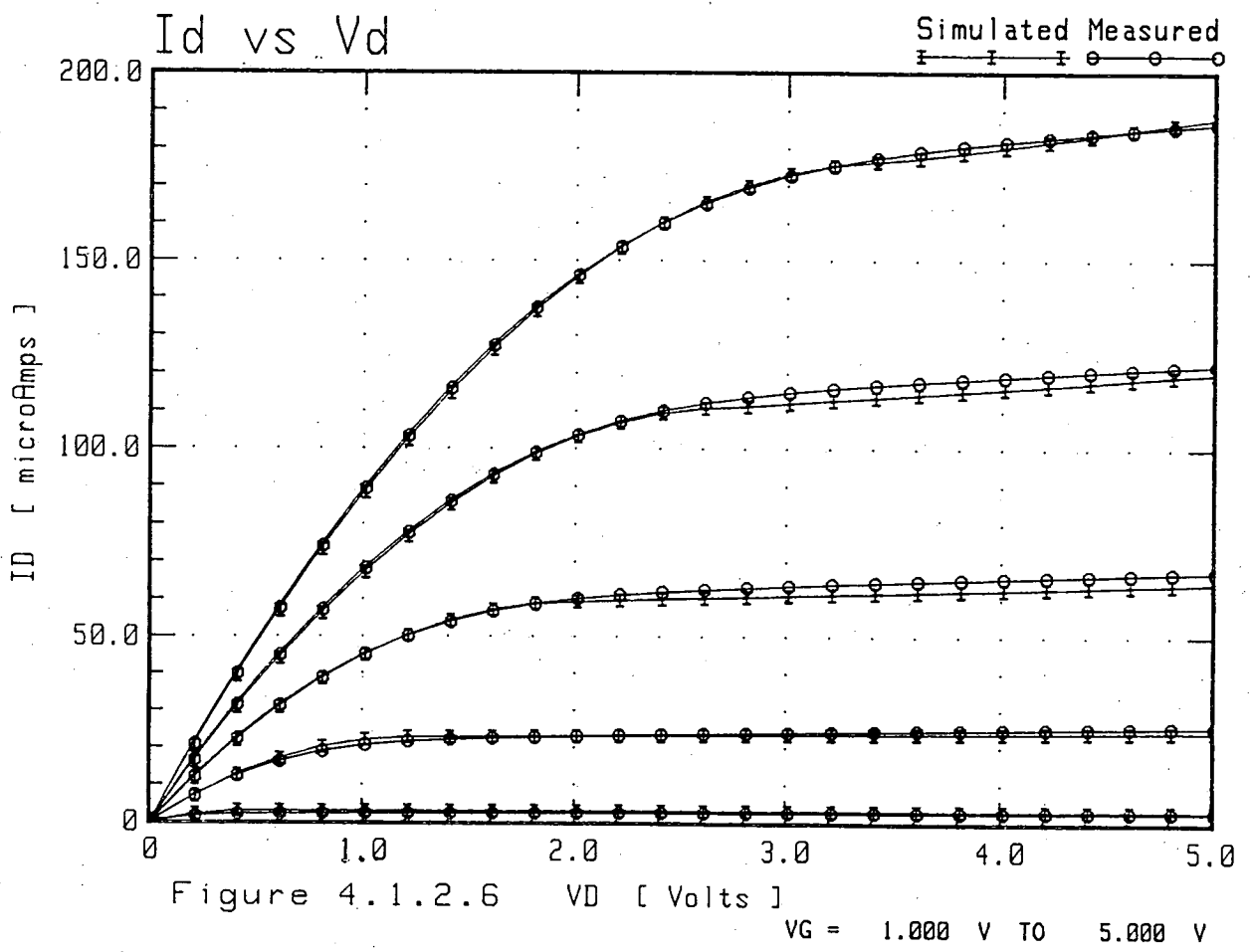


Figure 4.1.2.5

VB = 8.000 V

Final optimized parameters	Initial values
U0 = 100.0	500.0
VTO = 1.409	1.000
THETA = 331.0 m	30.00 m
Final MAX error = 21.55 %	Final RMS error = 9.60 %



4.1.3 The Physical Parameter Extraction Program:PARAMEX

The alternative to parameter extraction by numerical optimisation, which has just been discussed is to use physical parameter extraction. As mentioned in 4.1.1, this entails extracting each parameter according to a particular definition. The parameters are extracted one by one in a specific sequence from the same set of measurements each time. This allows the parameters which result to be compared with others extracted from other die and other wafers in order to provide comparison data which may be used for process control. If these parameters are also capable of accurately simulating devices then they can be used for circuit simulation as well as process control. In this way, they provide a valuable link between circuit design and fabrication. No commercially available software is able to fulfil this link.

In the EMF a computer program called PARAMEX^{*} has been written. The program implements the physical parameter extraction techniques described in Section 4.2. It runs on any HP series 200 controller using the BASIC 3.0 operating system. This advanced BASIC allows easy control of measuring instruments using the HPIB and reasonably fast program execution. The interactive nature of the language enables the user to interrupt the program, interrogate the computer as to the values of various variables, amend any variables as necessary, instruct instruments and then continue. In the experimental environment, this flexibility is most useful. Since the MOSFET is voltage driven, the measuring instruments include voltage sources and ammeters in order to measure d.c. characteristics. The oxide thickness can be found by measuring the capacitance of a gate oxide capacitor in accumulation using a capacitance meter. Information on the hardware used in the EMF is contained in the introduction of the Userguide which is included in APPENDIX C. Currently versions of the software have been supplied to Motorola Ltd (East Kilbride), Inmos (Newport), Jet Propulsion Labs (Pasadena, U.S.A.) and Rutherford Appleton Laboratories (Abingdon). The version supplied to Motorola included a section to yield parameters for their own MTIME model.

Basically the program is run in three stages: measure, extract and simulate. In the first part, various sets of device characteristics are measured and these can be stored on disc if desired. The second section extracts the parameters from the particular characteristics measured in the first section and the program then has the

* PARAMEX was written entirely by the author of this thesis.

capability to measure and simulate device characteristics in order to test the accuracy of the parameters. As well as a visual comparison, several different error figures can be listed. More precise details on how to use PARAMEX are given in the Userguide and the extraction algorithms are described in Section 4.2.

Parametric test is currently carried out at the end of the wafer fabrication process to determine whether or not the process is within specification. These parameters are measured on drop-in test die and an assessment is then made as to whether the wafer will yield working circuits. Typically sheet resistances, oxide capacitances, contact resistances and threshold voltages are monitored. Usually only a value well out of specification indicates a non-working wafer and the precise effect on the operation of circuits, in terms of output drive currents and speed of switching is unknown. By routinely measuring parameters for the SPICE model, they can be used for both process control and circuit simulation, since they physically represent different aspects of device operation and can be used in the model to accurately simulate devices. This link becomes increasingly important when optimum performance is sought and as device geometries are scaled down, so that smaller process variations have larger effects on device characteristics.^{62,63} A high speed version of PARAMEX has been developed which minimises the number of measurements and calculations which are required in order to obtain a complete set of parameters. This allows routine SPICE parameter measurement to be carried out in a production environment.

The SPICE models described in Chapter 2 use 15 d.c. parameters. These parameters which are listed below, do not all apply in the level 1 model. Those which apply to level 3 only are indicated by (3).

<i>Type</i>	N- or P-channel (1/-1)		
<i>Dep</i>	Enhancement or Depletion (0/1)		
t_{ox}	Oxide Thickness	m	
x_j	Junction Depth	m	(3)
N_{fs}	Fast State Density	m^{-2}	(3)
V_{to}	Threshold Voltage	V	
γ	Back Bias Coefficient	$V^{\frac{1}{2}}$	
L_d	Diffusion Length	m	
Δ_w	Width Reduction	m	
μ_o	Maximum Carrier Mobility	$m^2 V_s^{-1}$	
v_{max}	Maximum Carrier Velocity	$m s^{-1}$	(3)
η	Drain Feedback Coefficient		(3)
δ	Width Effect on Threshold		(3)
κ	Saturation Slope Coefficient		(3)

The first two parameters define the normal range and mode of operation of the device. These are used to determine the correct voltages to be applied in order to measure the electrical parameters. *Type* is 1 for n-channel and -1 for p-channel and *Dep* is 0 for an enhancement and 1 for a depletion device. These parameters are not used as input for the SPICE program.

Two other parameters are not calculated from measured current against voltage device characteristics. The gate oxide thickness per unit area, t_{ox} can be found from a measurement of the capacitance of an MOS capacitor in accumulation.

$$t_{ox} = \frac{\epsilon_{ox} Area}{C_{ox}} \quad 4.1.3.1$$

Alternatively it may be estimated from process specifications or physically measured in process. For level 3, the junction depth, x_j can be measured by profiling or estimated from process simulation.

4.2 SPICE 2 Parameter Extraction

4.2.1 Threshold Voltage , V_{to}

The first electrical parameter to be extracted, for both the level 1 and level 3 models, is the threshold voltage, V_{to} . To a first order, the equation for current in the linear region of operation is

$$I_d = \text{Beta} \left[V_g - V_{th} - \frac{V_d}{2} \right] V_d \quad 4.2.1.1$$

from the theory in section 2.3 based on Ihantola and Moll.³² Note that here *Beta* has been used to denote the gain of the MOSFET whereas the symbol β has previously been defined as $\frac{q}{k T}$. The Greek letter β is commonly used for both quantities. It is hoped that by distinguishing between β and *Beta*, confusion will be avoided. Then

$$\frac{I_d}{\text{Beta} V_d} = V_g - V_{th} - \frac{V_d}{2} \quad 4.2.1.2$$

On a $V_g : I_d$ graph, the intercept on the V_g -axis i.e. when $I_d = 0$ is $V_{th} + \frac{V_d}{2}$. For level 3, the exact expression for current is

$$I_d = \text{Beta} \left[V_g - V_{th} - \frac{(1 + F_b)}{2} V_d \right] V_d \quad 4.2.1.3$$

The parameters contributing to F_b , the body factor, have not been extracted at this stage. An average value for F_b is -0.008 and so equation 4.2.1.1 is assumed and the parameter V_{to} is identical in levels 1 and 3.

The device must be biased so that the gate voltage is both above threshold and above threshold plus drain voltage to ensure operation in the linear region. Drain voltage must be kept very low so that the depletion region around the drain is small and thus the appropriate term in the threshold voltage expression can be neglected. Having measured the $V_g : I_d$ characteristic (figure 4.2.1.1) under the above conditions, the next step is to pick out the turn-on point. A common method is to search for the two adjacent points between which there is greatest slope. With enhancement devices, the mobility of carriers and hence the gain of the transistor is greatest just above threshold, and carrier mobility is degraded as the transverse electric field increases. Unfortunately this doesn't work for depletion devices, where mobility continues to increase as the gate

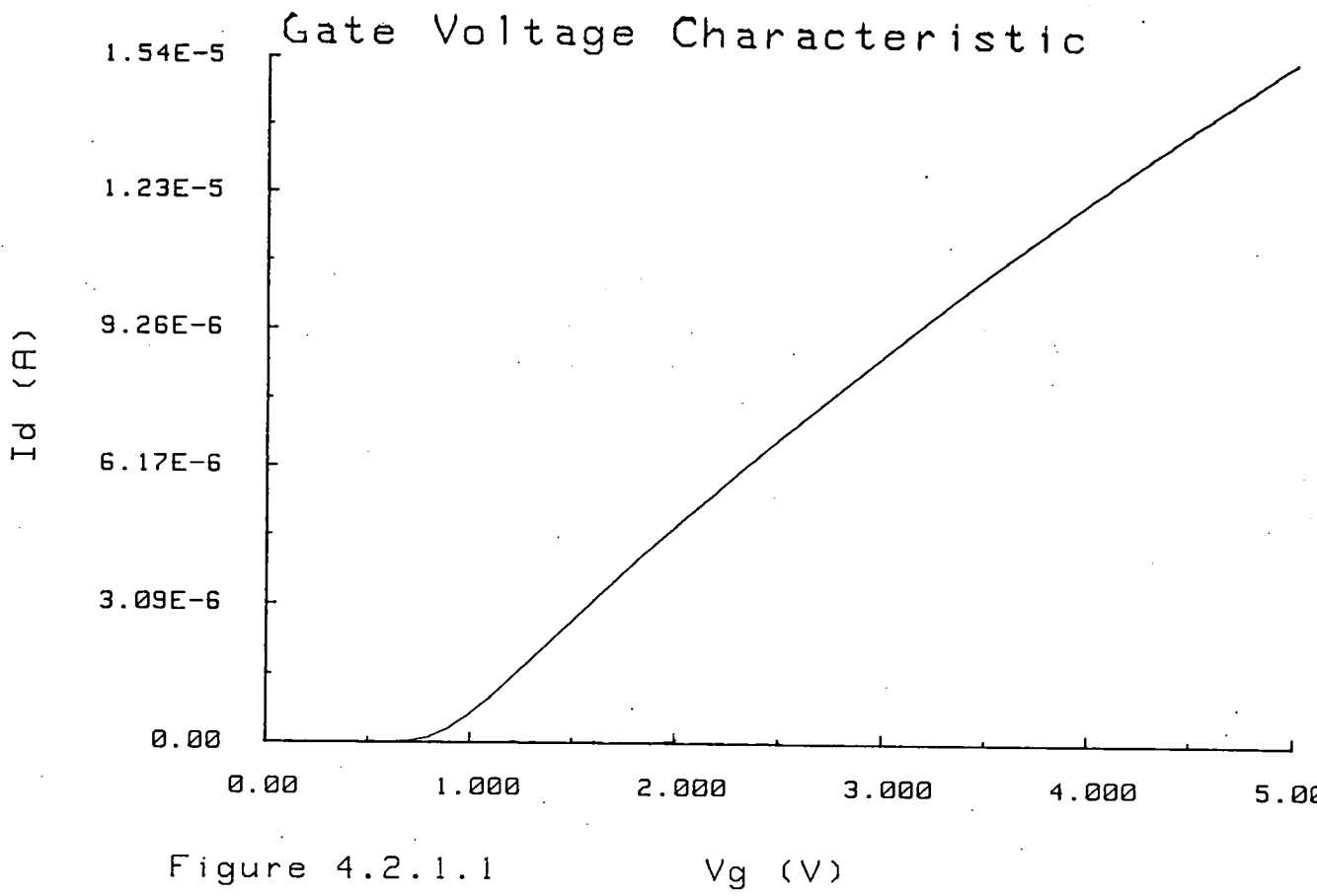


Figure 4.2.1.1

V_g (V)

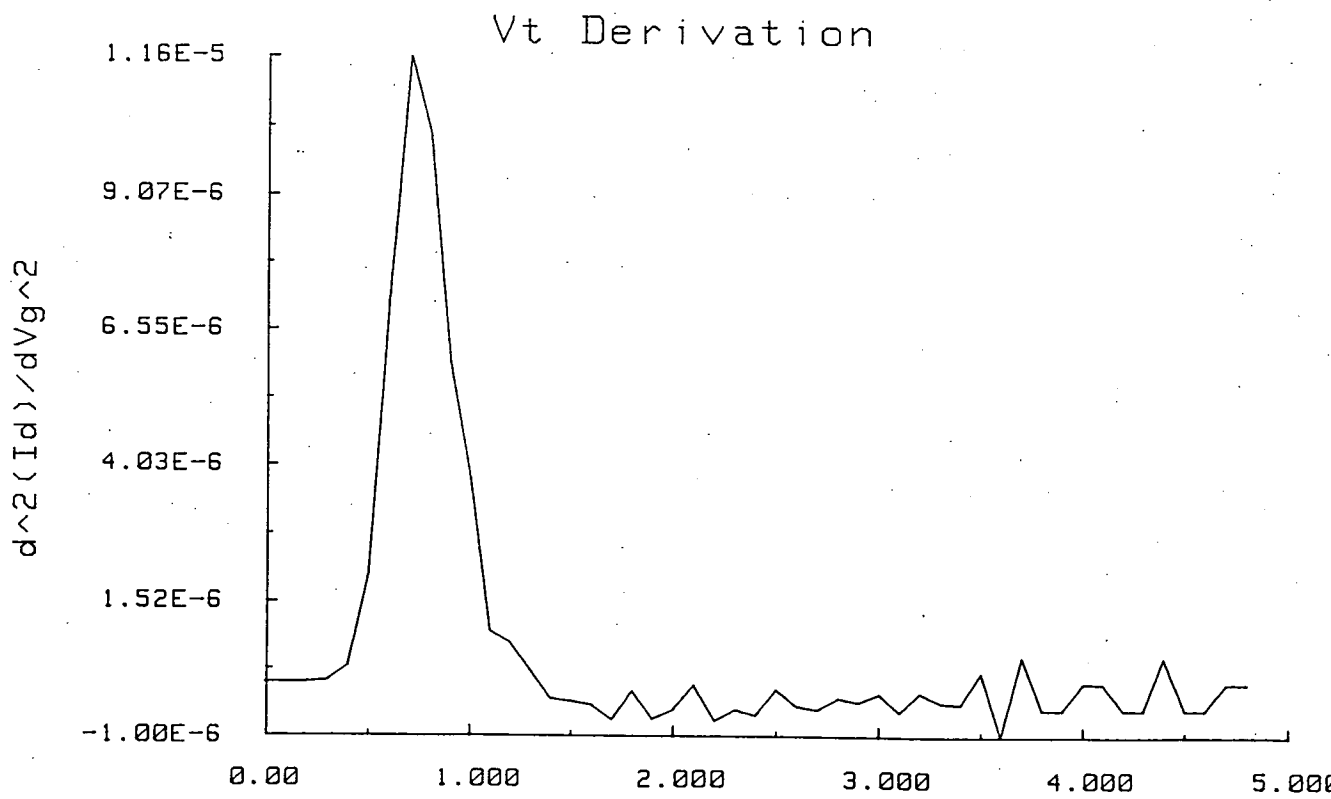


Figure 4.2.1.2

V_g (V)

voltage increases above threshold. An algorithm which is effective for enhancement and depletion devices is to look for the maximum second derivative of current. This is the rate of change of transconductance which is a maximum at the turn-on point (see figure 4.2.1.2). The $V_g:I_d$ graph is almost linear just above threshold and five points just above the turn-on voltage are extrapolated using linear regression to obtain the intercept with the gate voltage axis. From the equations above, it can be seen that this is equal to $V_{to} + \frac{V_d}{2}$ (see figure 4.2.1.3).

4.2.2 Diffusion Length , L_d

In the processing of wafers, the source and drain regions are implanted or diffused into areas defined by the gate. Inevitably there is some sideways diffusion under the gate which causes a reduction from the mask length L_m , of $2L_d$ (figure 4.2.2.1). The actual channel length L is given by

$$L = L_m - 2 L_d \quad 4.2.2.1$$

in both levels 1 and 3.

To find the diffusion length L_d , the $Beta$ s of different length transistors are found from their $V_g:I_d$ characteristics (figure 4.2.2.2). The drain voltage is kept low so that there is no depletion region around the drain which would affect the measured lateral diffusion. The gate voltage should be large enough to bias the device in the linear region but no higher. This enables the maximum $Beta$ to be measured and so reduces the influence of parasitic source and drain contact resistances on the measurement.⁵⁷

Biasing the devices in the linear region means that the $Beta$ values can be found from

$$Beta = \frac{I_d}{\left[V_g - V_{th} - \frac{V_d}{2} \right] V_d} \quad 4.2.2.2$$

The factor F_b in the level 3 equation makes negligible difference to the value of L_d which is calculated. Having found L_d , F_b can be taken into account in the other parameters whose extraction follows.

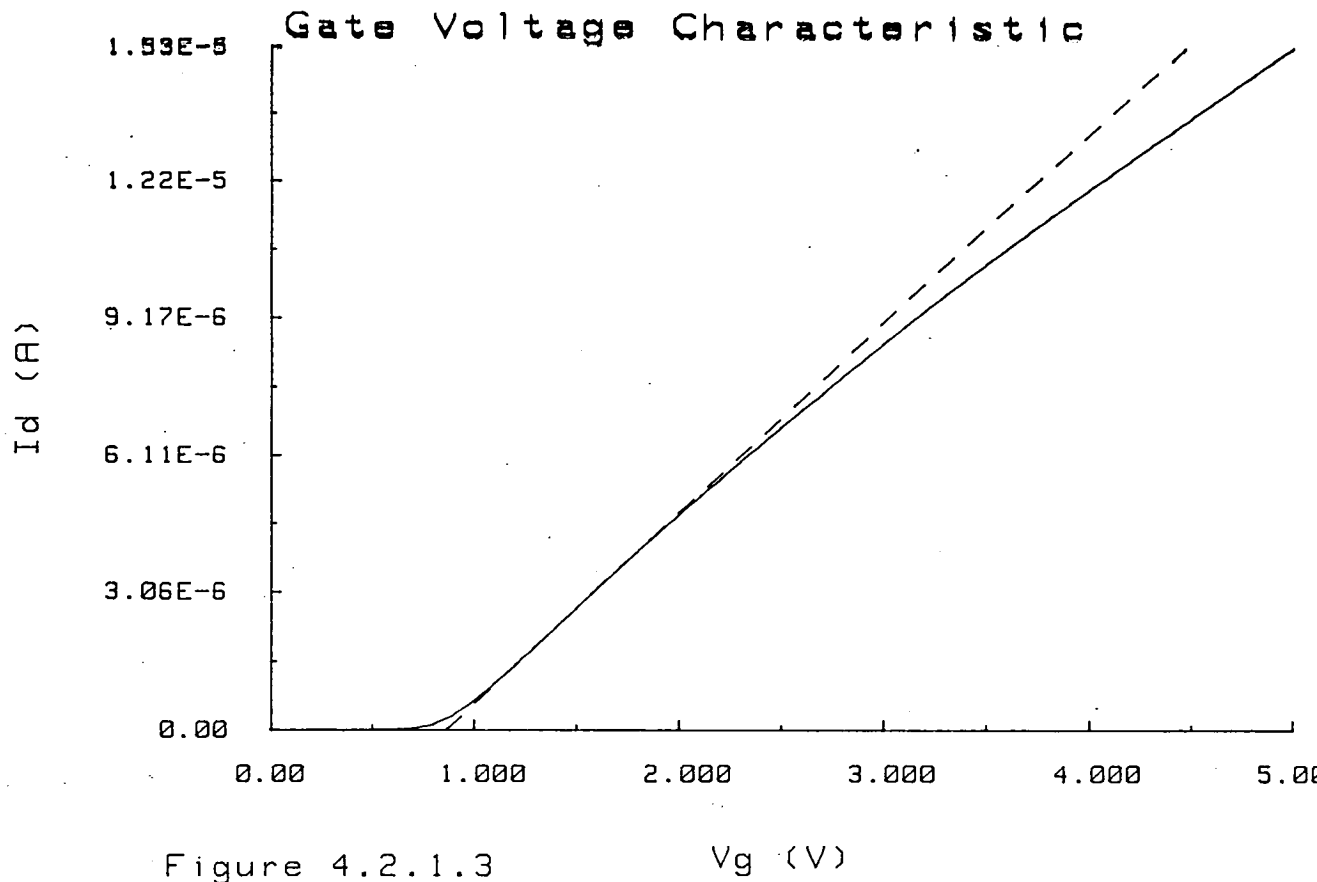
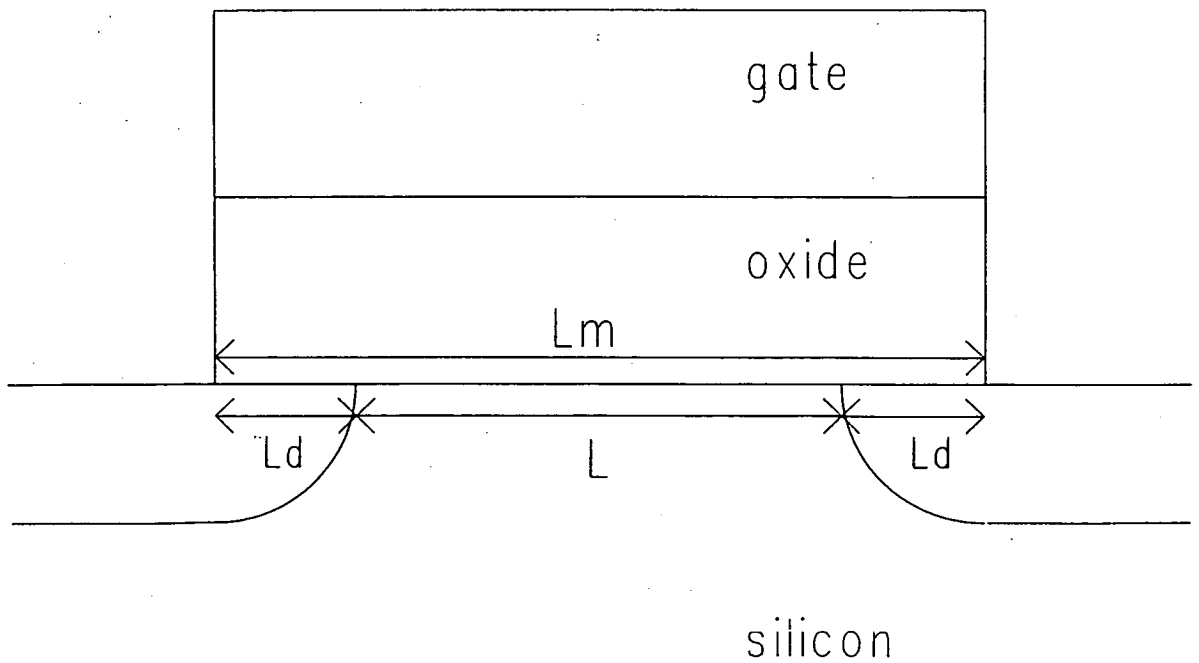


Figure 4.2.1.3

Figure 4.2.2.1 Channel Length Reduction



Beta can also be expressed as

$$Beta = \mu_{eff} C_{ox} \frac{W_m}{L_m - 2L_d}$$

$$\Leftrightarrow \frac{1}{Beta} = \frac{L_m}{\mu_{eff} C_{ox} W_m} - \frac{2L_d}{\mu_{eff} C_{ox} W_m} \quad 4.2.2.3$$

If $\frac{1}{Beta}$ is plotted against L_m then the intercept of the best fit straight line with the L_m axis is $2L_d$ (figure 4.2.2.3).

4.2.3 Substrate Bias Coefficient , γ

The inclusion of substrate bias in the calculation of threshold voltage is different for levels 1 and 3 and therefore the method of obtaining γ is different for each model. In level 1 the relationship is straightforward

$$V_{th} = V_{to} + \gamma |V_b|^{\frac{1}{2}} \quad 4.2.3.1$$

The same $V_g : I_d$ curve is measured as for V_{to} at several different substrate biases (figure 4.2.3.1). The threshold voltages are found from these (figure 4.2.3.2) and V_{th} is plotted against $|V_b|^{\frac{1}{2}}$ and the slope is γ (figure 4.2.3.3).

In the expression for threshold voltage in level 3, two terms are dependent upon substrate bias:

$$V_{th} = \dots + \gamma F_s (2\phi_b - V_b)^{\frac{1}{2}} + F_n (2\phi_b - V_b) \quad 4.2.3.2$$

It is assumed at this stage that threshold dependence on substrate bias is entirely due to the bulk depletion term and then the narrow channel factor, i.e. the parameter δ , can be calculated later to account for the threshold shift of a narrow device at non-zero substrate bias.

In order to extract the parameter γ for level 3, the threshold voltages at different substrate biases are found just as for level 1 and then three equations have to be solved:

$$N_{sub} = \frac{\gamma^2}{2q \epsilon_{si} \epsilon_o} C_{ox}^2 \quad 4.2.3.3$$

$$2\phi_b = \frac{2kT}{q} \ln \left[\frac{N_{sub}}{n_i} \right] \quad 4.2.3.4$$

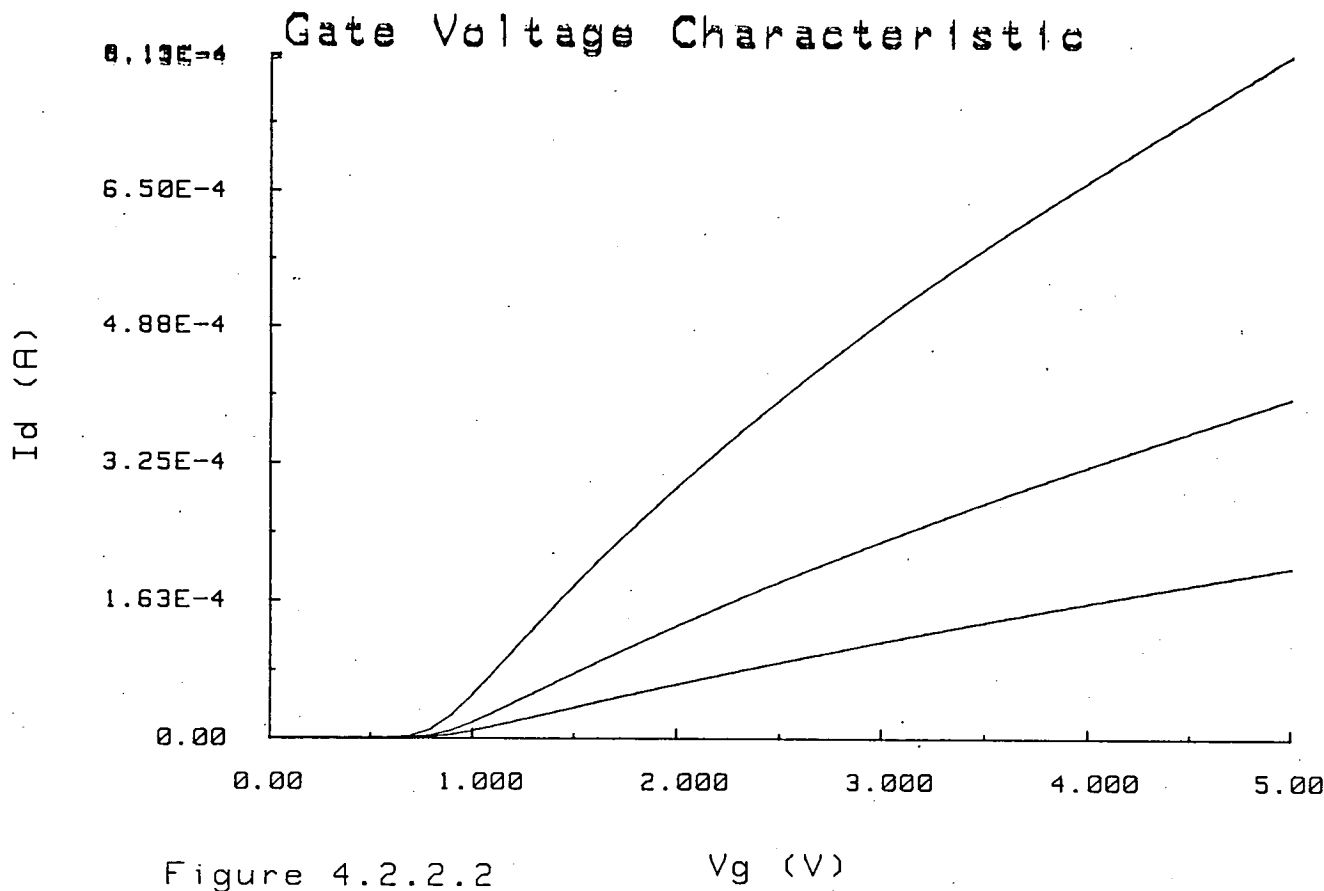


Figure 4.2.2.2

V_g (V)

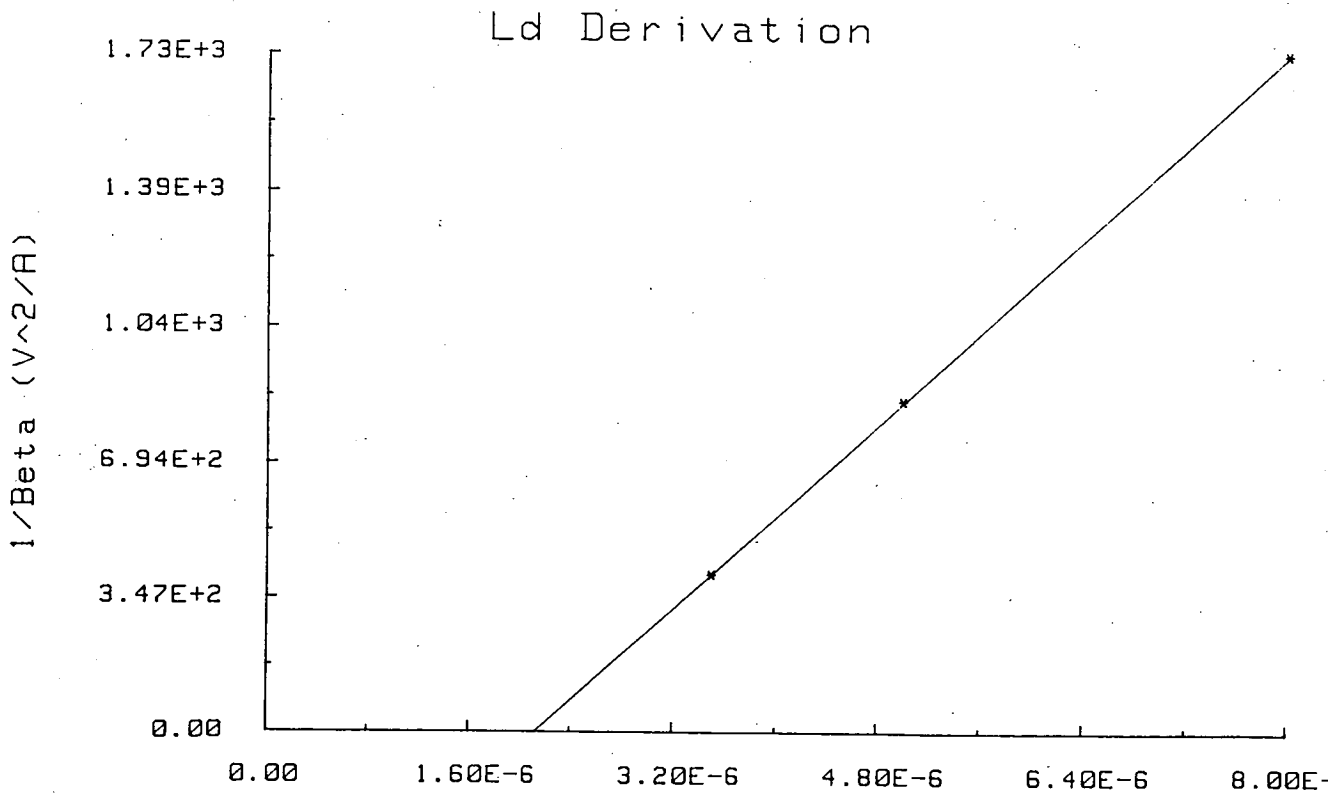


Figure 4.2.2.3

L_m (m)

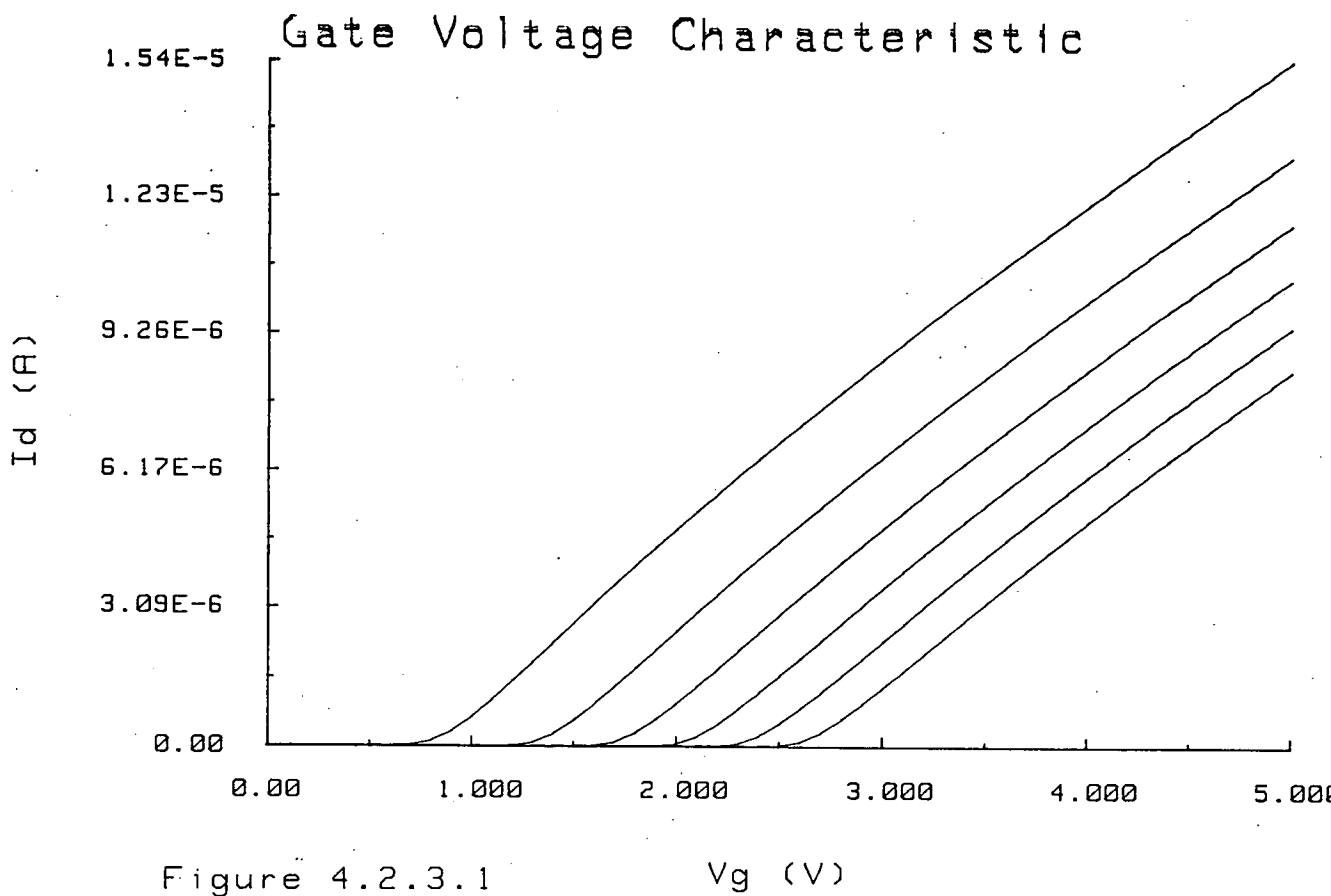


Figure 4.2.3.1 V_g (V)

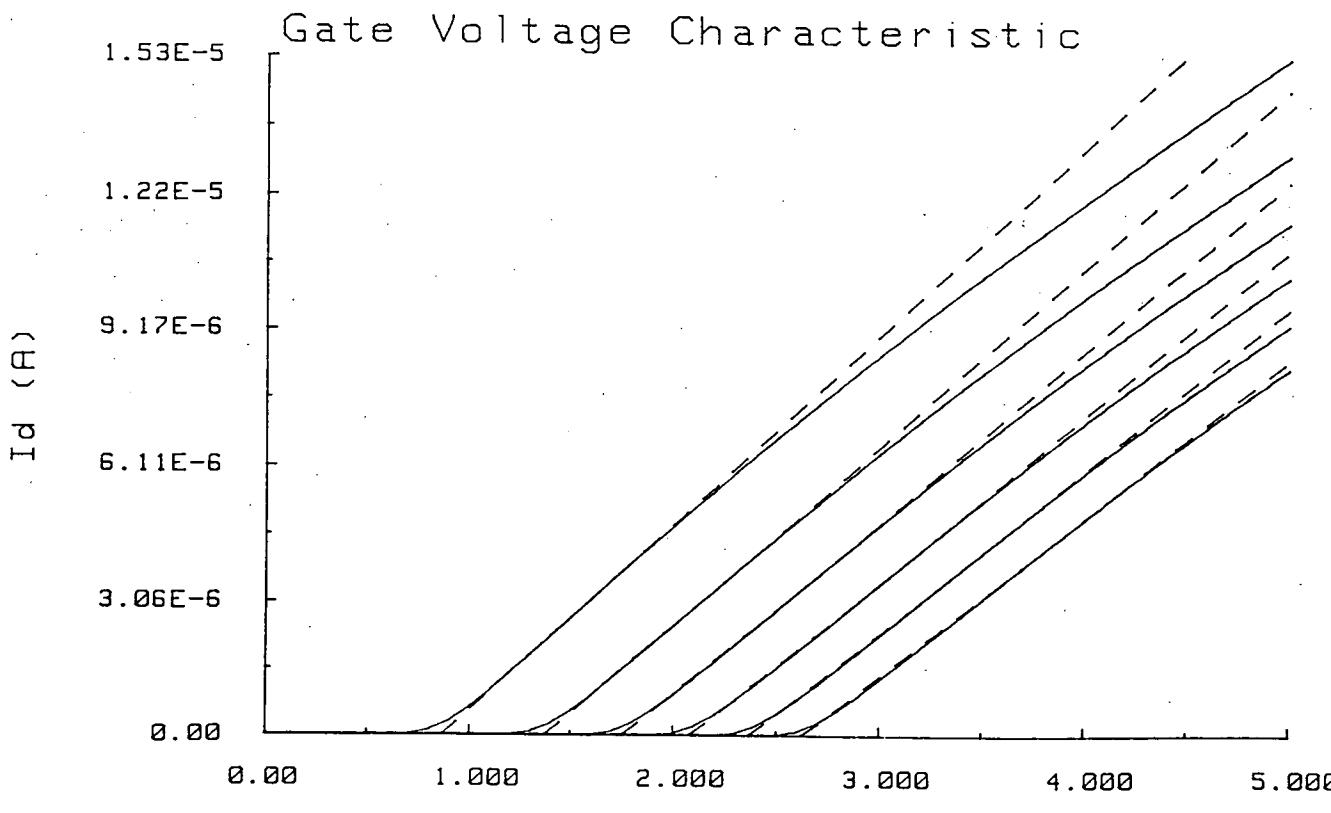


Figure 4.2.3.2 V_g (V)

and

$$V_{th} = V_{io} + \gamma F_s (2\phi_b - V_b)^{\frac{1}{2}} \quad 4.2.3.5$$

These equations are all dependent upon one another. The short channel factor, F_s , is dependent upon substrate concentration and substrate bias. The equations used to calculate F_s , are given in section 2.4.

An estimate is made of γ and the following procedure is repeated until a value satisfying the above equations is found. N_{sub} , X_d and $2\phi_b$ are all derived from γ . The term $F_s (2\phi_b - V_b)^{\frac{1}{2}}$ is then evaluated for each substrate bias and a new γ is determined as the slope of the best fit straight line between the values of $F_s (2\phi_b - V_b)^{\frac{1}{2}}$ and V_{th} (figure 4.2.3.4). After only a few iterations, a value satisfying all these equations is found.

4.2.4 Drain Feedback Coefficient, η

In level 3 only, threshold voltage is also dependent upon the drain voltage through the coefficient σ which is related to the parameter, η . This effect is known as static feedback and its magnitude varies considerably with the length of device. From the threshold voltage equation, it can be seen that σ is minus the slope of the $V_{th}:V_d$ line.

$$V_{th} = V_{fb} + 2\phi_b - \sigma V_d + \gamma F_s (2\phi_b - V_b)^{\frac{1}{2}} + F_n (2\phi_b - V_b) \quad 4.2.4.1$$

The value of η is deduced from

$$\eta = \frac{\sigma C_{ox} L^3}{8.15E - 22} \quad 4.2.4.2$$

To evaluate η , threshold voltage must be measured at different drain voltages. Gate voltage against drain current characteristics at different drain voltages are measured (figure 4.2.4.1). The geometrical factors which influence threshold can be ignored since all the measurements are made on the same device and their adjustments to threshold are not drain voltage dependent. Therefore the threshold offset is the same on each characteristic.

In order to evaluate V_{th} at each value of drain voltage, first of all the leakage current is found on the lowest drain voltage curve where $V_g = V_{th}$. If the threshold voltage lies between two gate voltage points, the leakage current is estimated

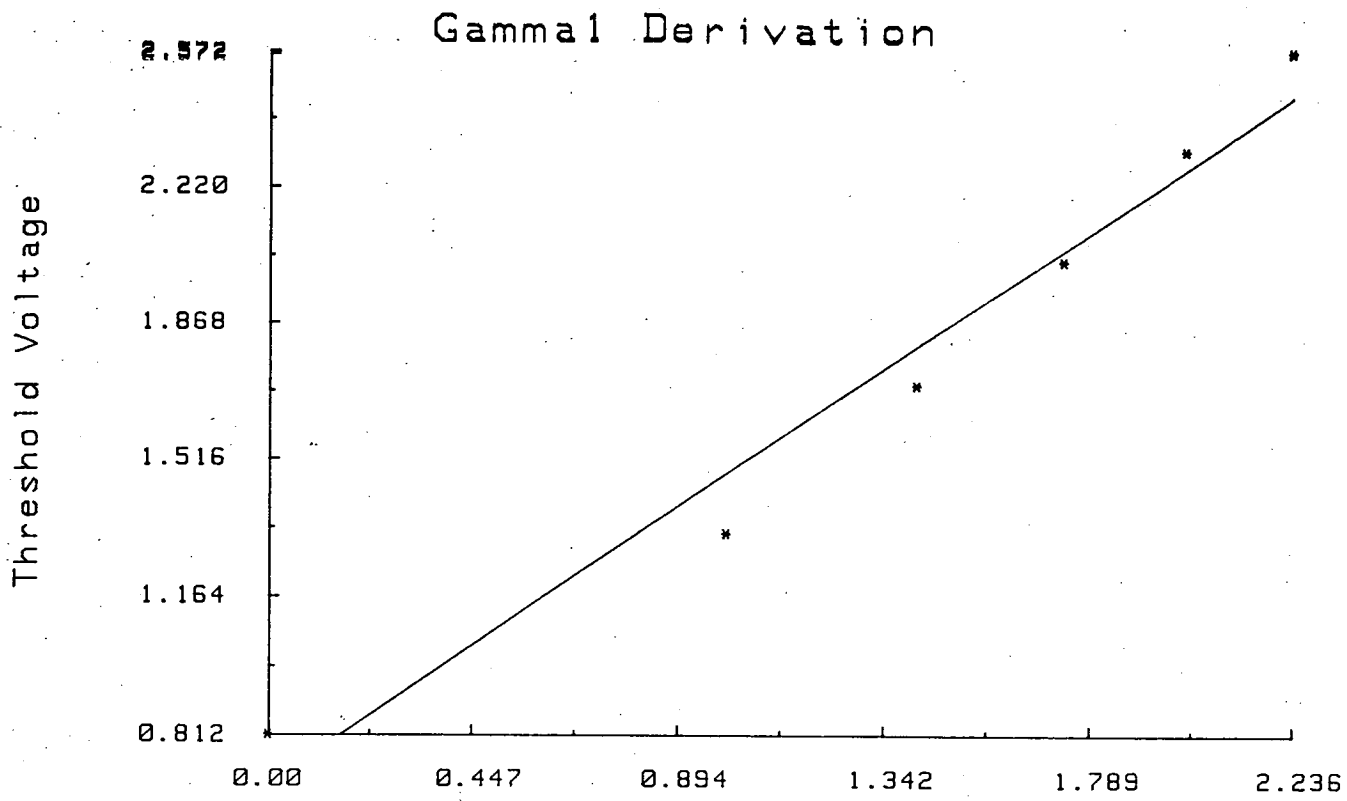


Figure 4.2.3.3 $(V_{bs})^{.5}$

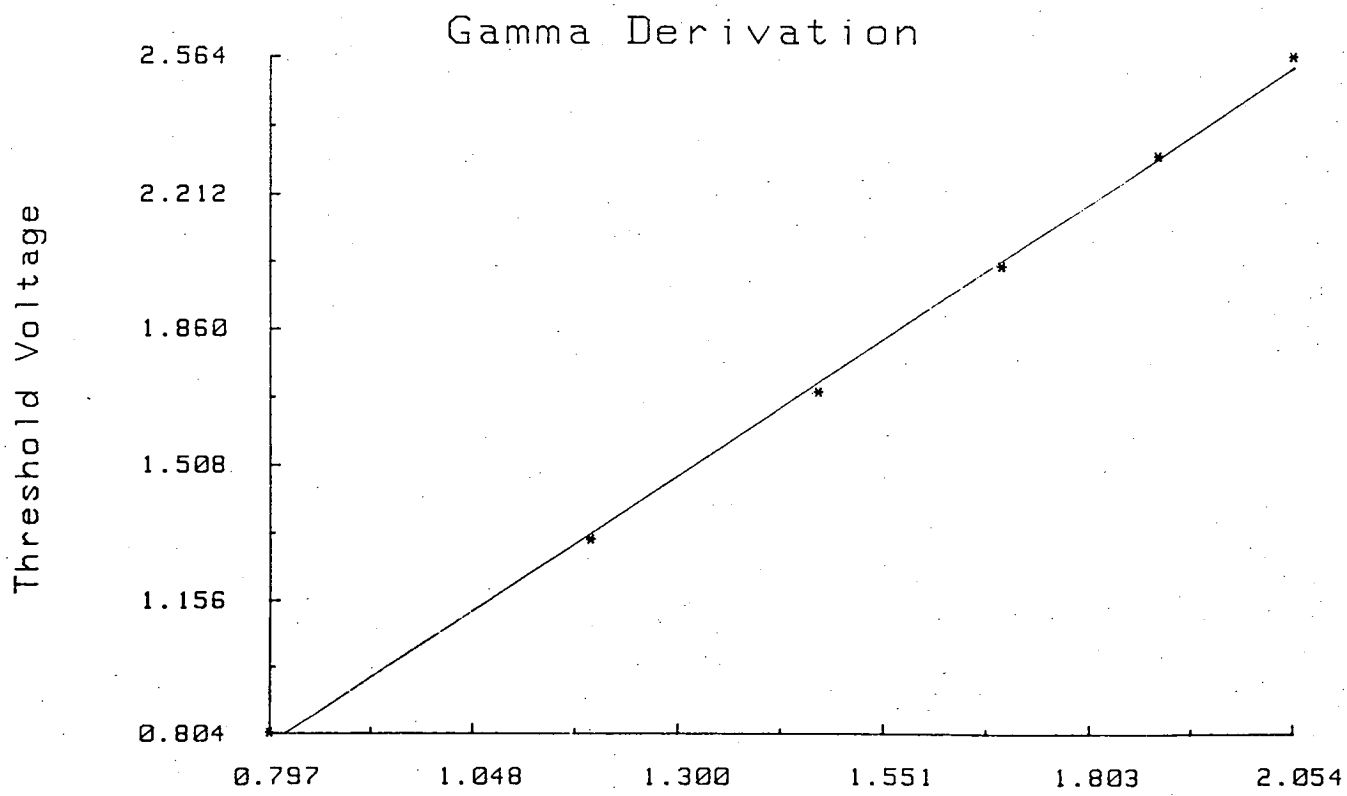


Figure 4.2.3.4 $F_s * (\Phi - V_{bs})^{.5}$

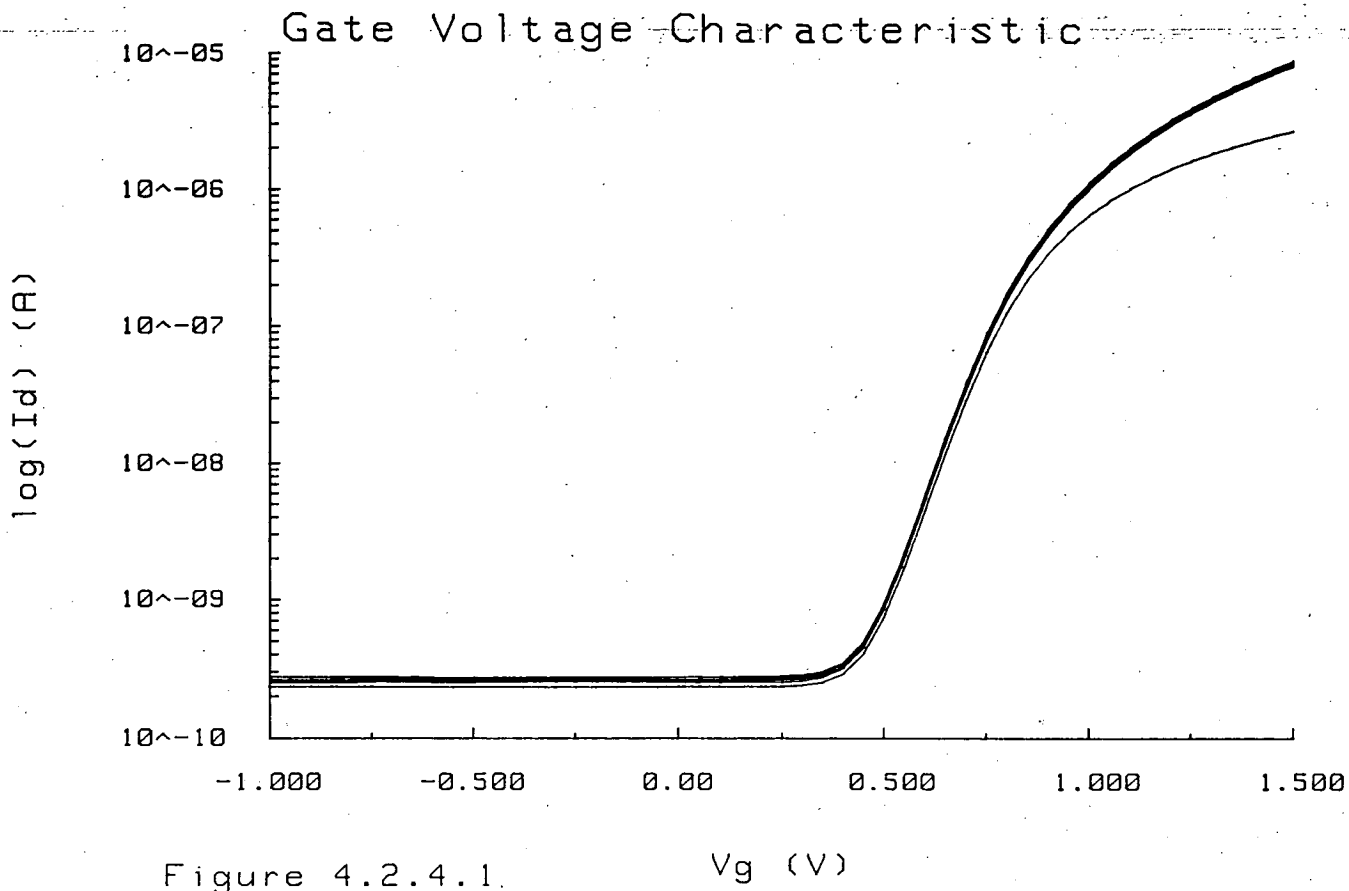


Figure 4.2.4.1. $V_g \text{ (V)}$

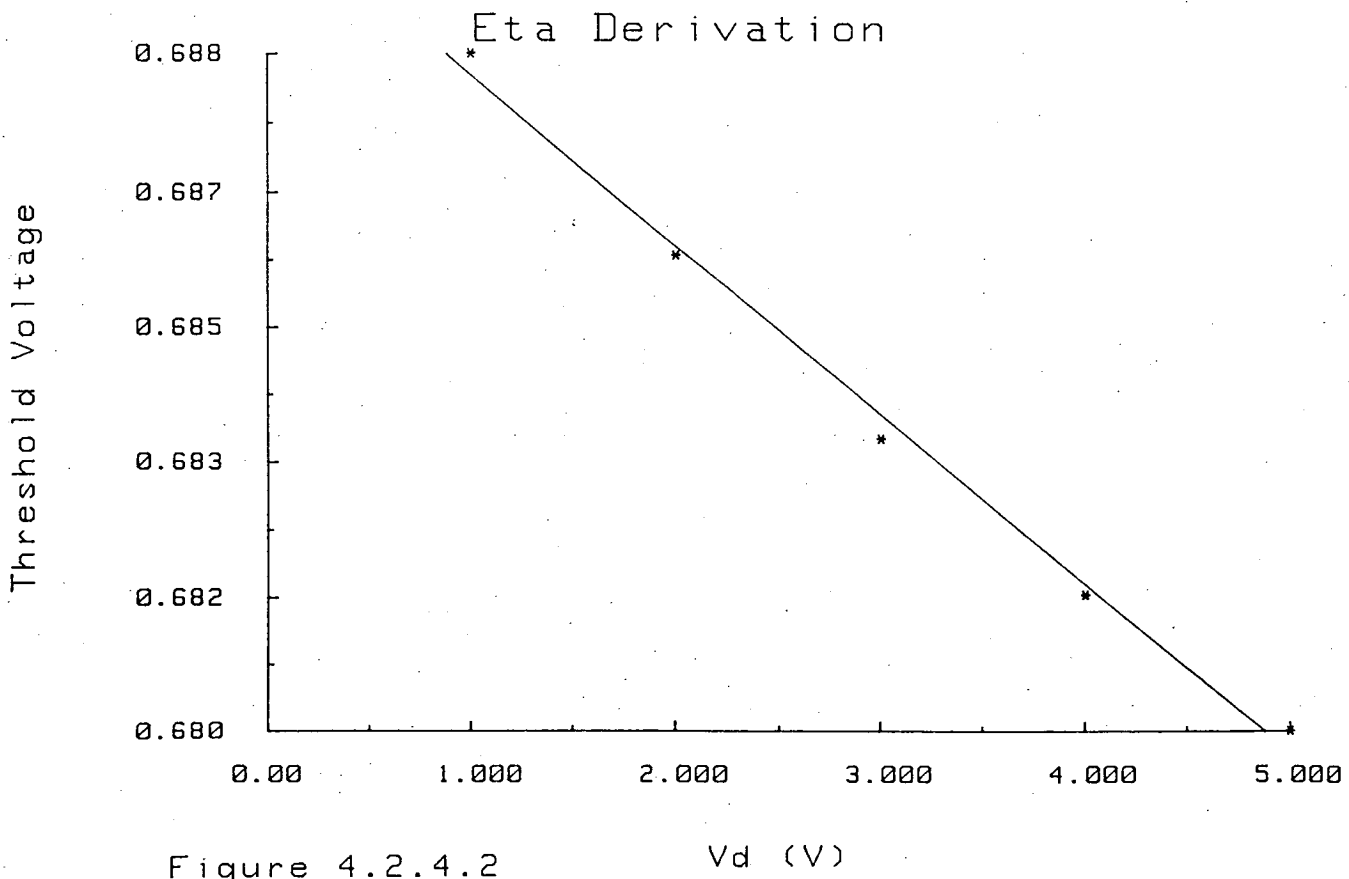


Figure 4.2.4.2 $V_d \text{ (V)}$

using linear interpolation. There is a question here as to whether linear or logarithmic interpolation should be used. Threshold is the point of transition between the subthreshold region where the current is logarithmically dependent on gate voltage and the regions above threshold where the dependence is linear. In practice, linear interpolation of current is best.

Threshold is the onset of strong inversion where the surface potential in the silicon is $2\phi_b$ away from the unbiased bulk potential. This corresponds to a carrier density equal to and of opposite type to that of the original substrate. Therefore the reverse process to the one outlined above is used at each drain voltage to find the gate voltage at which the leakage current at threshold flows in the device. These are the threshold voltages at different drain voltages.

V_{th} is plotted against V_d and σ and η are evaluated as described previously (figure 4.2.4.2).

4.2.5 Width Reduction , Δ_w

As well as the reduction in channel length, channel width is reduced below the mask width during processing. In this case, the difference is caused by the bird's beak effect in the field oxide and also the encroachment of the field implant into the channel region (figure 4.2.5.1).

In both level 1 and level 3, the actual channel width W is given by

$$W = W_m - 2 \Delta_w \quad 4.2.5.1$$

The extraction of Δ_w is similar to L_d and so the $Beta$ values for various width devices have to be found from their $V_g:I_d$ characteristics (figure 4.2.5.2). The drain voltage is kept low to keep the depletion region small. The gate voltage is chosen, just as for L_d , to maximise $Beta$.⁵⁷

The devices are biased in the linear region so that

$$Beta = \frac{I_d}{\left[V_g - V_{th} - \frac{(1+F_b)}{2} V_d \right] V_d} \quad 4.2.5.2$$

Figure 4.2.5.1 Channel Width Reduction

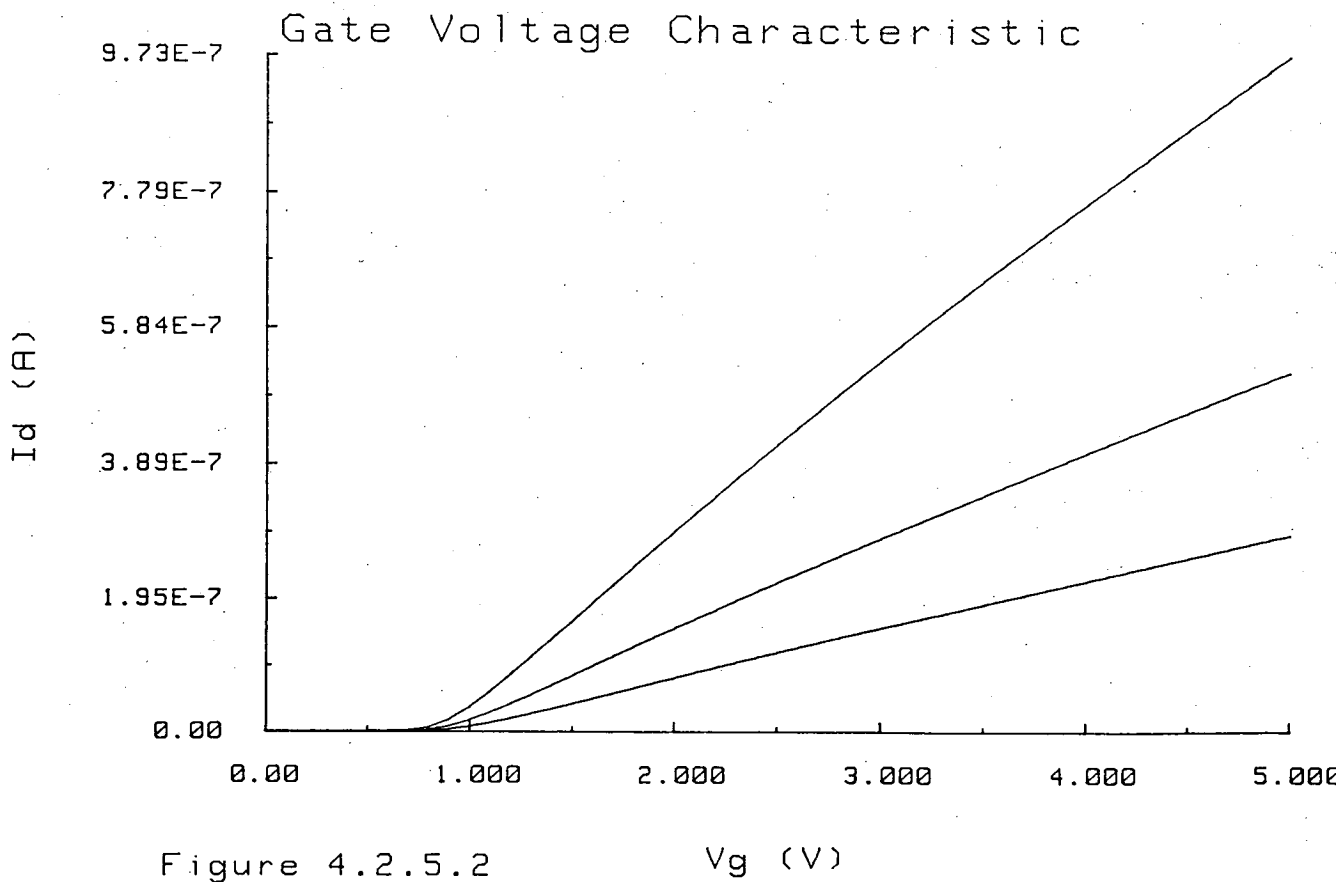
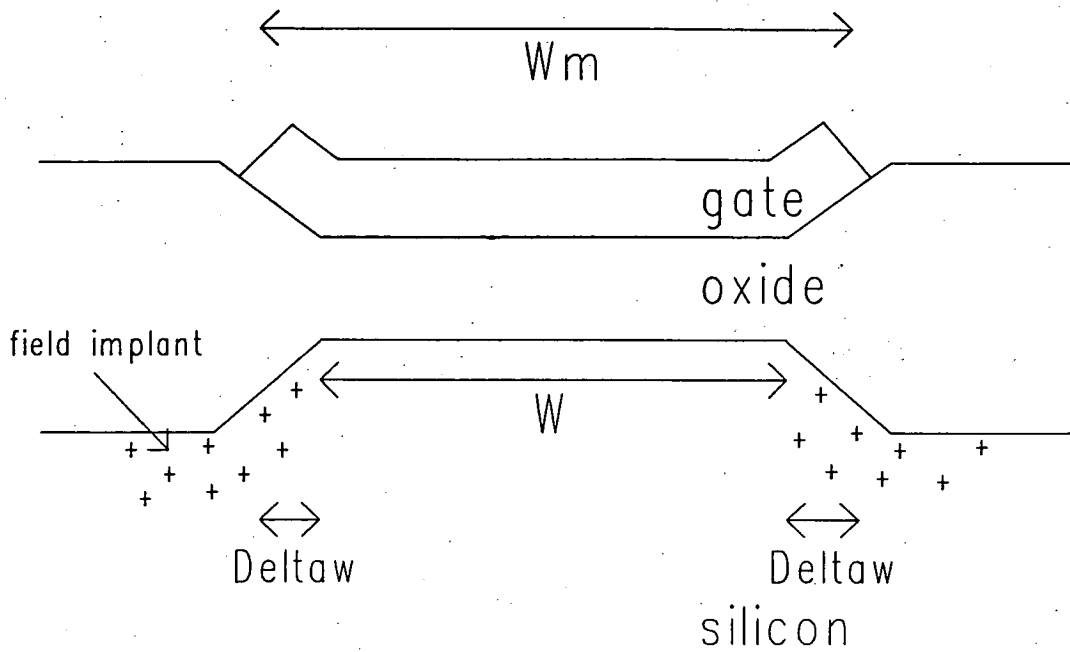


Figure 4.2.5.2

V_g (V)

and also

$$Beta = \mu_{eff} C_{ox} \frac{W_m - 2 \Delta_w}{L_m - 2 L_d}$$

$$\Leftrightarrow Beta = \mu_{eff} C_{ox} \frac{W_m}{L_m - 2 L_d} - \mu_{eff} C_{ox} \frac{2 \Delta_w}{L_m - 2 L_d} \quad 4.2.5.3$$

This shows that $2 \Delta_w$ is the intercept on the W_m axis when plotting $Beta$ against W_m (figure 4.2.5.3).

4.2.6 Narrow Channel Factor , δ

As stated in the section on the extraction of γ , this level 3 parameter is used to take account of threshold of a narrow device at non-zero substrate bias. Therefore the $V_g : I_d$ characteristic of a narrow device is measured with a substrate voltage applied (figure 4.2.6.1). The threshold voltage is extracted using linear regression (figure 4.2.6.2).

The threshold voltage predicted using the model without any narrow channel correction is evaluated using

$$V_{th} = V_{to} - \gamma 2\phi_b^{\frac{1}{2}} + \gamma F_s (2\phi_b - V_b)^{\frac{1}{2}} - \sigma V_d \quad 4.2.6.1$$

The difference between the measured threshold voltage and the one calculated by the model is equated to the narrow channel term.

$$F_n (2\phi_b - V_b) = V_{th}(meas) - V_{th}(calc) \quad 4.2.6.2$$

$$\Leftrightarrow \delta = \frac{4 C_{ox} W (V_{th}(meas) - V_{th}(calc))}{2 \pi \epsilon_{si} \epsilon_o (2\phi_b - V_b)} \quad 4.2.6.3$$

The parameter δ can then be calculated from the expression above.

4.2.7 Fast State Density , N_{fs}

N_{fs} is a measure of the slope of the subthreshold gate voltage:drain current characteristic in level 3. (In level 1, the subthreshold current is assumed to be zero.) The form of the drain current expression is

$$I_d = I_o \exp \left[\frac{q (V_g - V_{on})}{N k T} \right] \quad 4.2.7.1$$

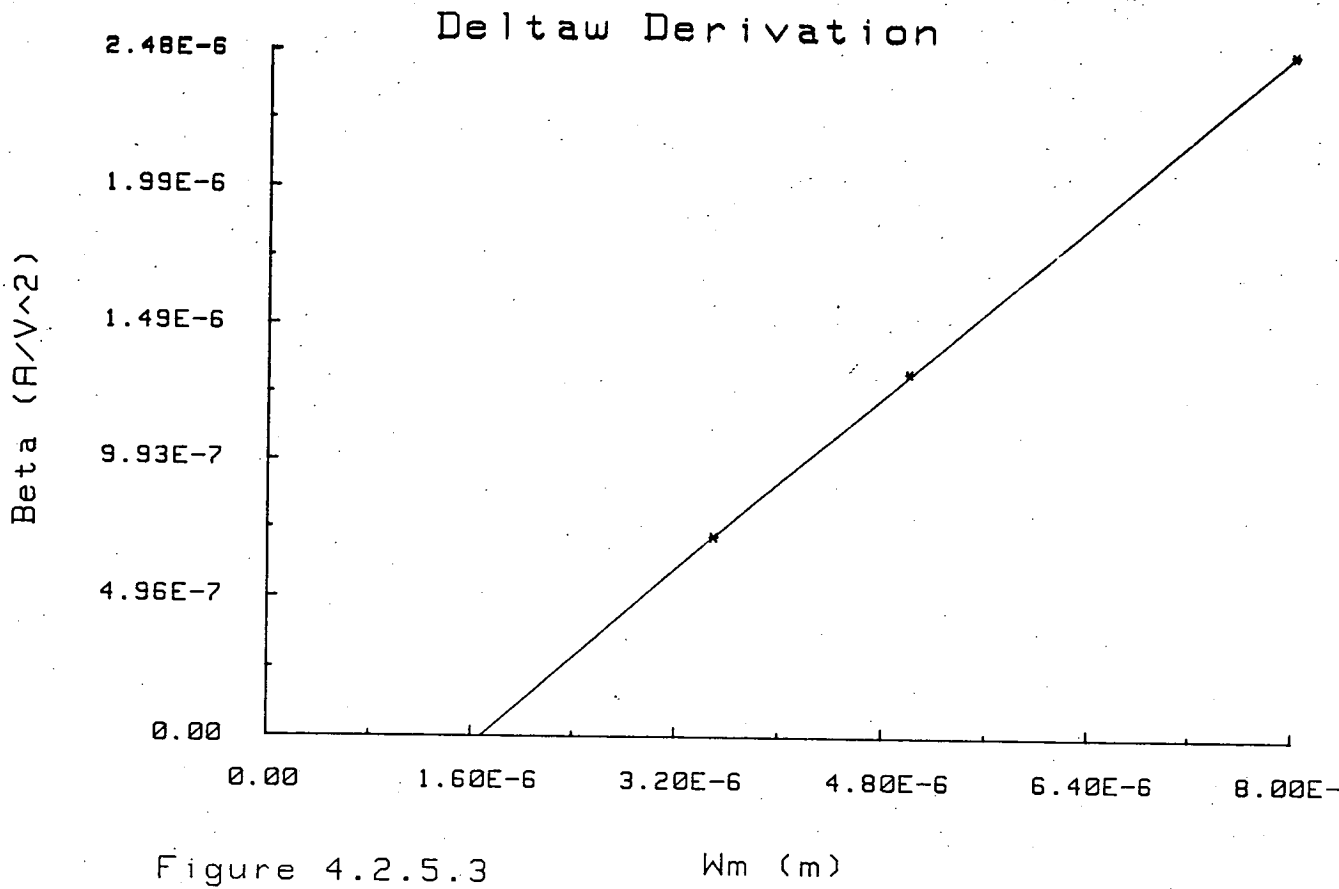


Figure 4.2.5.3 Wm (m)

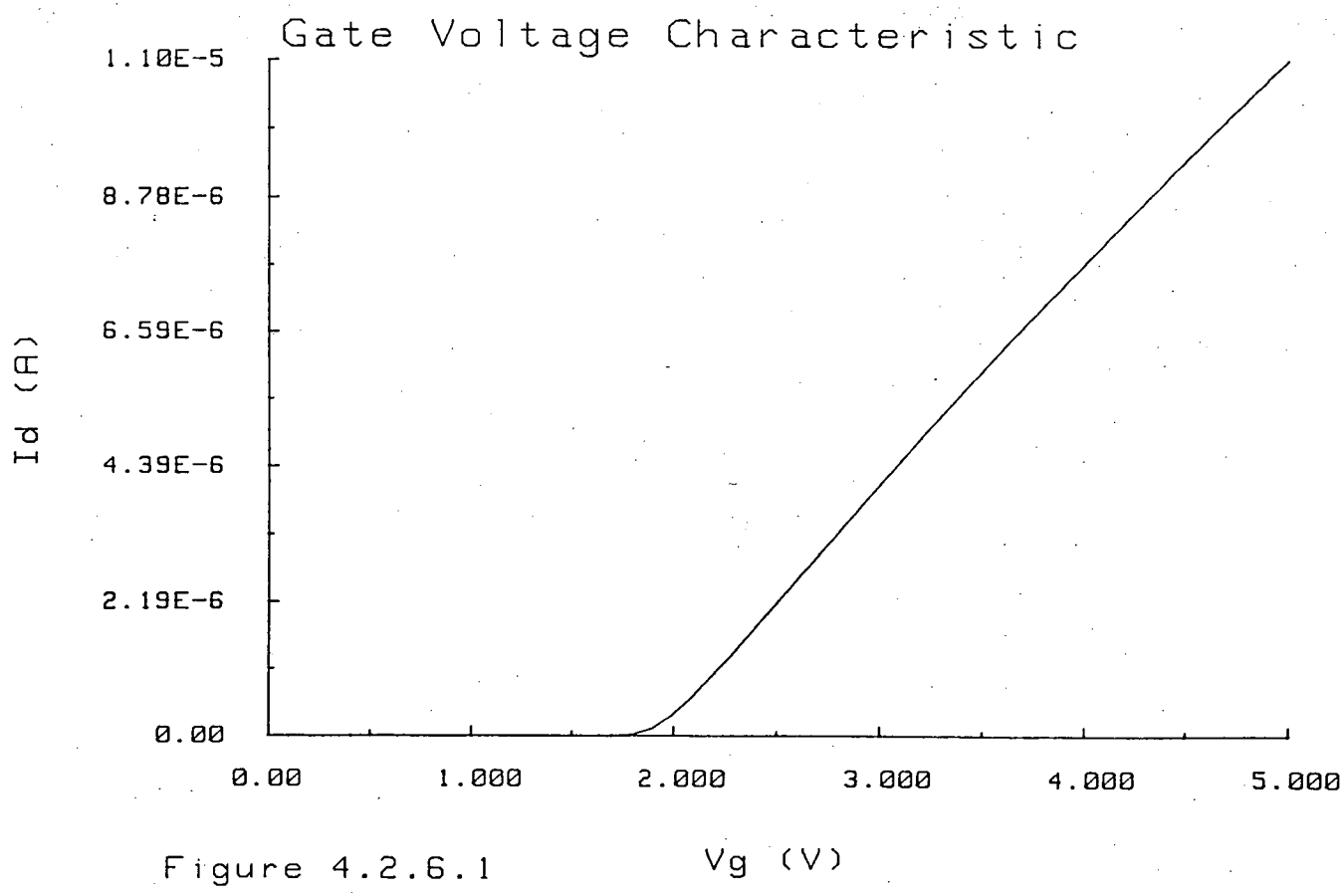
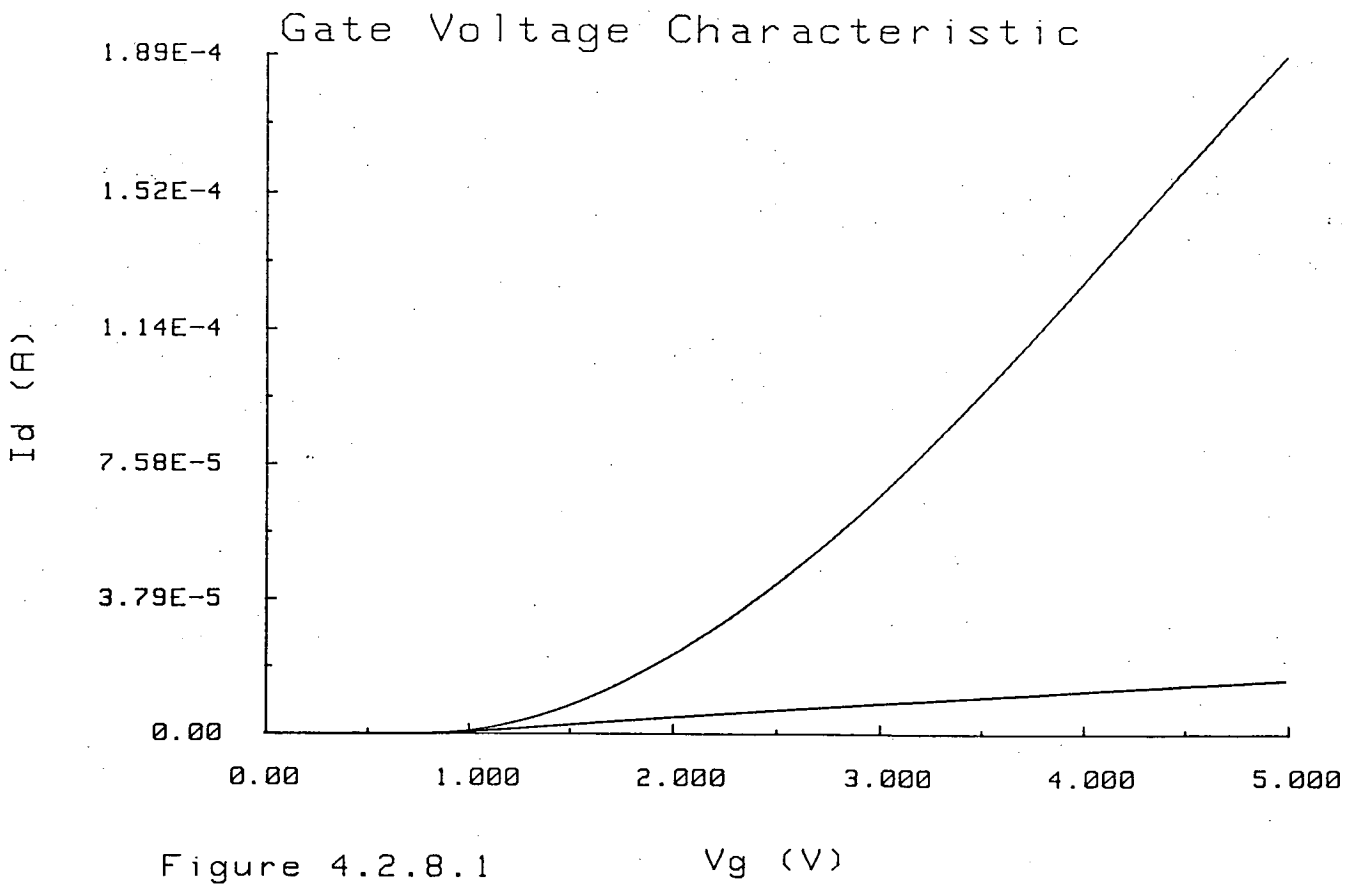
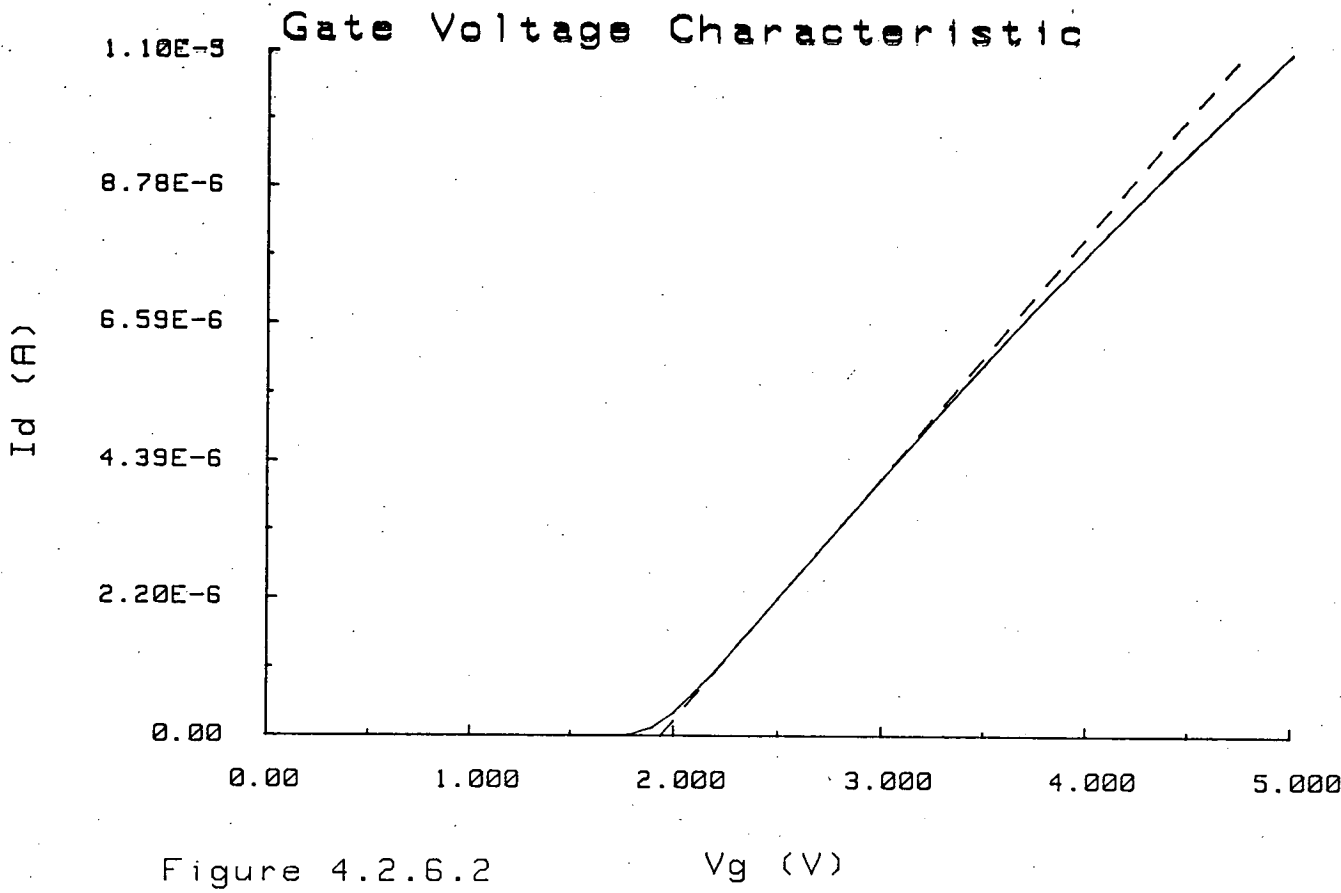


Figure 4.2.6.1 Vg (V)



where I_o is a base current given by the expression for current in the linear region with $V_g = V_{on}$. This is modified by the exponential term which varies with V_g . Turning the above around yields

$$\frac{q (V_g - V_{on})}{N k T} = \ln \left[\frac{I_d}{I_o} \right] \quad 4.2.7.2$$

Currents are measured at two gate voltages in the middle of the subthreshold region. These two points are V_{g1}, I_{d1} and V_{g2}, I_{d2} . Then

$$N = \frac{q (V_{g1} - V_{g2})}{k T \ln \left[\frac{I_{d1}}{I_{d2}} \right]} \quad 4.2.7.3$$

From the model

$$N = 1 + \frac{C_s}{C_{ox}} + \frac{C_d}{C_{ox}} \quad 4.2.7.4$$

where

$$\frac{C_s}{C_{ox}} = \frac{q N_{fs}}{C_{ox}} \quad 4.2.7.5$$

and $\frac{C_d}{C_{ox}}$ can be calculated by using the parameters extracted so far. Finally

$$N_{fs} = \left[N - 1 - \frac{C_d}{C_{ox}} \right] \frac{C_{ox}}{q} \quad 4.2.7.6$$

Two points are chosen from the middle of the subthreshold region on the low drain voltage characteristic as measured for η (figure 4.2.4.1). The procedure outlined above yields the parameter, N_{fs} .

4.2.8 Carrier Mobility, μ_o and θ

Carrier mobility dependence upon gate voltage is modelled by the same relationship in levels 1 and 3. In level 3, mobility is dependent upon drain voltage as well. The model assumes that carrier mobility is a maximum, μ_o at threshold voltage and that there is a gradual degradation of mobility as gate voltage increases which is modelled by θ . The lesser effect, carrier mobility reduction with increasing drain voltage which becomes significant for shorter channels, is modelled in level 3 by v_{max} and its extraction will be described later. Mobility parameters are highly dependent

upon the length of device due to the drain depletion region.

Firstly, the values of mobility at different gate voltages are found. The gate voltage:drain current characteristic is measured at very low V_d for level 3 and at a higher V_d for level 1 (figure 4.2.8.1) in order that the parameters are more accurate over a greater proportion of the range of operating voltages. The parameters extracted so far are used in the model to provide V_{th} and F_b . Mobilities can then be calculated (figures 4.2.8.2 and 4.2.8.3) using the expression:

$$\mu_{eff} = \frac{I_d}{C_{ox} \frac{W}{L} \left[V_g - V_{th} - \frac{1+F_b}{2} V_d \right] V_d} \quad 4.2.8.1$$

In level 1, F_b is zero and in level 3 these mobilities are taken to be the surface mobilities since the drain voltage was kept low. Surface mobility is defined as the carrier mobility before any reduction because of the drain voltage takes place.

The equation governing mobility as a function of gate voltage is

$$\mu_s = \frac{\mu_o}{1 + \theta (V_g - V_{th})} \quad 4.2.8.2$$

$$\Leftrightarrow \frac{\mu_o}{\mu_s} = \theta V_g - \theta V_{th} + 1 \quad 4.2.8.3$$

If the best fit straight line is fitted on the graph of $\frac{1}{\mu_s} : V_g$ then $\frac{1}{\mu_s} = \frac{1}{\mu_o}$ when $V_g = V_{th}$. Theta is the slope of the graph $\frac{\mu_o}{\mu_s}$ against V_g and examples for level 1 and level 3 are shown in figures 4.2.8.4 and 4.2.8.5 respectively.

4.2.9 Maximum Carrier Velocity, v_{max}

As stated in 4.2.8, this effect becomes increasingly significant as channel lengths are reduced and is only modelled in level 3. Drain current is measured as a function of drain voltage for a high gate voltage so that most of the measured characteristic lies in the linear region of operation (figure 4.2.9.1). Mobility is calculated from the drain current values in the same way as it was calculated for the evaluation of μ_o and θ using all the parameters extracted so far (figure 4.2.9.2).

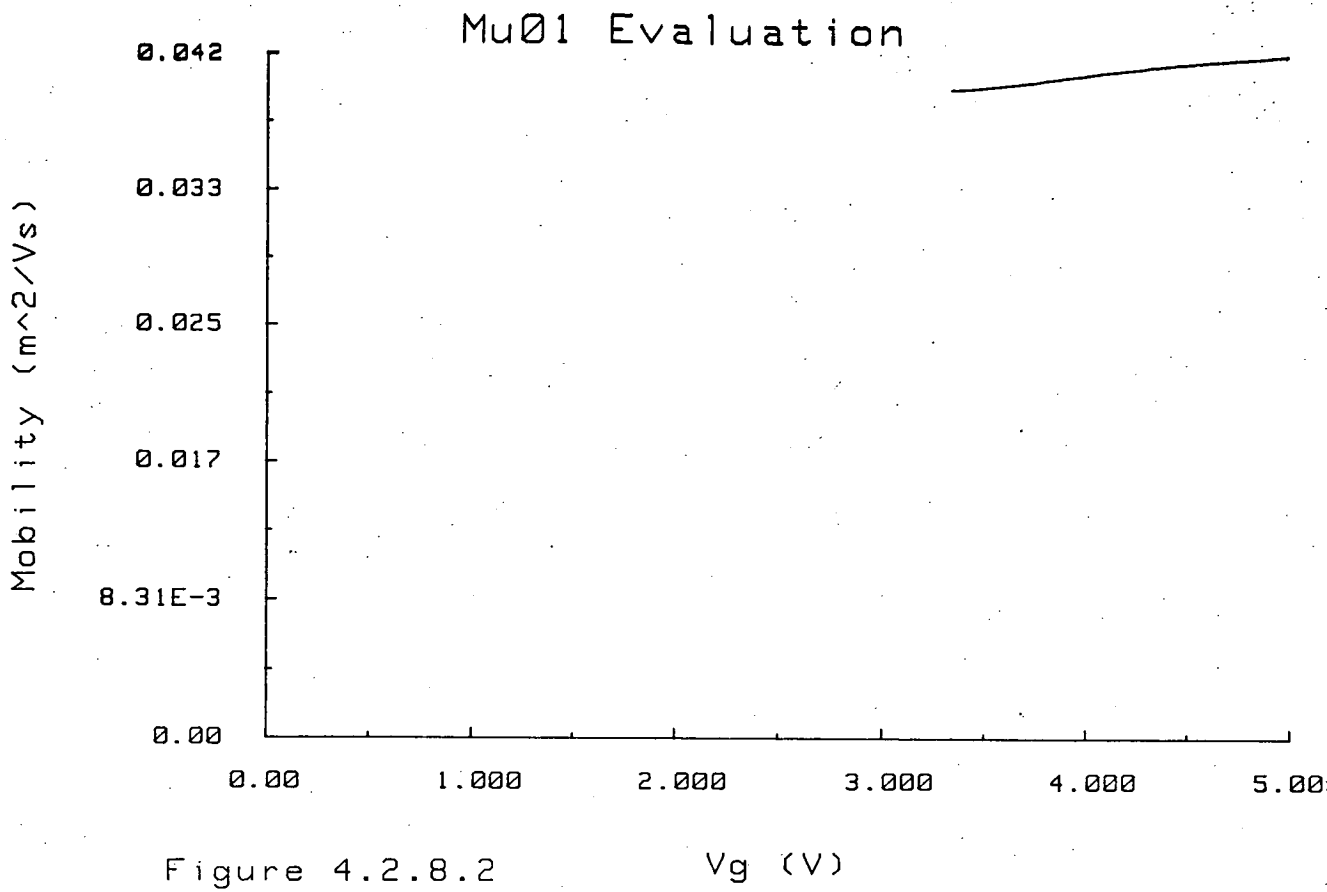


Figure 4.2.8.2 Vg (V)

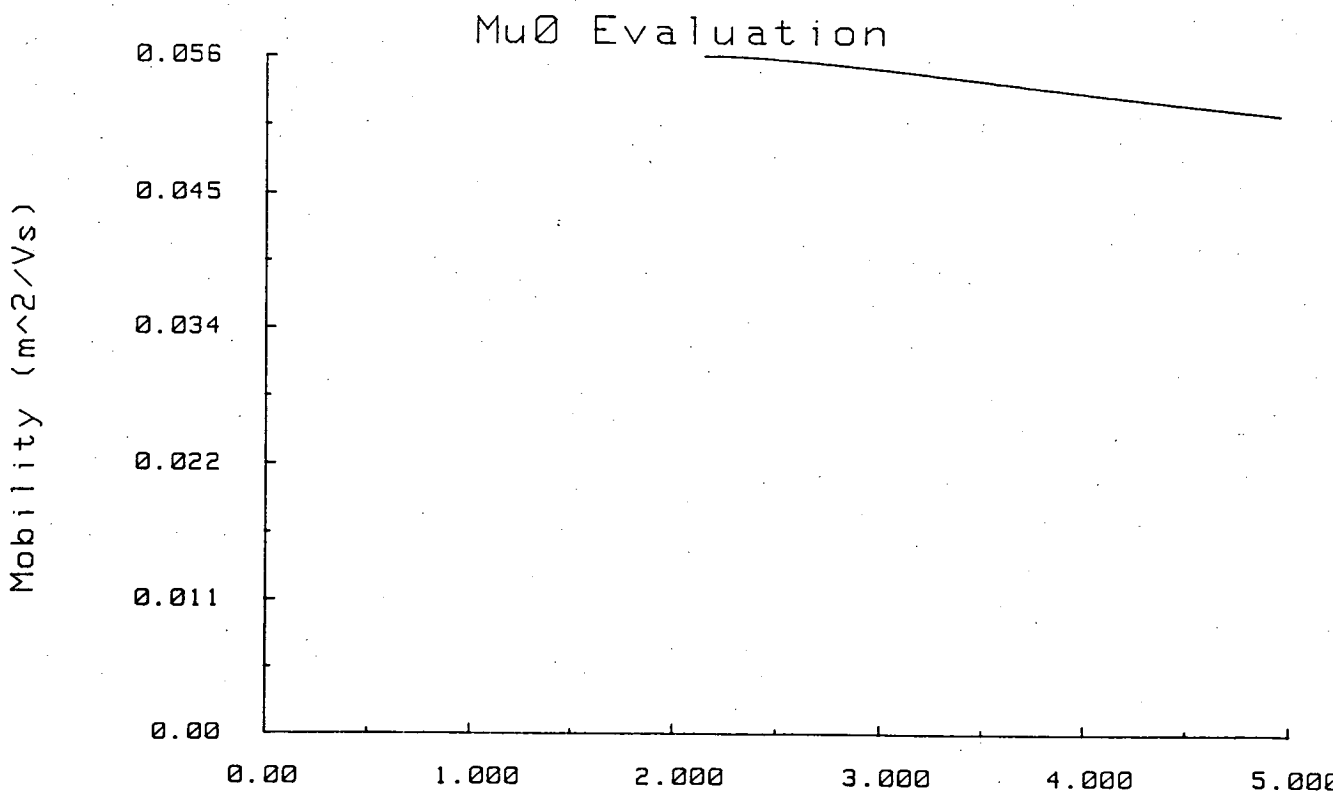


Figure 4.2.8.3 Vg (V)

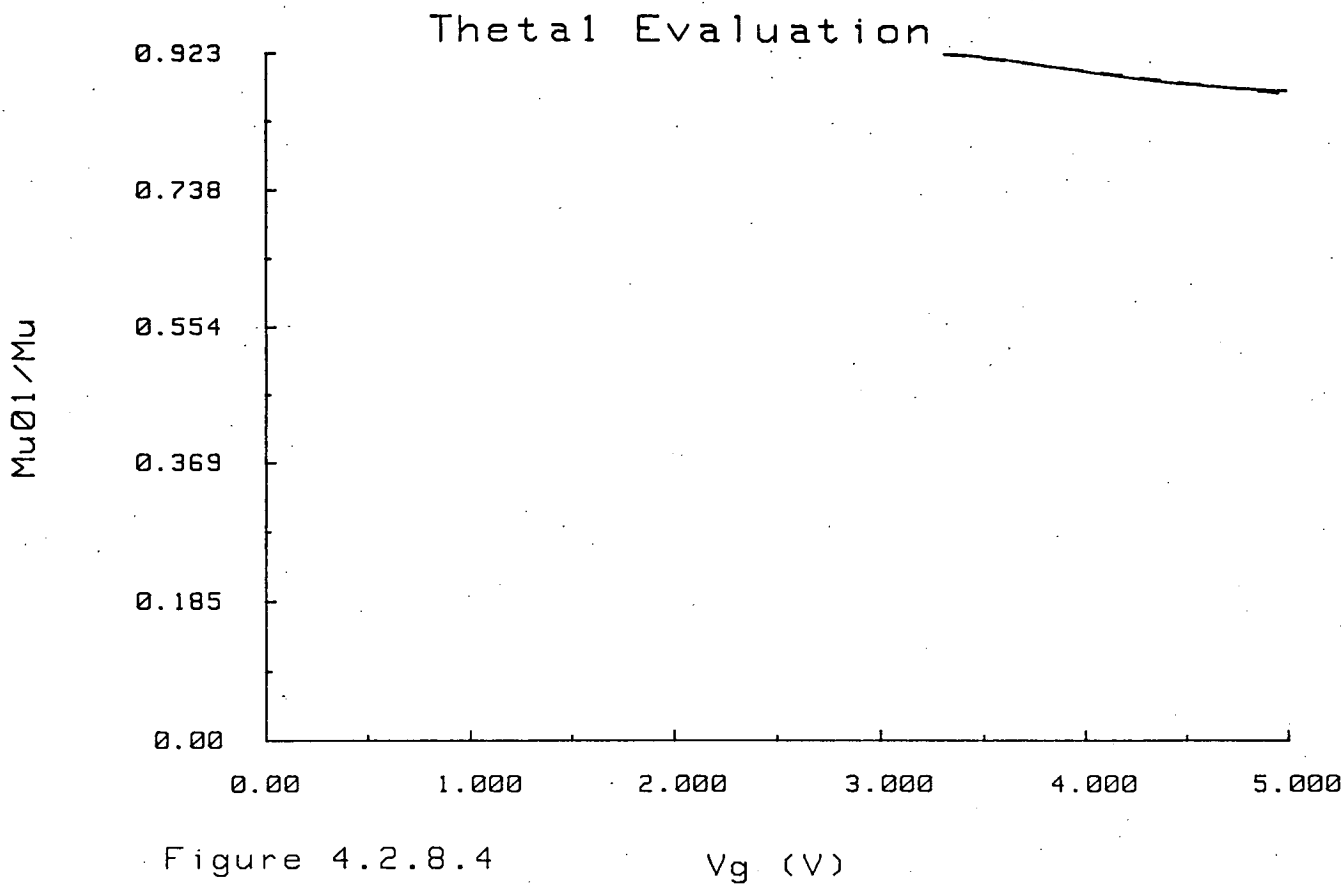


Figure 4.2.8.4 V_g (V)

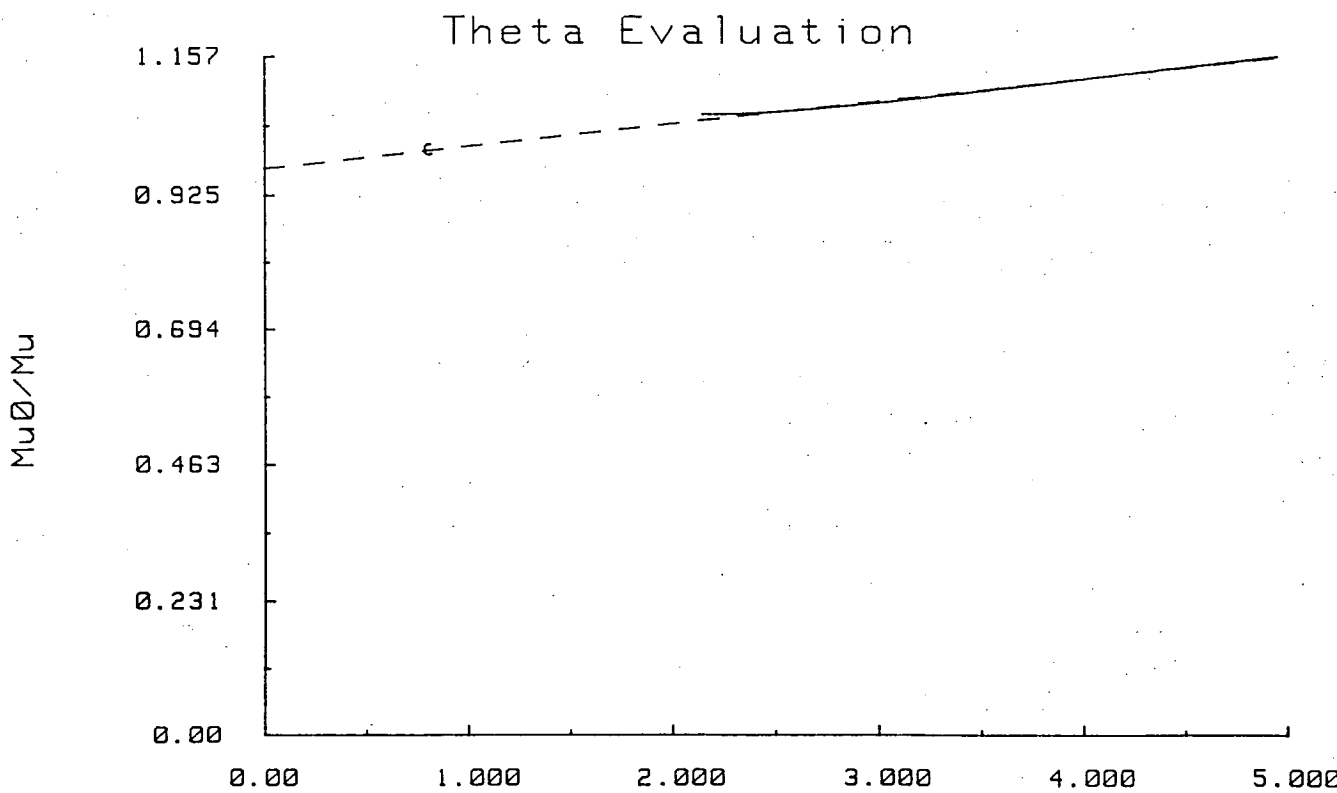


Figure 4.2.8.5 V_g (V)

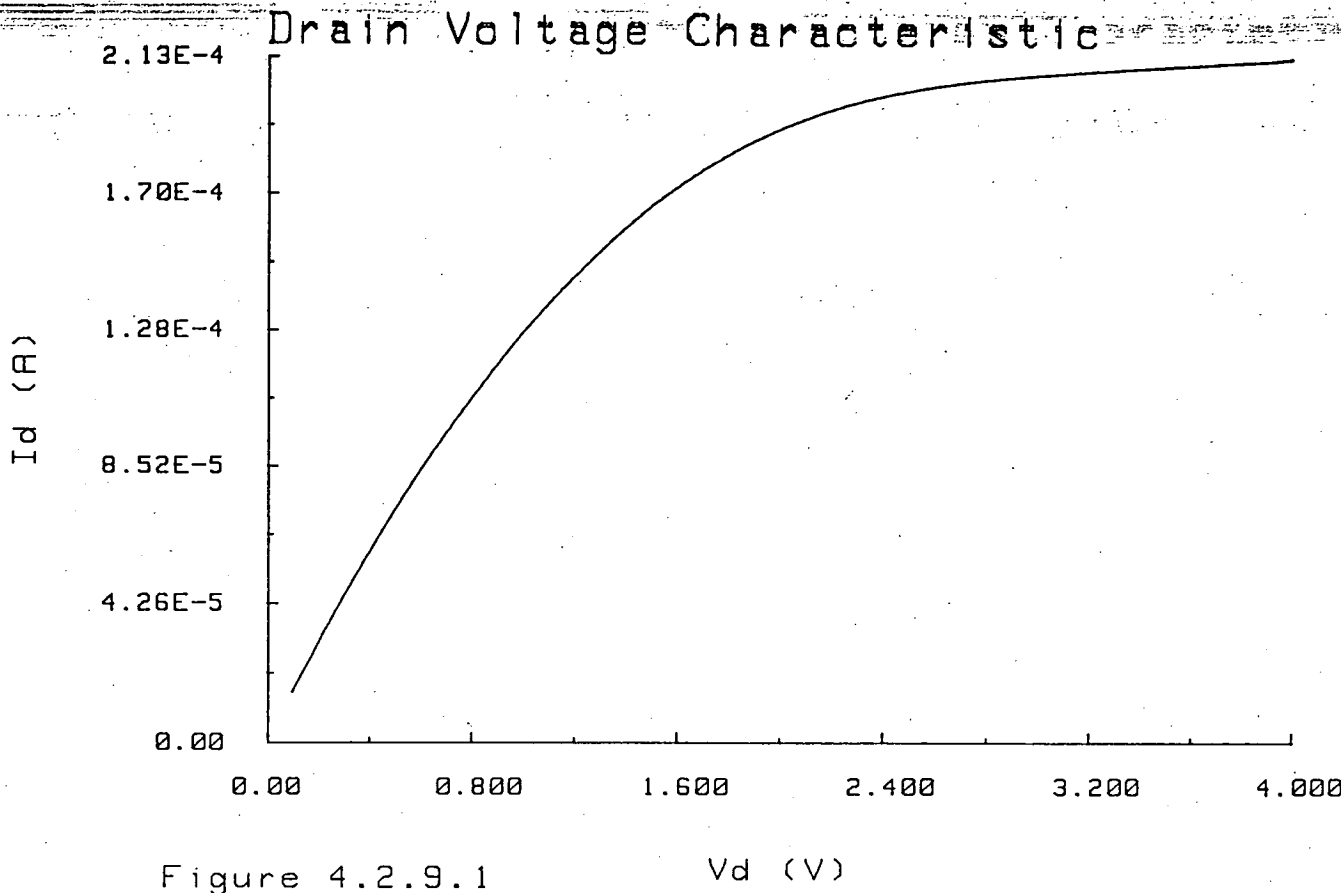


Figure 4.2.9.1 V_d (V)

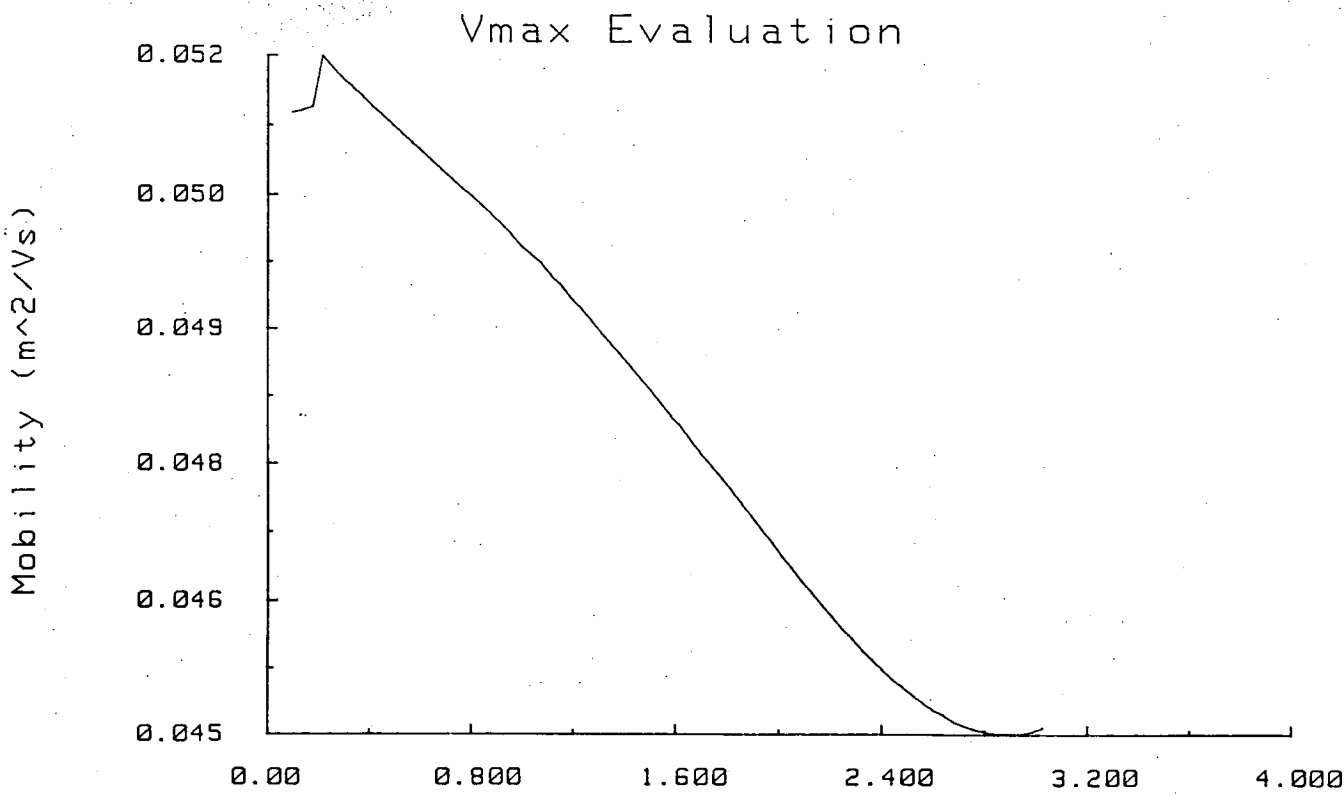


Figure 4.2.9.2 V_d (V)

The governing expression for drain modulation of mobility in the linear region is

$$\mu_{eff} = \frac{\mu_s}{1 + \frac{V_d \mu_s}{L v_{max}}} \quad 4.2.9.1$$

$$\Leftrightarrow \frac{V_d}{L} = \left[\frac{1}{\mu_{eff}} - \frac{1}{\mu_s} \right] v_{max} \quad 4.2.9.2$$

At low drain voltage, the effective carrier mobility μ_{eff} is approximately the surface mobility, μ_s . The parameter v_{max} is the slope of the line obtained when plotting $\frac{V_d}{L}$ against $\frac{1}{\mu_{eff}} - \frac{1}{\mu_s}$ for $V_d < V_{dsat}$ (figure 4.2.9.3).

An iterative technique is required here since V_{dsat} is a function of v_{max} . The iteration is terminated when V_{dsat} is close to its value from the previous iteration.

4.2.10 Saturation Slope Coefficient, κ

In level 3, κ is the parameter which characterises the slope of the drain voltage : drain current curve in saturation. As the voltage on the drain is increased with respect to the gate, the point is reached where carrier inversion cannot be sustained at the drain end of the channel (figure 4.2.10.1). This effect which is called channel pinch-off, combined with carrier velocity saturation explains why at a certain drain voltage the device current stops increasing at the same rate.⁴¹ As drain voltage further increases, the length of channel which is no longer inverted, L_{del} , increases and thus the actual length of the gate-controlled channel is reduced. Conduction continues because of the high parallel electric field across the pinched off region. There is a slight increase in current due to the shortening of the channel. This slight increase is much greater for a shorter channel device where the reduction in length is much more significant since it is a larger proportion of the channel.

Measurements of drain current are made at a medium gate voltage over a range of high drain voltages (figure 4.2.10.2). This should form the characteristic of a device in saturation. The model is then implemented with the parameters extracted to find V_{dsat} and I_{dsat} . The saturation drain current I_{dsat} and the actual measured drain

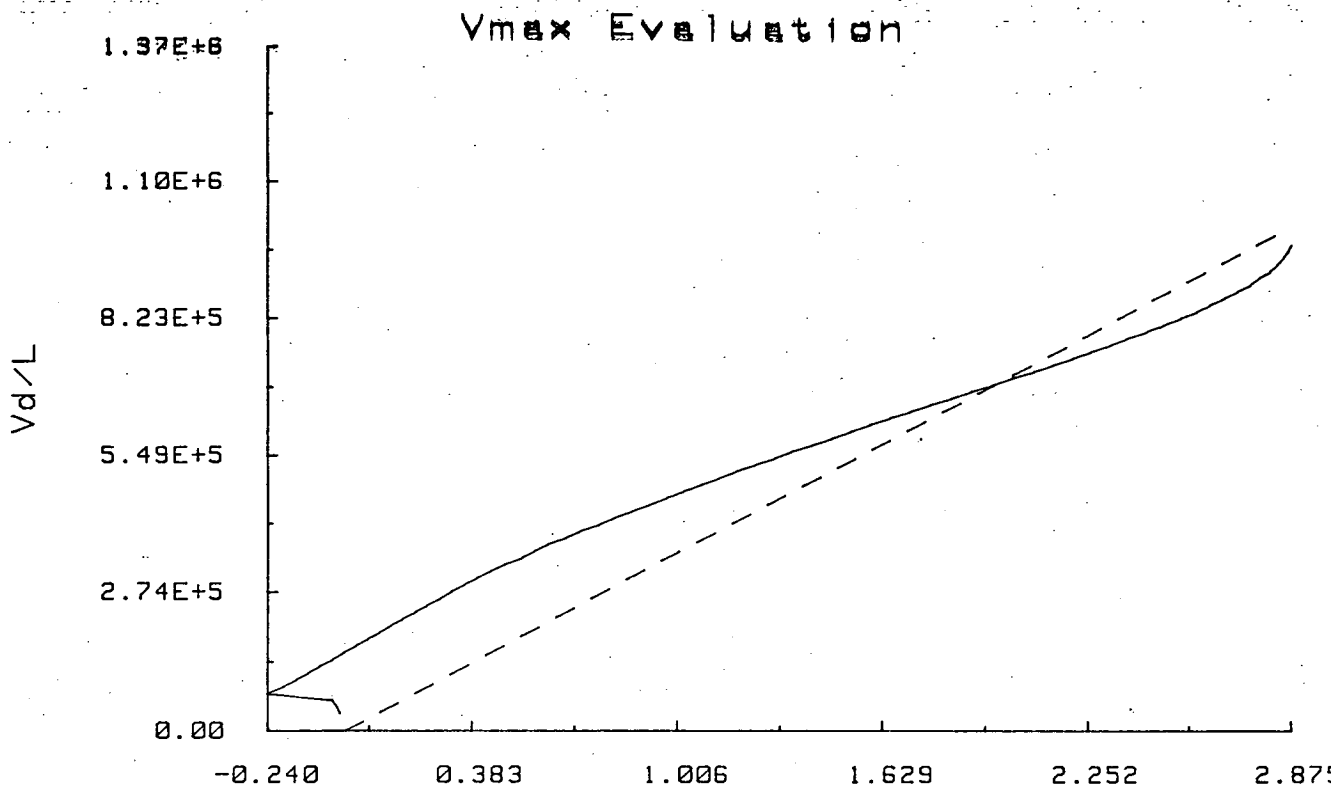
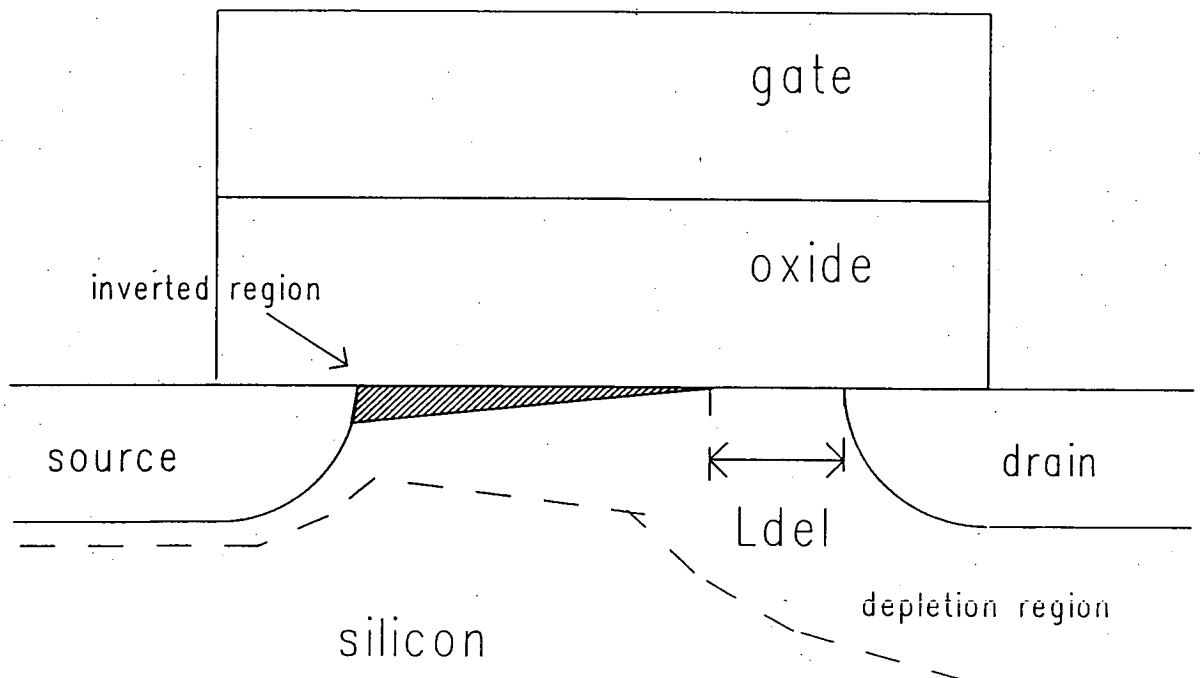


Figure 4.2.9.3 $1/M_{eff} - 1/M_{us}$

Figure 4.2.10.1 Channel Pinch-Off



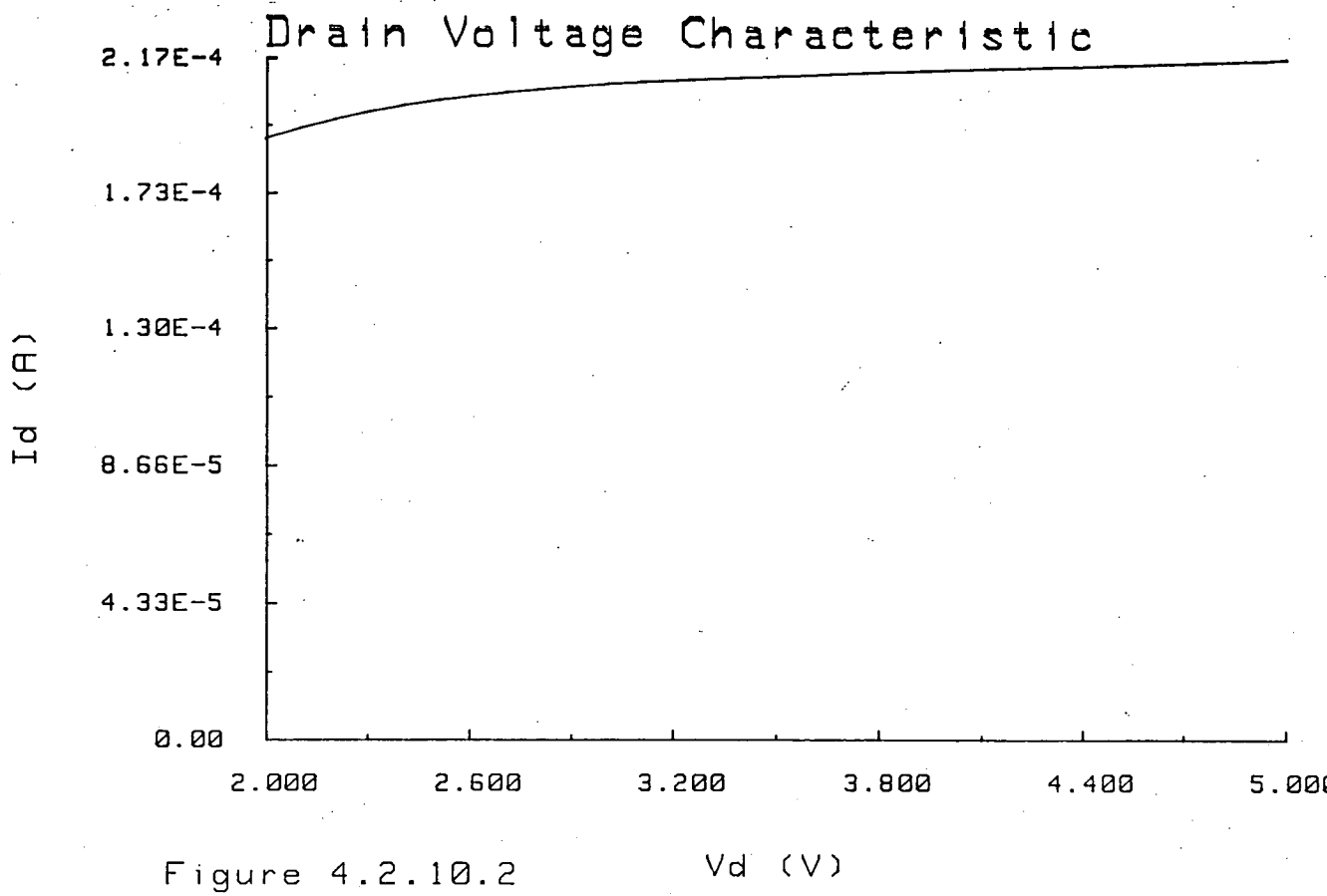


Figure 4.2.10.2 V_d (V)

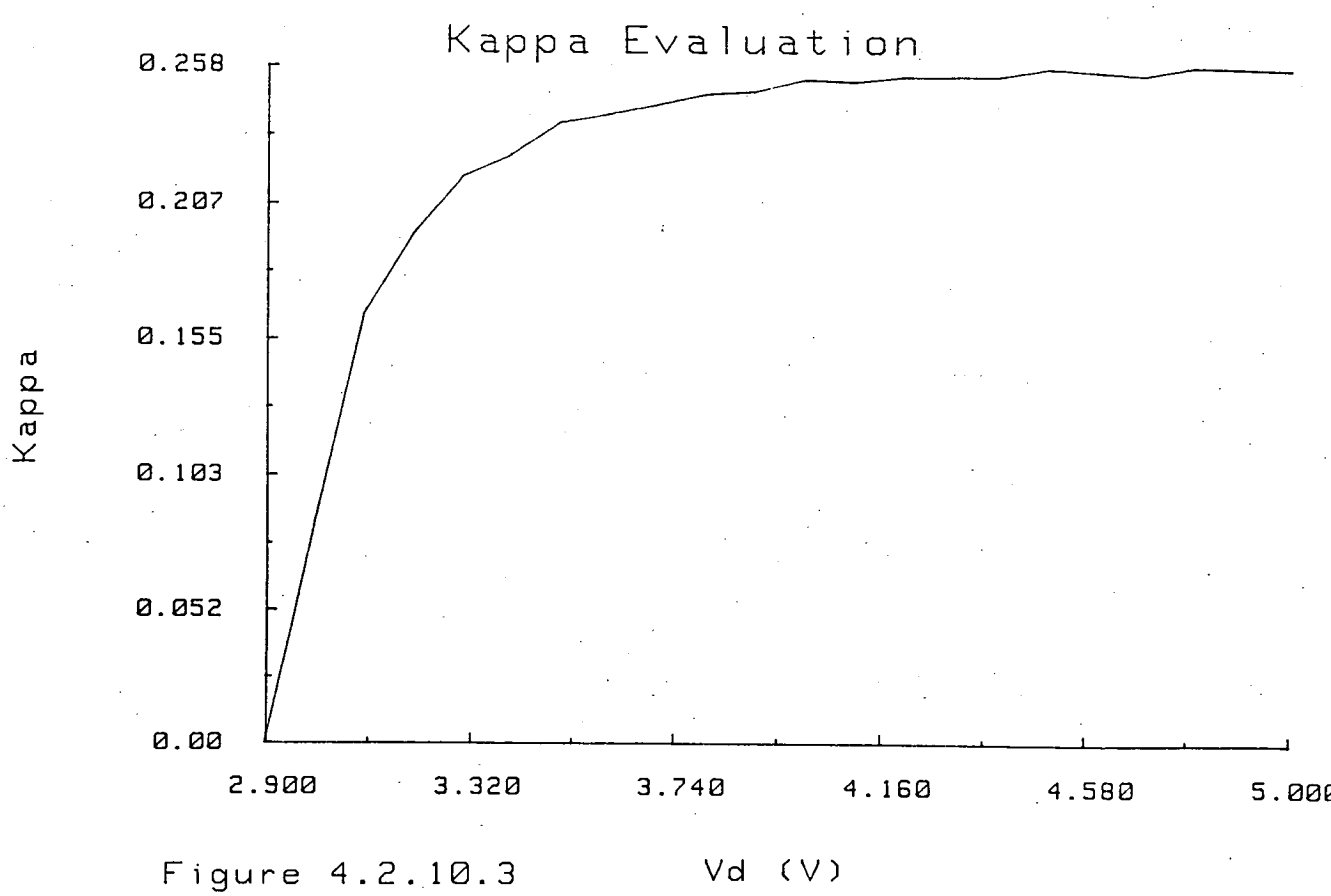


Figure 4.2.10.3 V_d (V)

current, I_d are compared to find the change in length, L_{del} using

$$L_{del} = L \left[1 - \frac{I_{dsat}}{I_d} \right] \quad 4.2.10.1$$

If as in some cases, I_{dsat} is greater than I_d then for that portion of the measured curve, $L_{del}=0$ since a negative value of L_{del} is unrealistic. The slope coefficient, κ for various drain values is calculated (figure 4.2.10.3) from the L_{del} values using

$$\kappa = \frac{L_{del}^2}{X_d^2 (V_d - V_{dsat})} \quad 4.2.10.2$$

It is found that κ almost becomes a constant at higher drain voltages. Therefore, the value of κ found at the largest normal operating V_d is chosen as the parameter.

4.3 Simulation of Characteristics

4.3.1 NMOS Enhancement Results

In the Parameter Extraction Section (4.2), the graphs illustrated the extraction carried out using an NMOS $5\mu\text{m} \times 5\mu\text{m}$ enhancement device. The level 1 and level 3 parameters which result are listed in table 4.3.1.1.

In order to test the accuracy of these parameters for circuit simulation purposes, measured and simulated characteristics from different regions of device operation were compared. In the figures which follow, the solid lines are the measured characteristics and the dashed lines are simulated characteristics. Figures 4.3.1.1 and 4.3.1.2 show the results for drain voltage characteristics and gate voltage characteristics respectively using the level 1 model. The complete set of error figures for the drain voltage characteristics are shown in Table 4.3.1.2. PARAMEX yields all these figures when an error analysis is requested. It can be seen that level 1 provides only a rough approximation of the measured curves. One of the major inadequacies of the model is the lack of any sort of variation of mobility with drain voltage. The mobility variation parameters were evaluated when the drain voltage was 2V and this can be seen in the simulation.

The average percentage error in the gate voltage curve when $V_b=0$ (figure 4.3.1.2) is 9.8%, most of which is due to the large errors around threshold voltage. At

Table 4.3.1.1 NMOS Enhancement $5\mu m \times 5\mu m$ Parameters

	Level 1		Level 3	
Type	1		1	
Dep	0		0	
t_{ox}	558	A	558	A
x_j			1	μm
N_{fs}			1.43×10^{15}	m^{-2}
V_{to}	0.81	V	0.81	V
γ	0.79	$V^{\frac{1}{2}}$	1.39	$V^{\frac{1}{2}}$
L_d	1.06	μm	1.06	μm
Δ_w	0.84	μm	0.84	μm
μ_o	360	$cm^2 V_s^{-1}$	590	$cm^2 V_s^{-1}$
θ	-0.031	V^{-1}	0.037	V^{-1}
v_{max}			3.51×10^5	$m s^{-1}$
η			0.037	
δ			0.211	
κ			0.257	

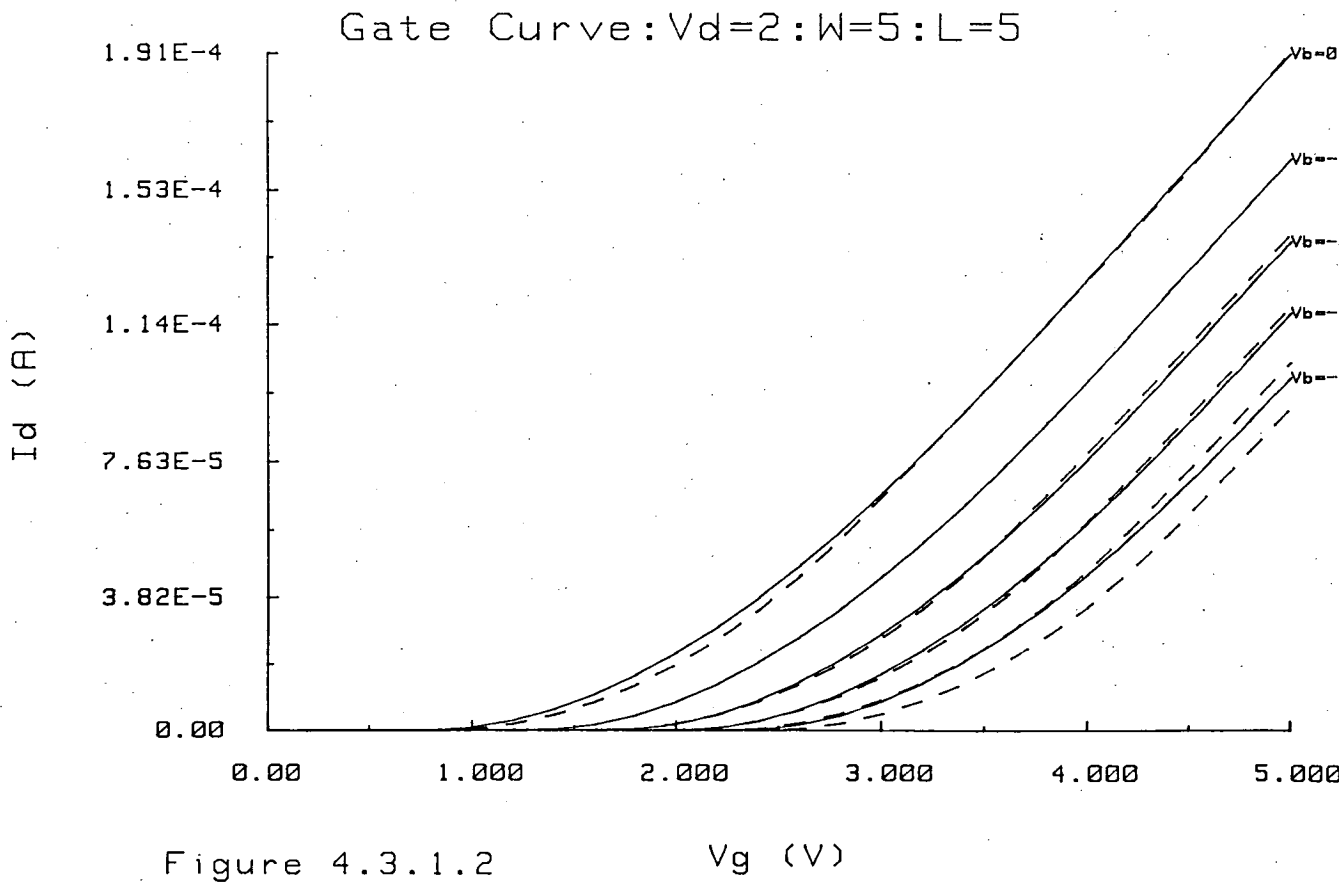
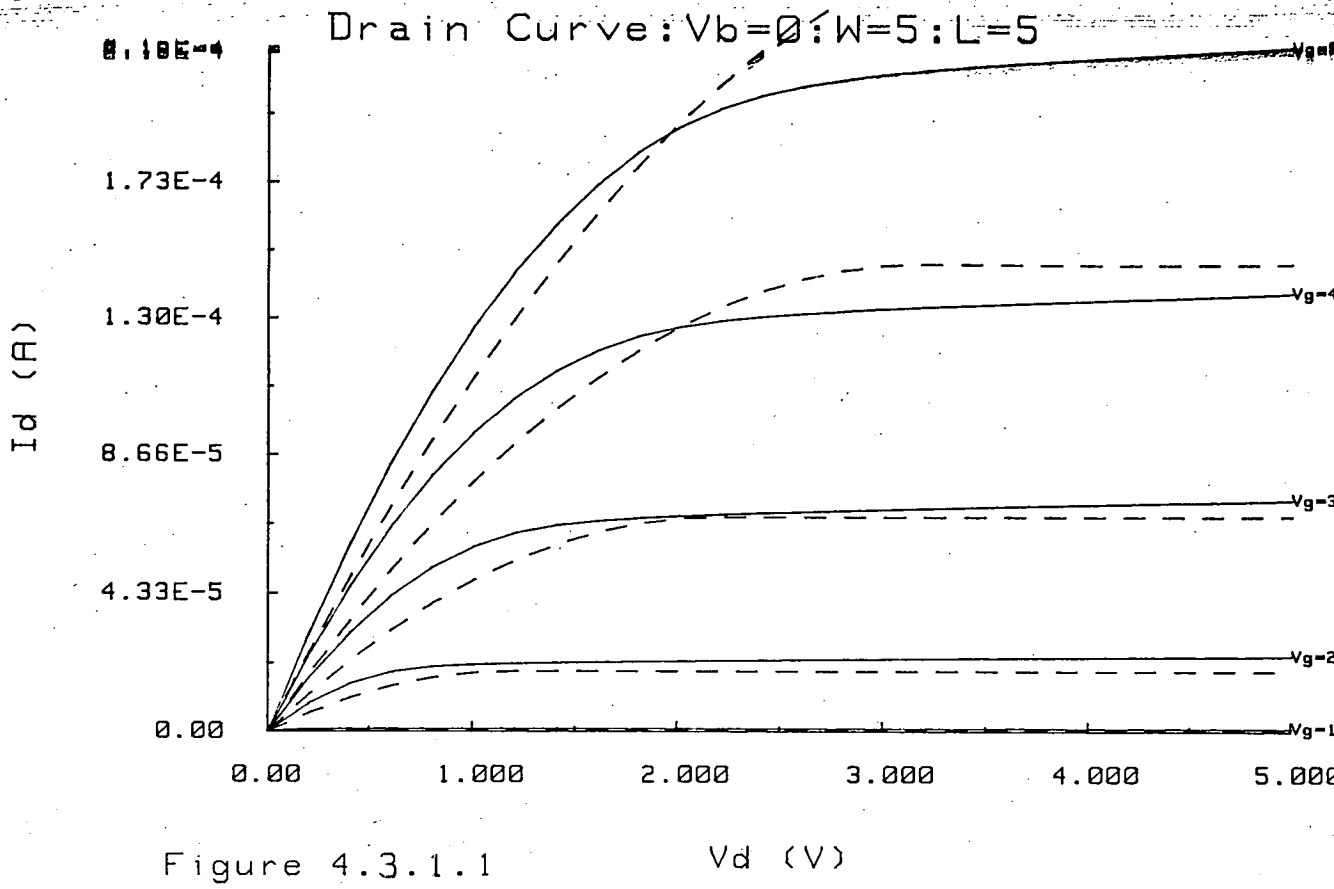


Table 4.3.1.2 Level 1 Error figures for NMOS Enhancement Device

V_g V	Average r.m.s. Error (X 10 ⁻⁶)	Max Absolute Error (X 10 ⁻⁶)	Average Percentage Error %
1	0.543	0.628	54
2	3.74	4.80	18
3	5.37	11.2	8.9
4	11.0	15.4	10
5	31.8	49.9	15

Table 4.3.1.3 Level 3 Error figures for NMOS Enhancement Device

V_g V	Average r.m.s. Error (X 10 ⁻⁶)	Max Absolute Error (X 10 ⁻⁶)	Average Percentage Error %
1	0.110	0.220	10.8
2	0.536	1.45	2.0
3	1.40	2.35	1.9
4	1.54	2.55	1.3
5	1.36	2.80	0.8

non-zero substrate biases the fit is not good because of the fact that threshold voltage is assumed to vary linearly with the square root of substrate bias. For a non-uniformly doped channel, this is not the case. The result can be seen: the simulation when $V_b = -1$ almost coincides with the measured curve when $V_b = -2$.

The extra factors included in the level 3 model result in much more accurate simulation as shown in figures 4.3.1.3, 4.3.1.4, 4.3.1.5 and 4.3.1.6. In this case the error figures for the drain voltage characteristics (figure 4.3.1.3) are shown in Table 4.3.1.3.

The gate voltage characteristics (figure 4.3.1.4) reveal larger errors. The main reason for this is that the mobility variation with gate voltage changes with substrate bias. At $V_b = -4$ mobility degrades with gate voltage more rapidly than the simulation predicts. When the drain voltage increases to 5V (figure 4.3.1.5) so that the device is in saturation, the simulation is good; when $V_b = 0$, the maximum absolute error is $2.83 \mu\text{A}$ and when $V_b = -4$, it is $3.15 \mu\text{A}$.

Figure 4.3.1.6 shows a typical prediction of device operation in subthreshold. The transition at threshold is not modelled well although the parameter N_f , leads to an accurate simulation of the slope in subthreshold. However, since current varies logarithmically with gate voltage, the simulated current can easily be an order of magnitude out. The average percentage error for $V_b = 0$ is 24.5%.

The techniques can be readily extended to smaller geometries. Figures 4.3.1.7 and 4.3.1.8 show the measured and level 3 simulated characteristics for a drawn $1.5\mu\text{m} \times 1.5\mu\text{m}$ device. The parameters which are listed in Table 4.3.1.4, show that the actual width of the device is less than $1\mu\text{m}$ and the actual length is about $1.15\mu\text{m}$. The average percentage errors and maximum absolute errors for figure 4.3.1.7 are given in Table 4.3.1.5. Figure 4.3.1.8 again demonstrates how the relationship between carrier mobility and gate voltage changes as a substrate bias is applied; this effect is not included in the model. When the substrate bias is 0V, the average percentage error is 4.7%.

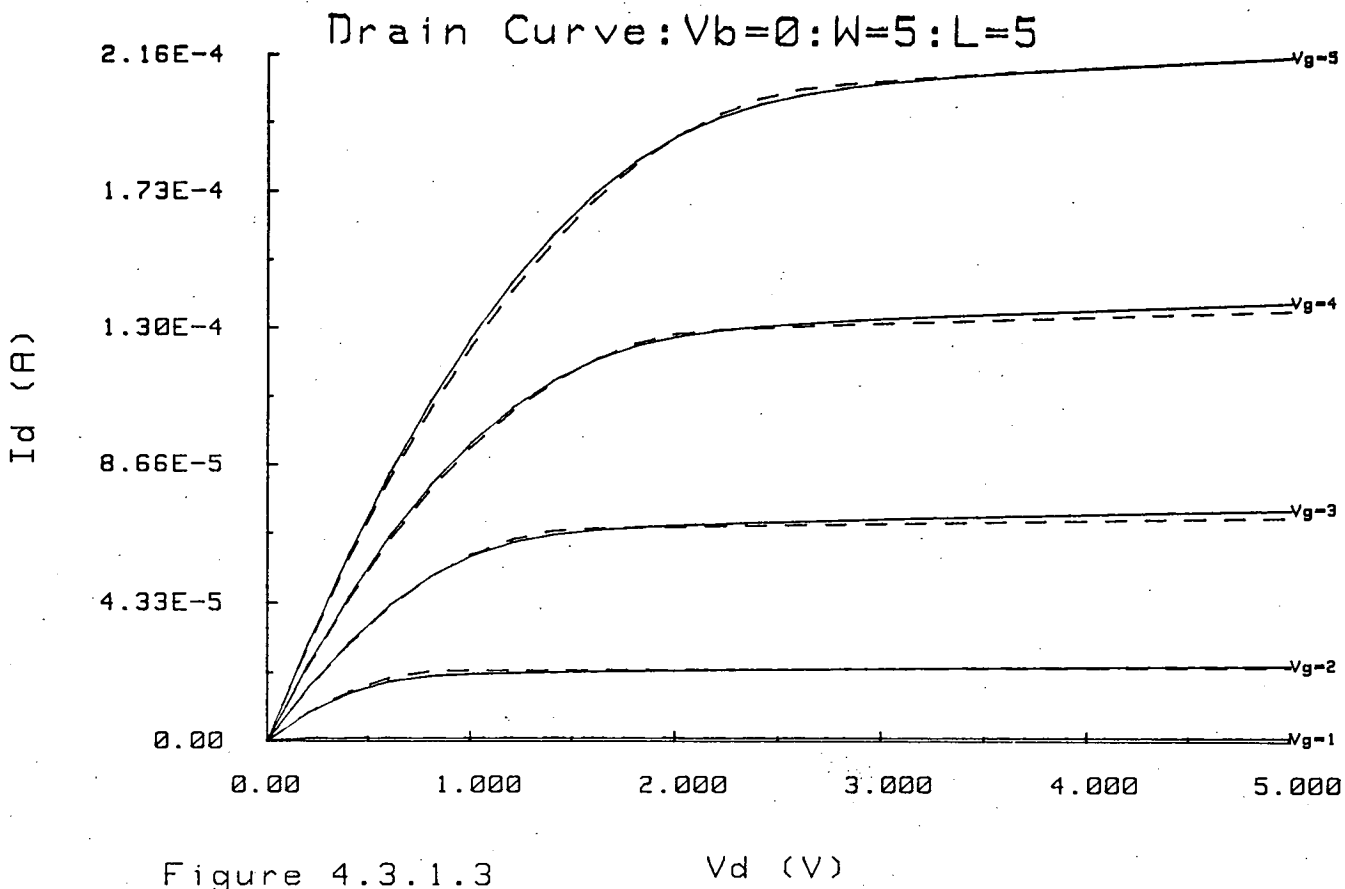


Figure 4.3.1.3

V_d (V)

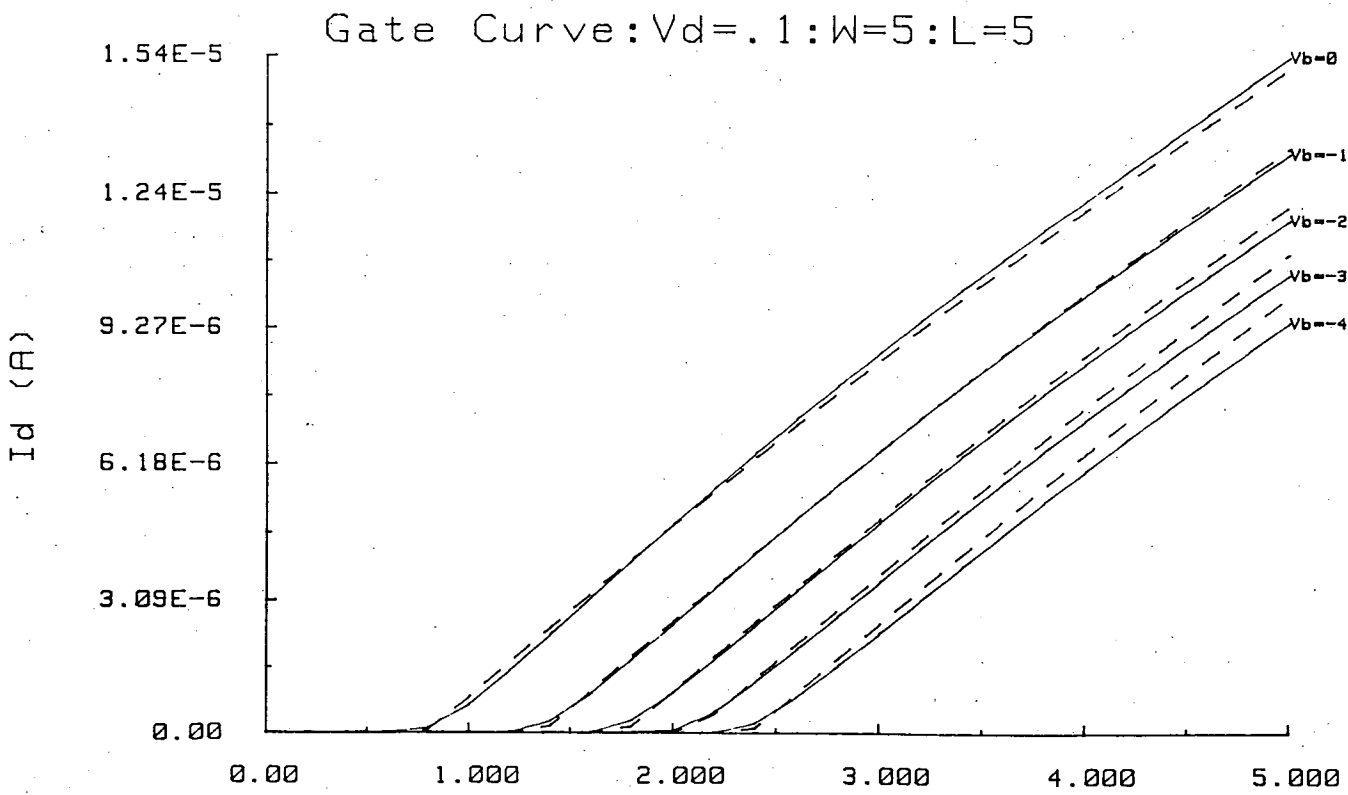


Figure 4.3.1.4

V_g (V)

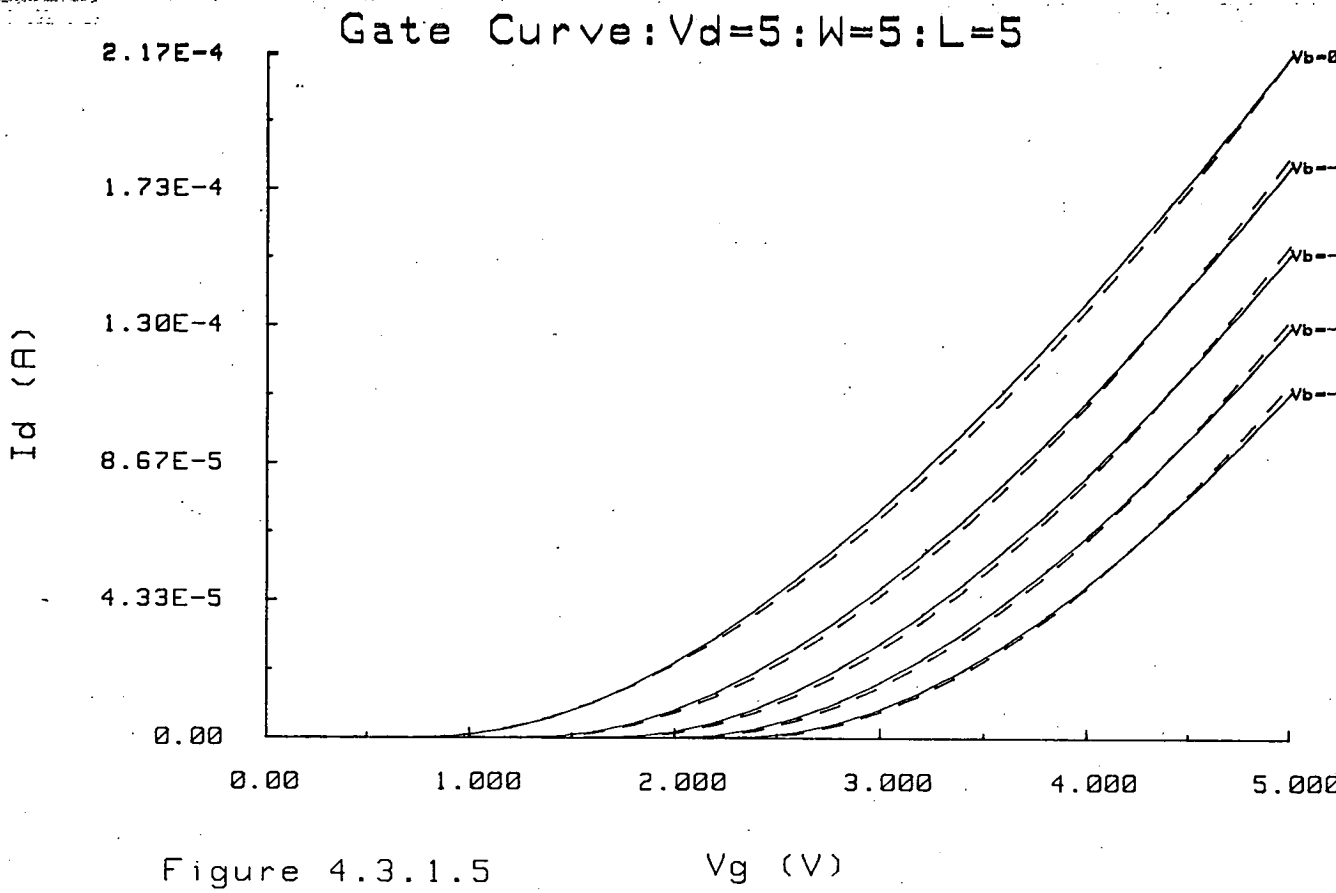


Figure 4.3.1.5 V_g (V)

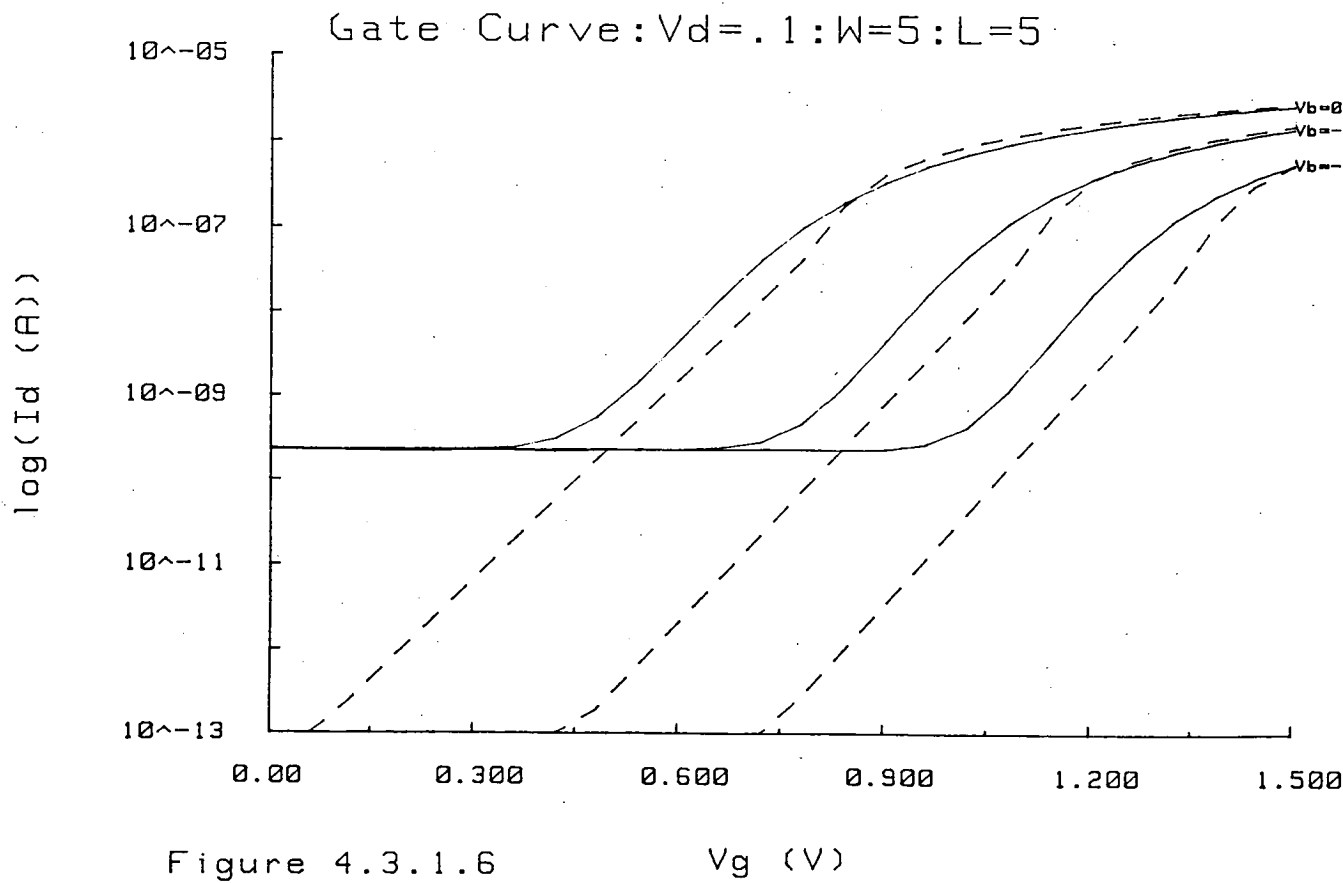


Figure 4.3.1.6 V_g (V)

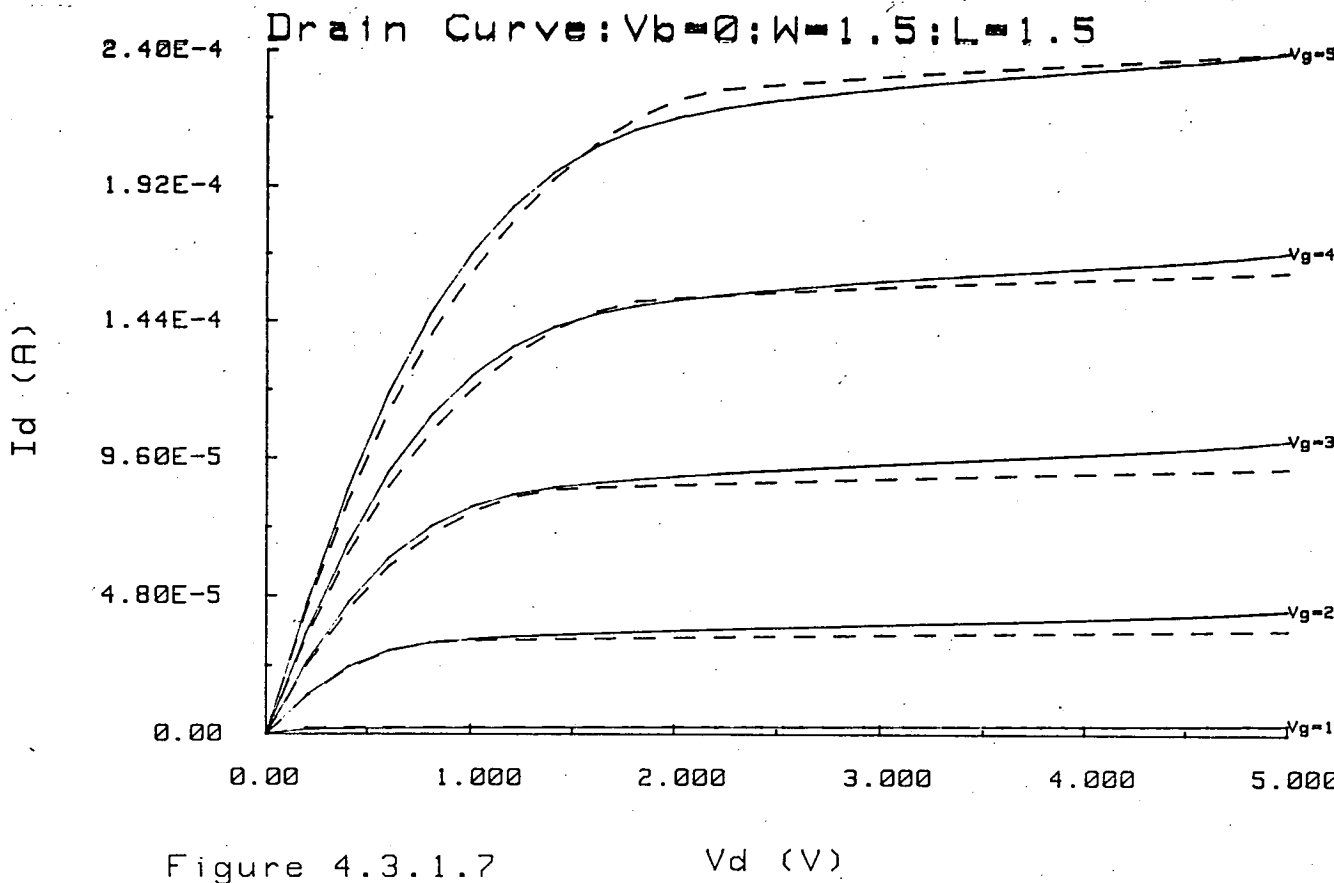


Figure 4.3.1.7

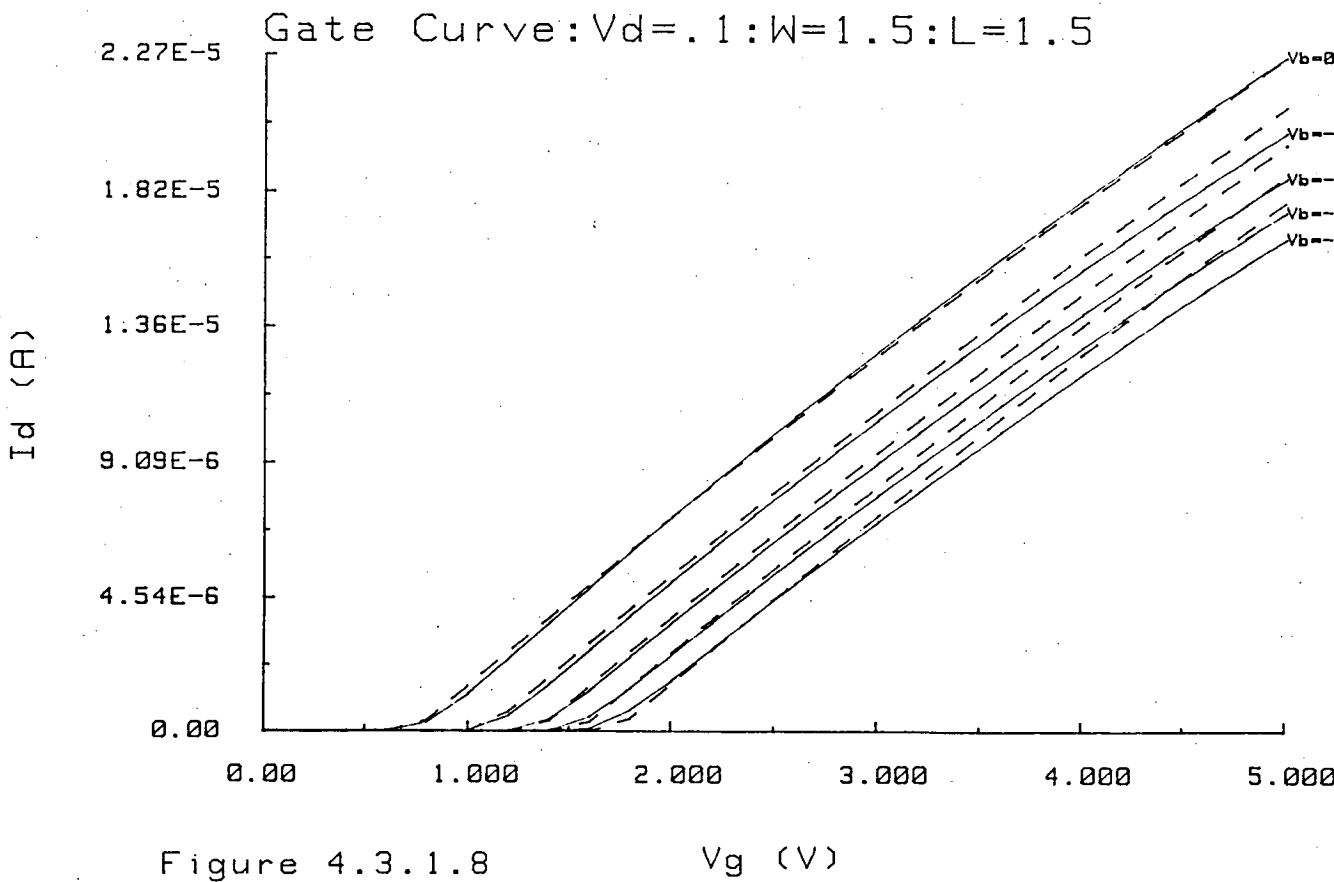


Figure 4.3.1.8

Table 4.3.1.4 NMOS Enhancement $1.5\mu m \times 1.5\mu m$ Parameters

Type	1	
Dep	0	
t_{ox}	250	Å
x_j	0.3	μm
N_{fs}	7.47×10^{15}	m^{-2}
V_{to}	0.73	V
γ	0.88	$V^{\frac{1}{2}}$
L_d	0.17	μm
Δ_w	0.33	μm
μ_o	610	$cm^2 V_s^{-1}$
θ	0.029	V^{-1}
v_{max}	1.82×10^5	$m s^{-1}$
η	0.029	
δ	0.261	
κ	0.088	

Table 4.3.1.5 Level 3 Error figures for Small Geometry NMOS

V_g	Maximum Absolute Error ($\times 10^{-6}$)	Average Percentage Error %
1	0.45	5.3
2	6.9	8.0
3	9.8	5.1
4	7.1	2.5
5	7.6	2.2

4.3.2 Depletion Device Simulation

Although depletion devices do not operate entirely in accordance with the SPICE MOS enhancement models, the models are frequently used to simulate depletion devices when NMOS circuits are being developed. There is no depletion device model in SPICE and so these extraction techniques have been used to reveal some aspects of depletion device operation which are different from those of enhancement devices. The simulation results, which are presented below, illustrate how accurately the designer can expect depletion devices to be modelled in SPICE.

As was explained in section 4.2.1, the extraction of depletion device threshold is complicated by the fact that mobility increases as the gate voltage increases from threshold to 0V. This is due to the diminishing transverse electric field. A technique was developed whereby the second derivative is used to pick out the turn-on point and so threshold can be deduced. This was illustrated in Section 4.2.1.

The variation of carrier mobility is shown in figure 4.3.2.1. The resulting values, deduced using figure 4.3.2.2, are $\mu_o = 0.026 \text{ m}^2 \cdot \text{Vs}^{-1}$ and $\theta = -0.063 \text{ V}^{-1}$. The low value of μ_o is the result of the high impurity concentration and the transverse electric field. Subsequently mobility increases as the absolute gate voltage and transverse electric field decreases and therefore θ is negative (figure 4.3.2.2).

Modulation of carrier mobility with drain voltage, which is shown in figures 4.3.2.3 and 4.3.2.4 is very similar to the enhancement device apart from the lower values of carrier mobility. There is a reduction in carrier mobility until saturation is reached.

The complete set of Level 3 parameters for the $6\mu\text{m} \times 6\mu\text{m}$ NMOS depletion device used in the example above is in Table 4.3.2.1. The device did not turn off at $V_b = 0$ and so the parameters derived in the subthreshold region, N_{fs} and η could not be extracted.

A typical comparison of measured and simulated depletion device

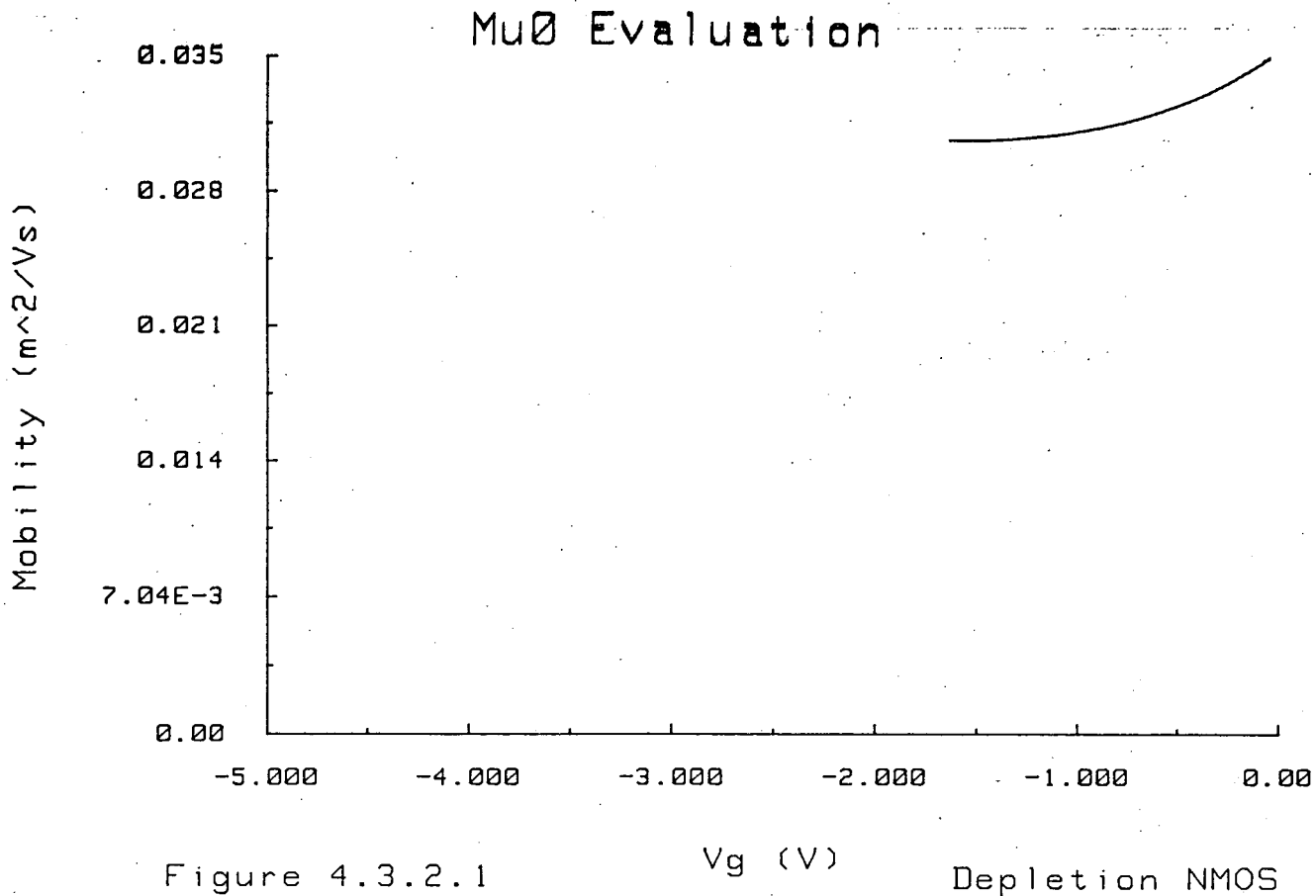


Figure 4.3.2.1

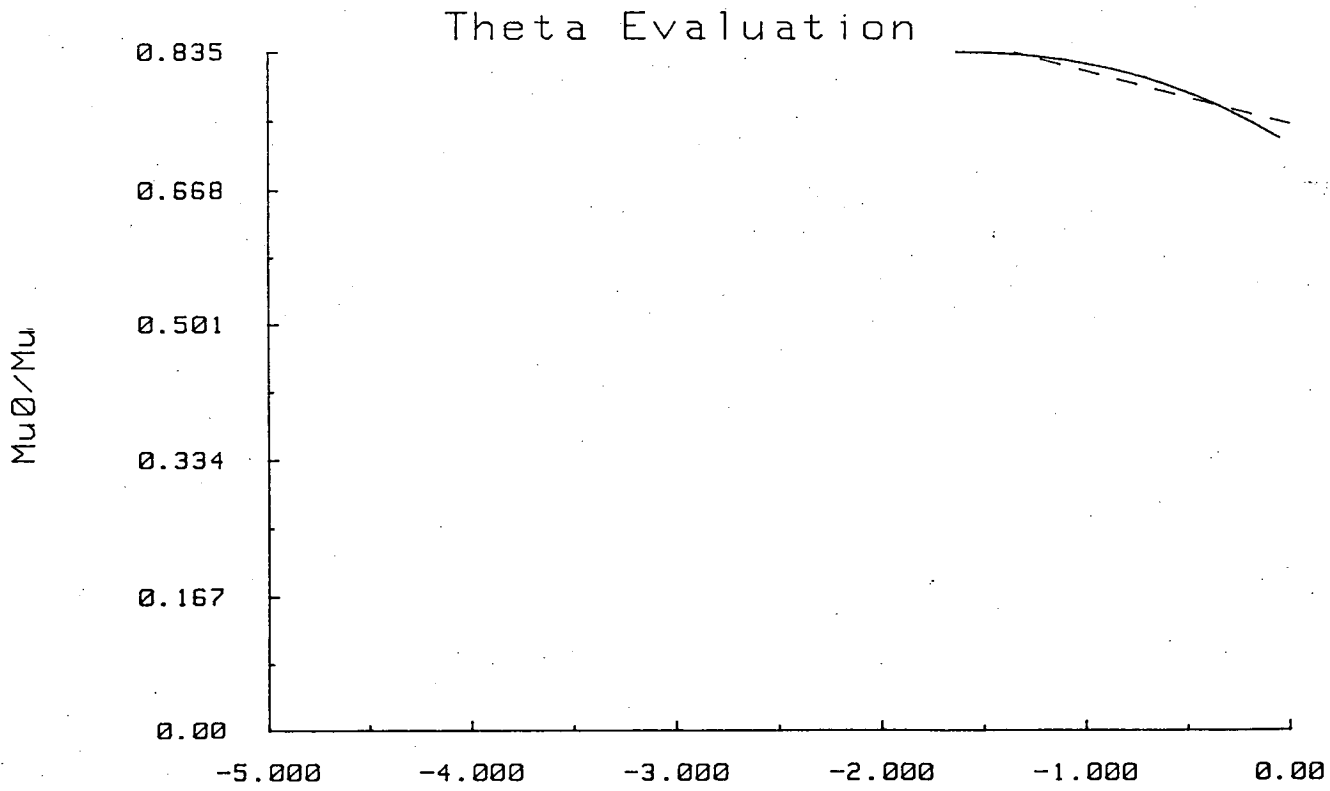


Figure 4.3.2.2

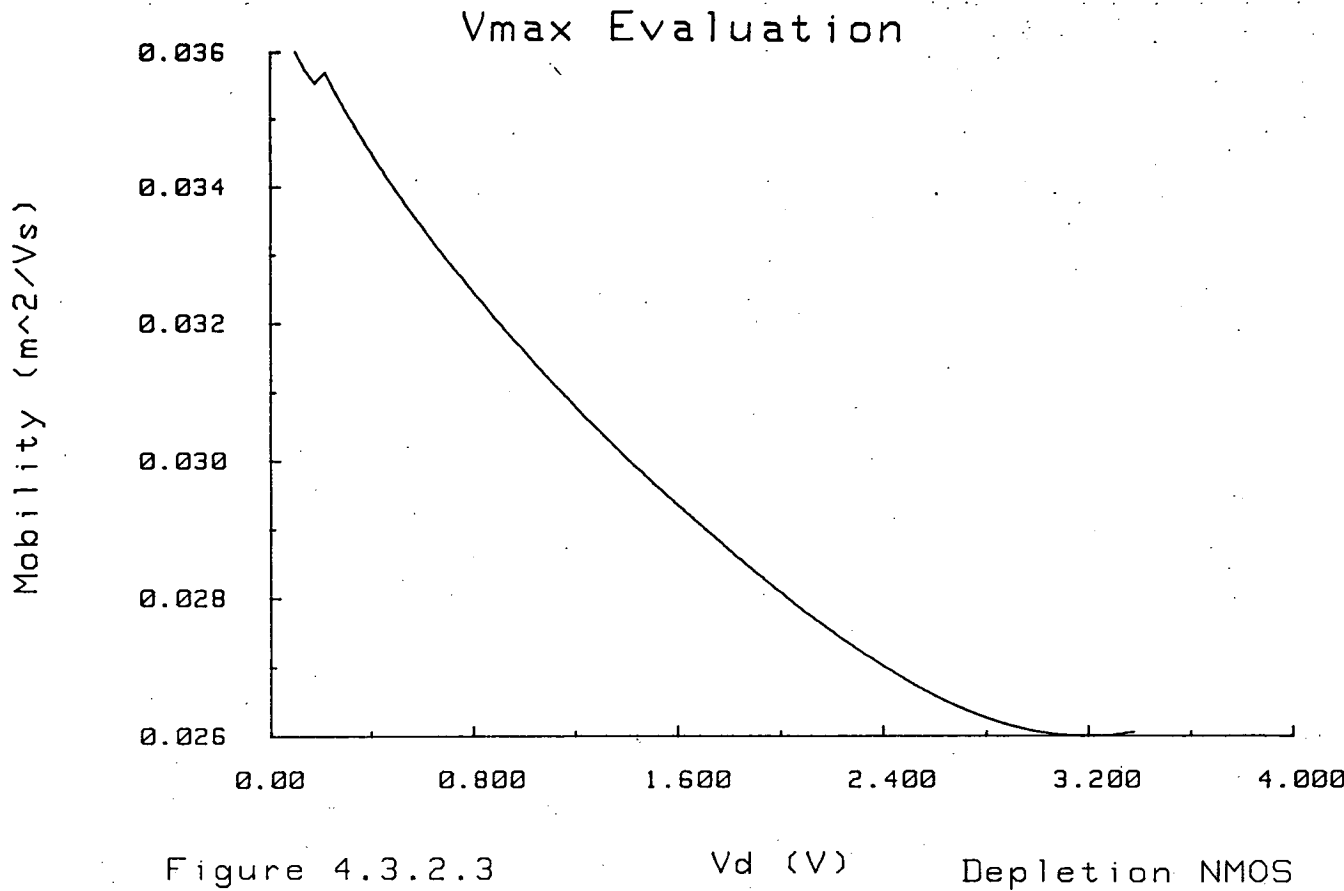


Figure 4.3.2.3 V_d (V) Depletion NMOS

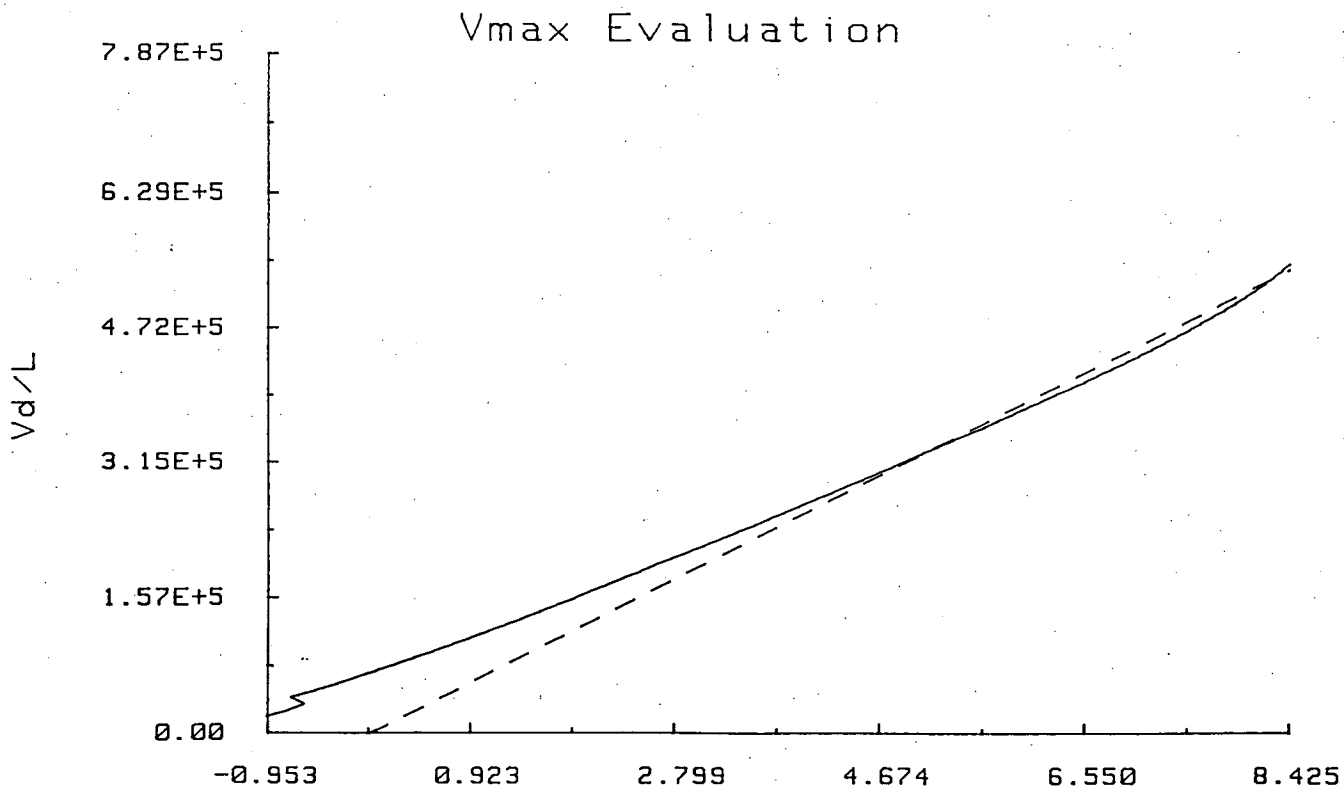


Figure 4.3.2.4 $1/\mu_{eff} - 1/\mu_s$ Depletion NMOS

Table 4.3.2.1 NMOS Depletion $6\mu m \times 6\mu m$ Parameters

<i>Type</i>	1	
<i>Dep</i>	1	
t_{ox}	250	Å
x_j	0.3	μm
N_{fs}	-	
V_{to}	-3.96	V
γ	0.57	$V^{\frac{1}{2}}$
L_d	0.46	μm
Δ_w	0.78	μm
μ_o	260	$cm^2 V_s^{-1}$
θ	-0.063	V^{-1}
v_{max}	6.43×10^4	$m s^{-1}$
η	-	
δ	0	
κ	0.760	

characteristics is shown in figure 4.3.2.5. The inaccuracy of the model can be seen more readily in the gate voltage characteristic in figure 4.3.2.6. The model takes no account of the fact that the device has a leakage current. However by the use of the second derivative method for extracting threshold voltage and a negative θ to allow for the increasing carrier mobility, an accurate representation of device operation is available at gate voltages above $-2V$. The maximum absolute error for figure 4.3.2.6 is $5.7\mu A$, which occurs before the model reaches its threshold voltage. In figure 4.3.2.5, the average percentage errors when $V_g = -1$ and 0 are 0.9% and 0.5% respectively.

4.3.3 P-channel Device Simulation

The application of the SPICE level 3 model to p-channel devices is very similar to its application to n-channel devices. As was mentioned in Section 2.6, Wright²⁵ points out that the relationship between carrier mobility and drain voltage is based on the relationship between carrier velocity and field in p-channel devices rather than n-channel devices. Below are some examples of extraction and simulation using a PMOS $5\mu m \times 5\mu m$ device. The graphs in figures 4.3.3.1 and 4.3.3.2 demonstrate how threshold is deduced using the second derivative to identify the turn on point. The mobility versus gate voltage graphs, shown in figures 4.3.3.3 and 4.3.3.4, show that mobility goes down as the transverse electric field increases and the mobility of holes is significantly lower than electrons; $\mu_o = 170 cm^2.Vs^{-1}$. The low mobility means that carrier velocity does not reach the saturation velocity under normal applied voltages at $5\mu m$ channel lengths. Figure 4.3.3.5 indicates that there is no significant carrier velocity saturation and consequently v_{max} is set to zero so that the effect is not modelled.

The parameters which result for the p-channel devices are listed in Table 4.3.3.1 and the simulated device characteristics are compared with the measured device characteristics in figures 4.3.3.6, 4.3.3.7, 4.3.3.8 and 4.3.3.9. The error figures for the characteristics in figures 4.3.3.5 and 4.3.3.6 are given in Tables 4.3.3.2 and 4.3.3.3 respectively. Again the relationship between carrier mobility and substrate bias is not accounted for and the effect of this can be seen in figure 4.3.3.7. The same effect is evident in figure 4.3.3.8 where V_d has been set at $-5V$ to examine the accuracy with which the device can be simulated in saturation. The average percentage errors when $V_b = 0$ and $V_b = 1$ are 2.7% and 4.4% respectively. Finally the subthreshold region is

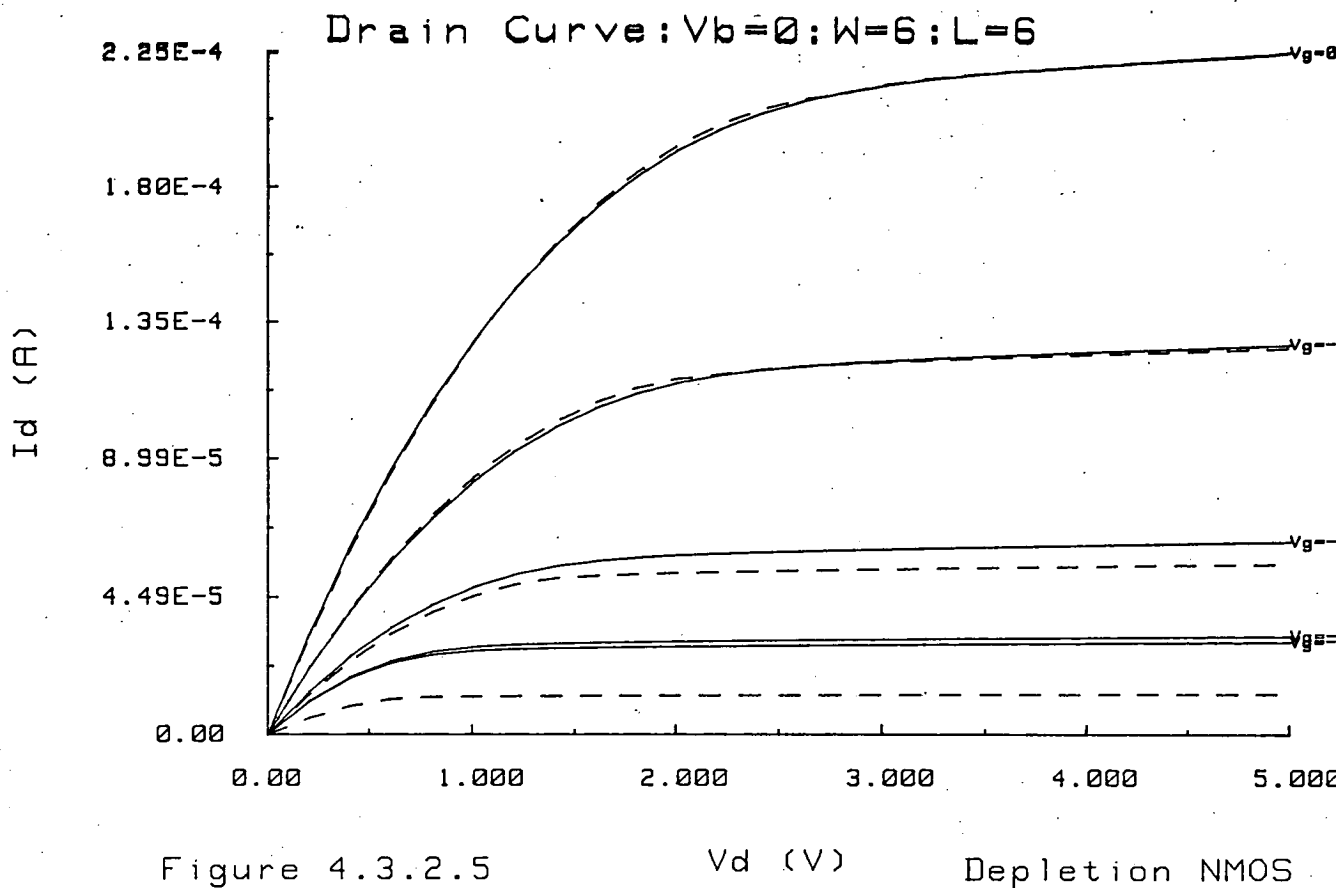


Figure 4.3.2.5 $V_d (V)$ Depletion NMOS

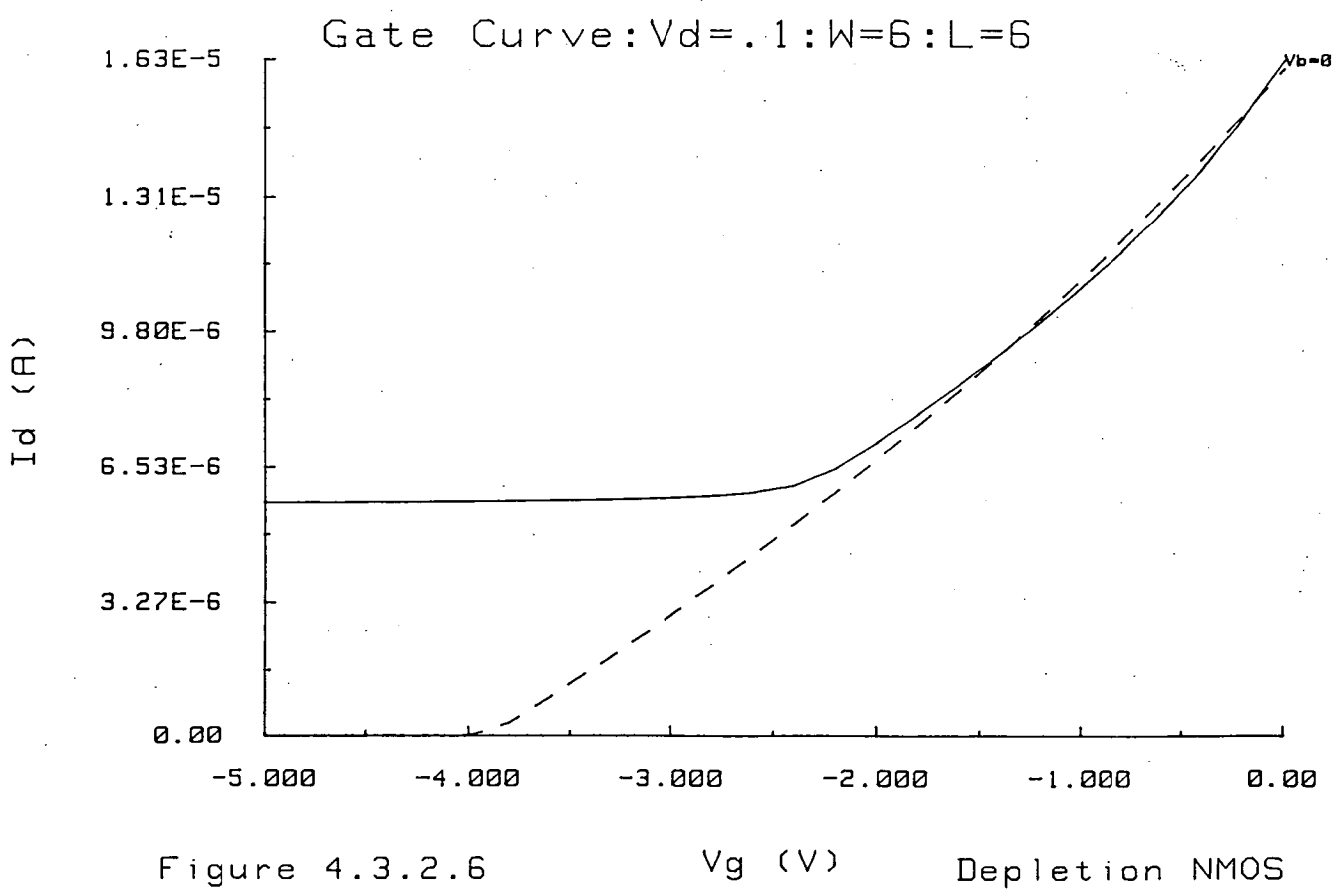


Figure 4.3.2.6 $V_g (V)$ Depletion NMOS

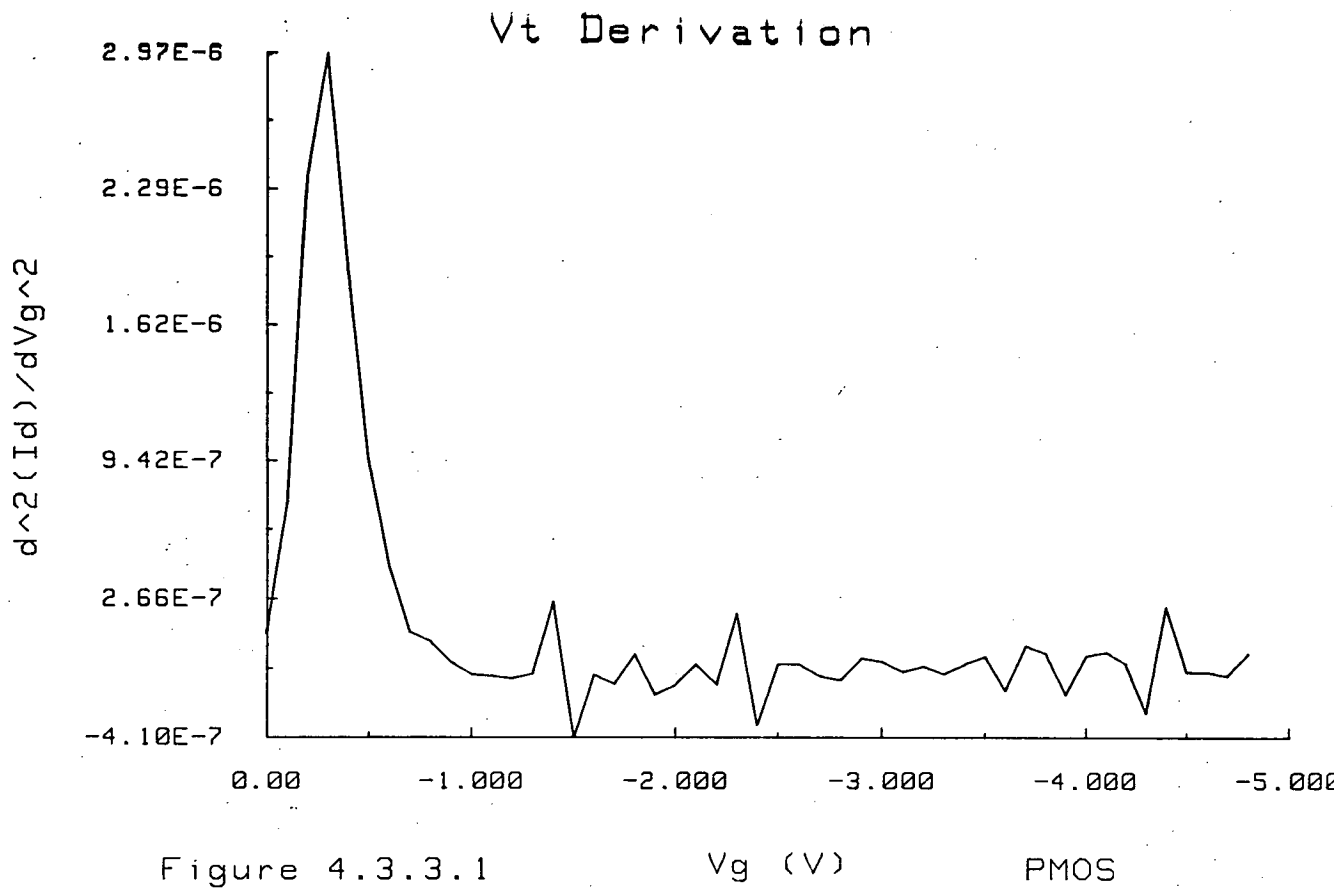


Figure 4.3.3.1 V_g (V) PMOS

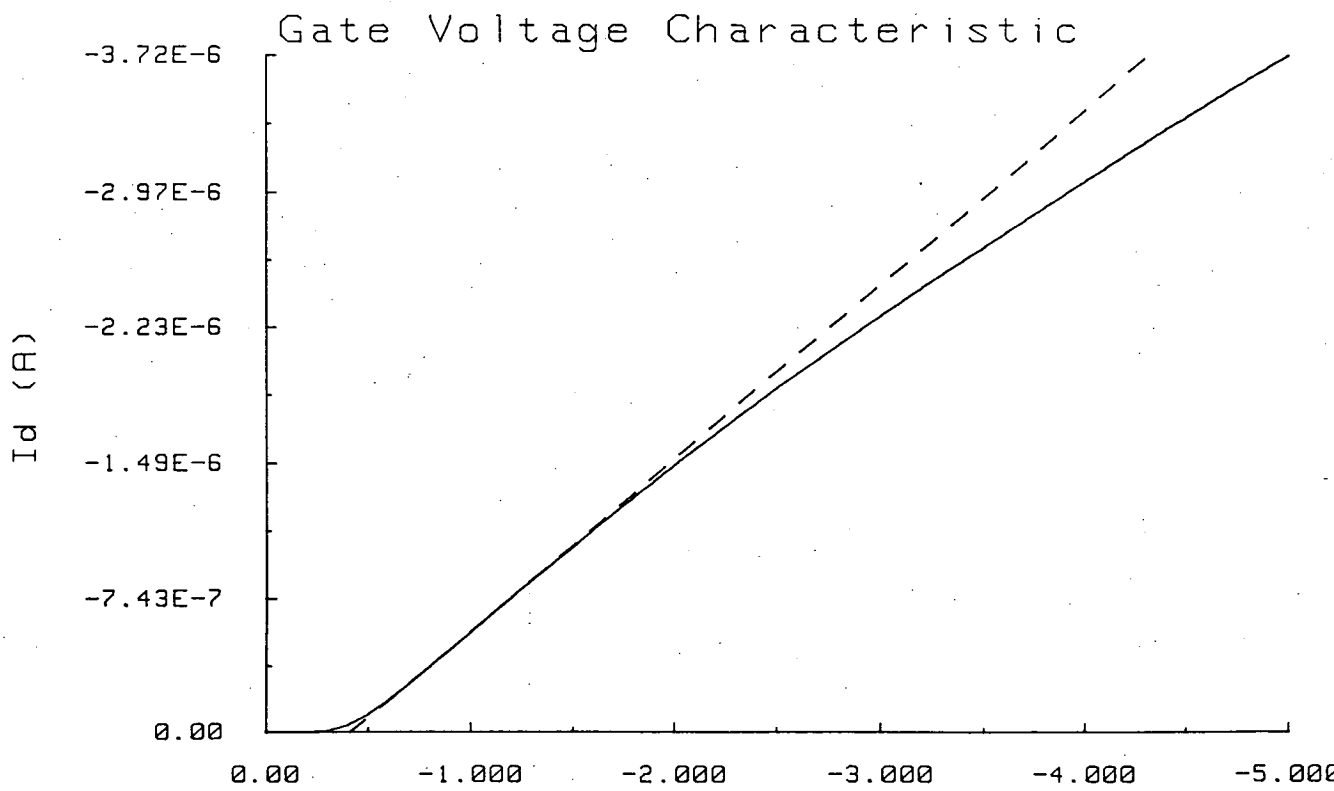


Figure 4.3.3.2 V_g (V) PMOS

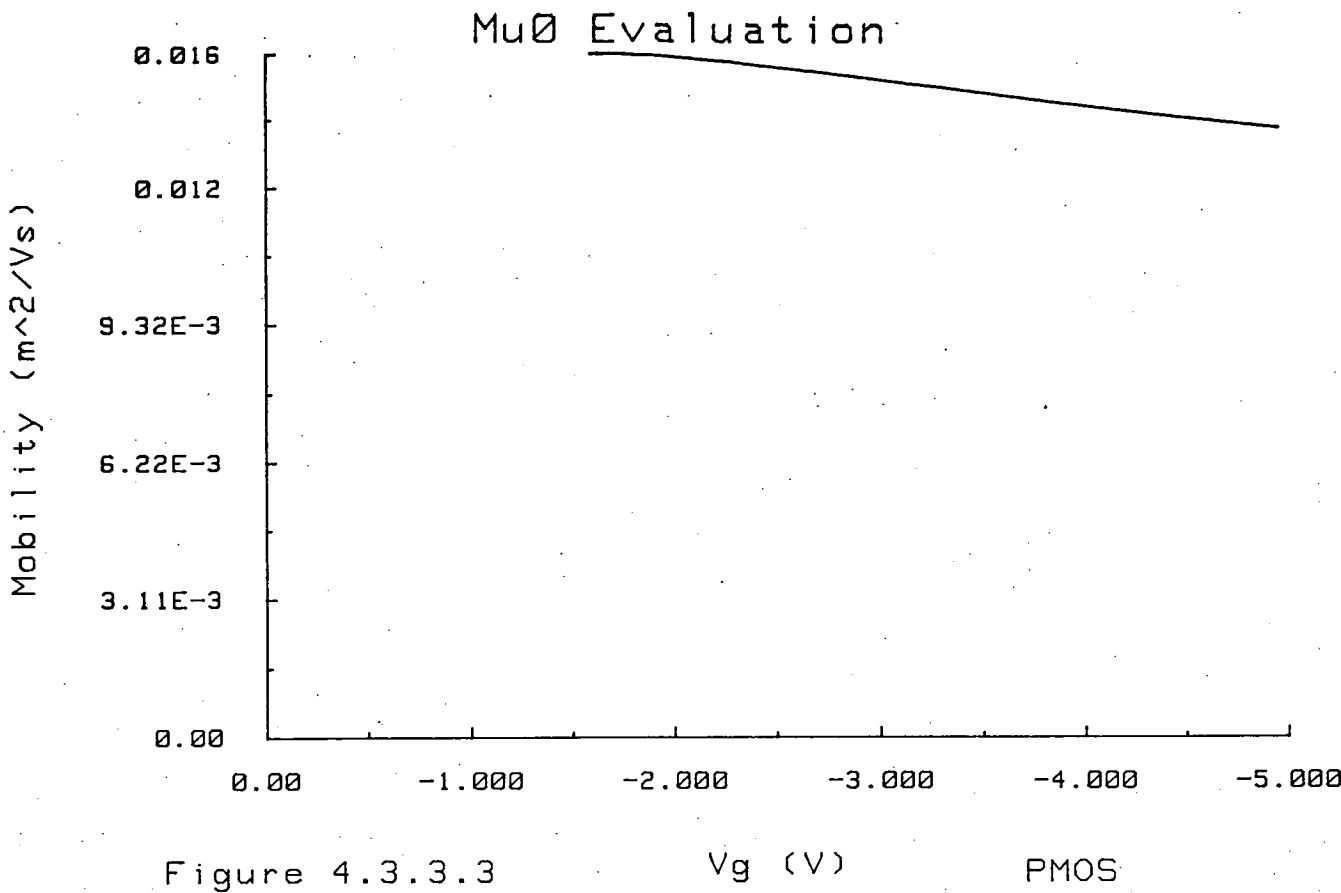


Figure 4.3.3.3 Vg (V) PMOS

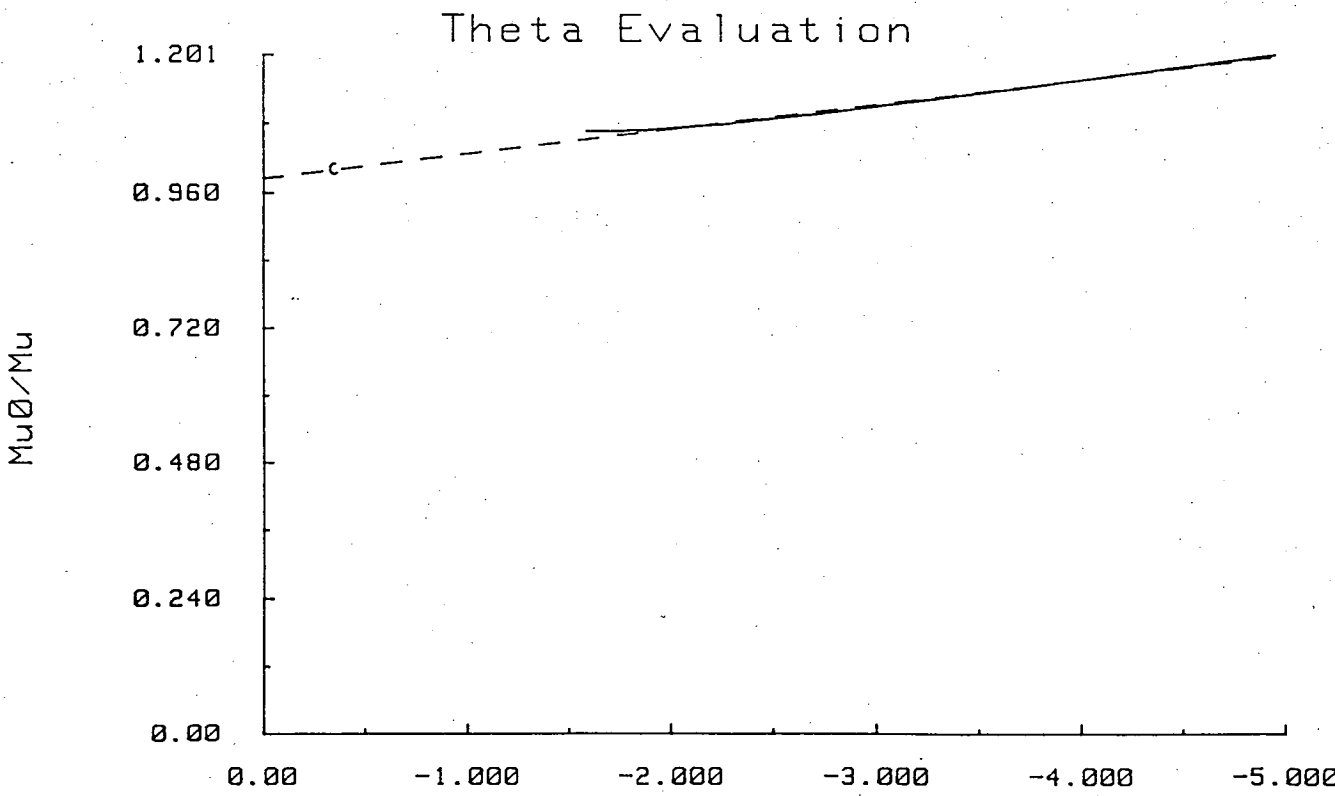


Figure 4.3.3.4 Vg (V) PMOS

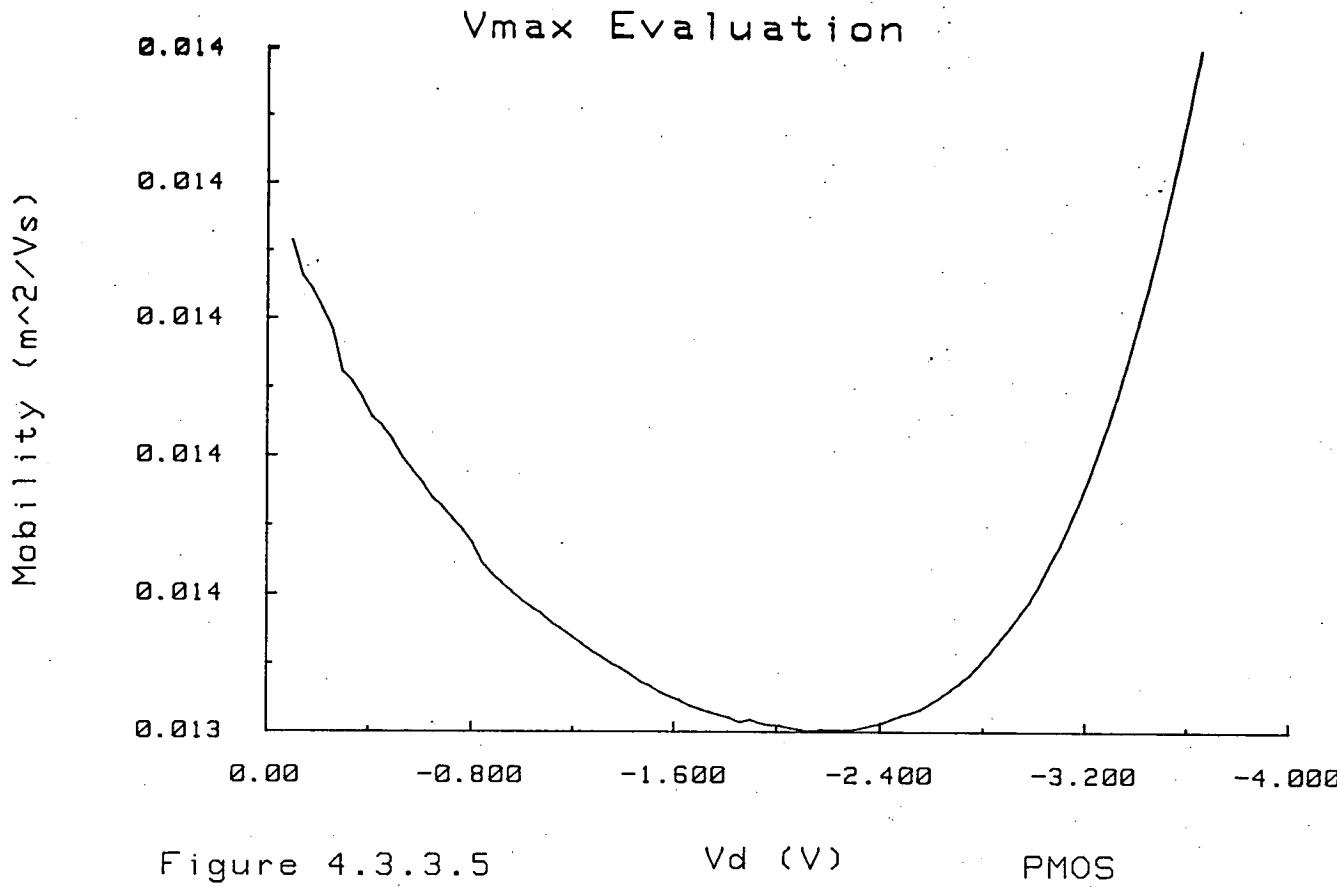


Figure 4.3.3.5 V_d (V) PMOS

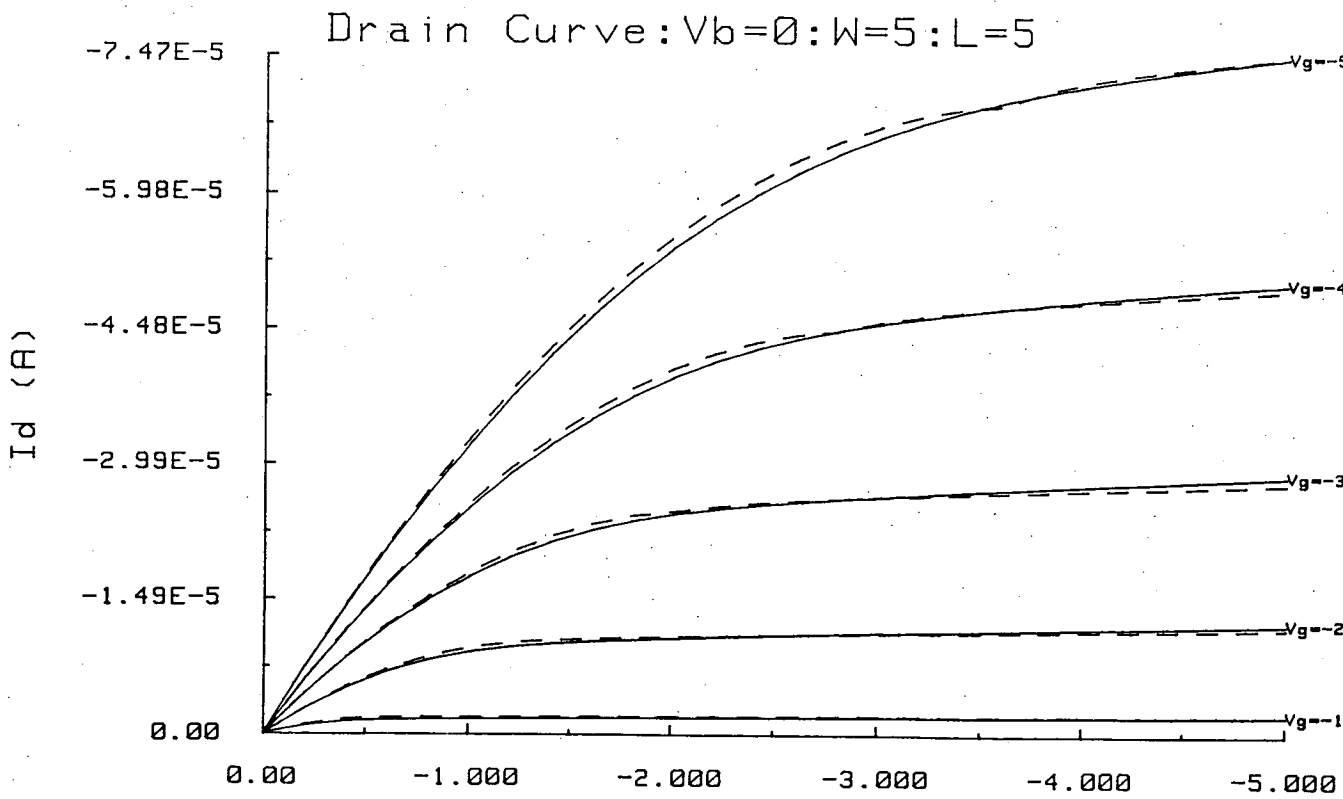


Figure 4.3.3.6 V_d (V) PMOS

Table 4.3.3.1 PMOS Enhancement $5\mu m \times 5\mu m$ Parameters

<i>Type</i>	-1	
<i>Dep</i>	0	
<i>t_{ox}</i>	683	Å
<i>x_j</i>	1	μm
<i>N_{fs}</i>	3.04×10^{15}	
<i>V_{to}</i>	-0.35	V
γ	0.96	$V^{\frac{1}{2}}$
<i>L_d</i>	1.00	μm
Δ_w	0.80	μm
μ_o	170	$cm^2 V_s^{-1}$
θ	0.043	V^{-1}
<i>v_{max}</i>	0	$m s^{-1}$
η	0.067	
δ	0.260	
κ	0.180	

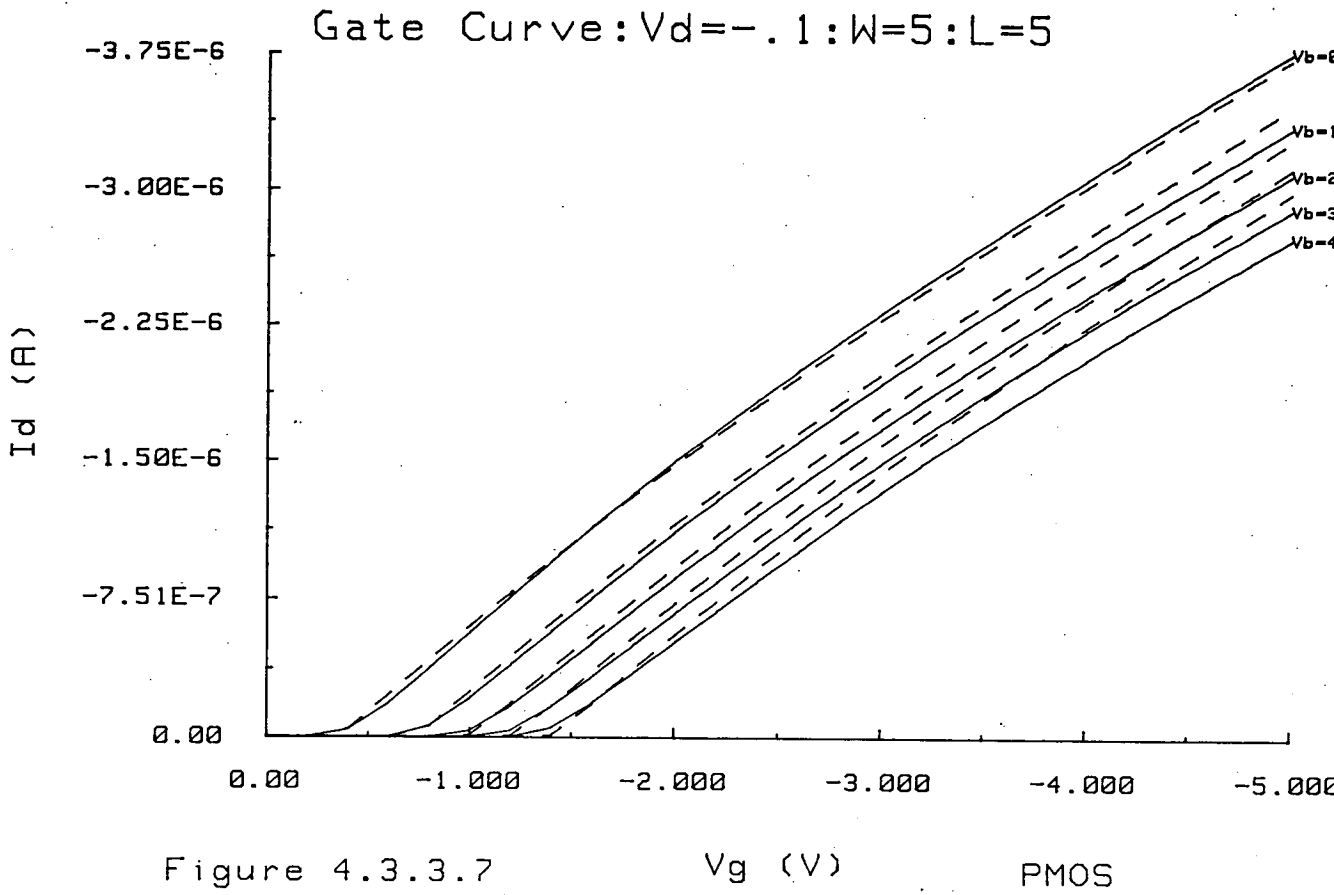


Figure 4.3.3.7

V_g (V)

PMOS

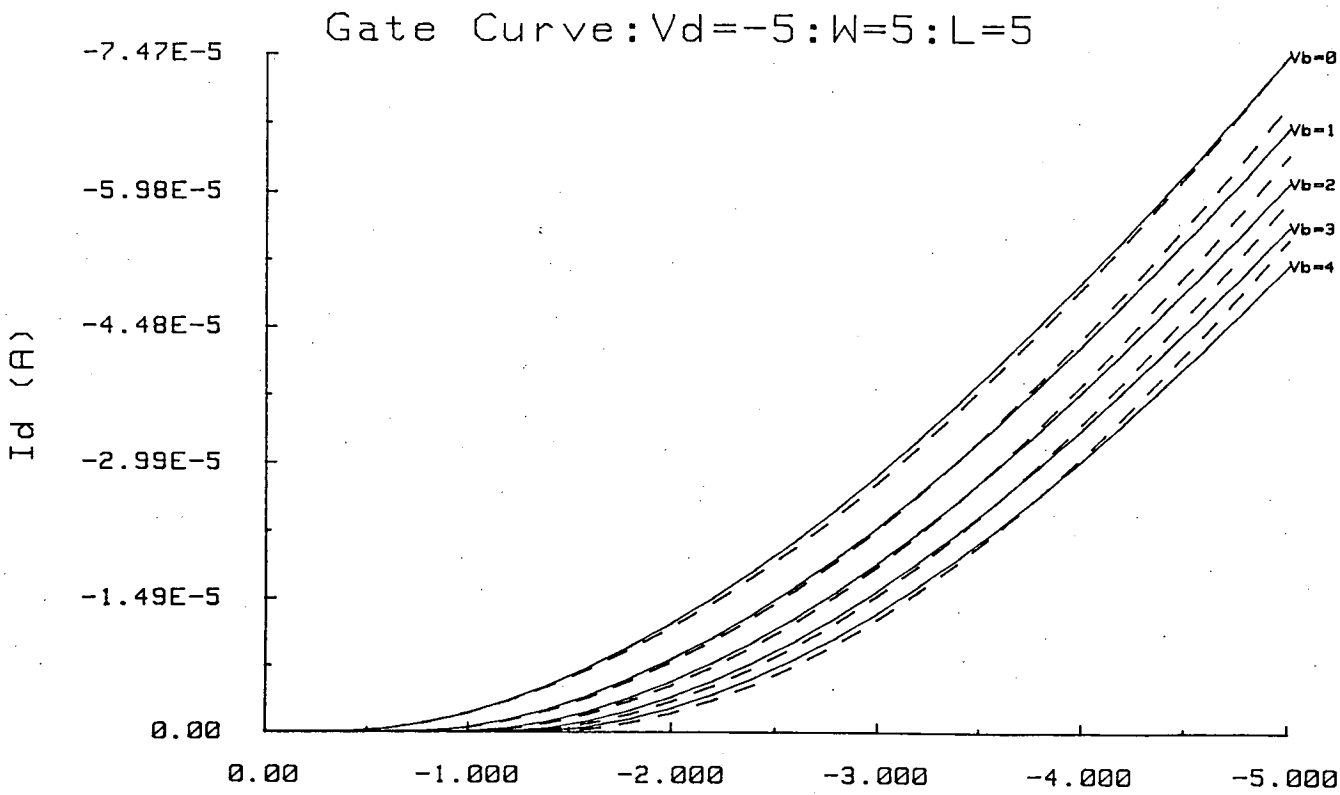


Figure 4.3.3.8

V_g (V)

PMOS

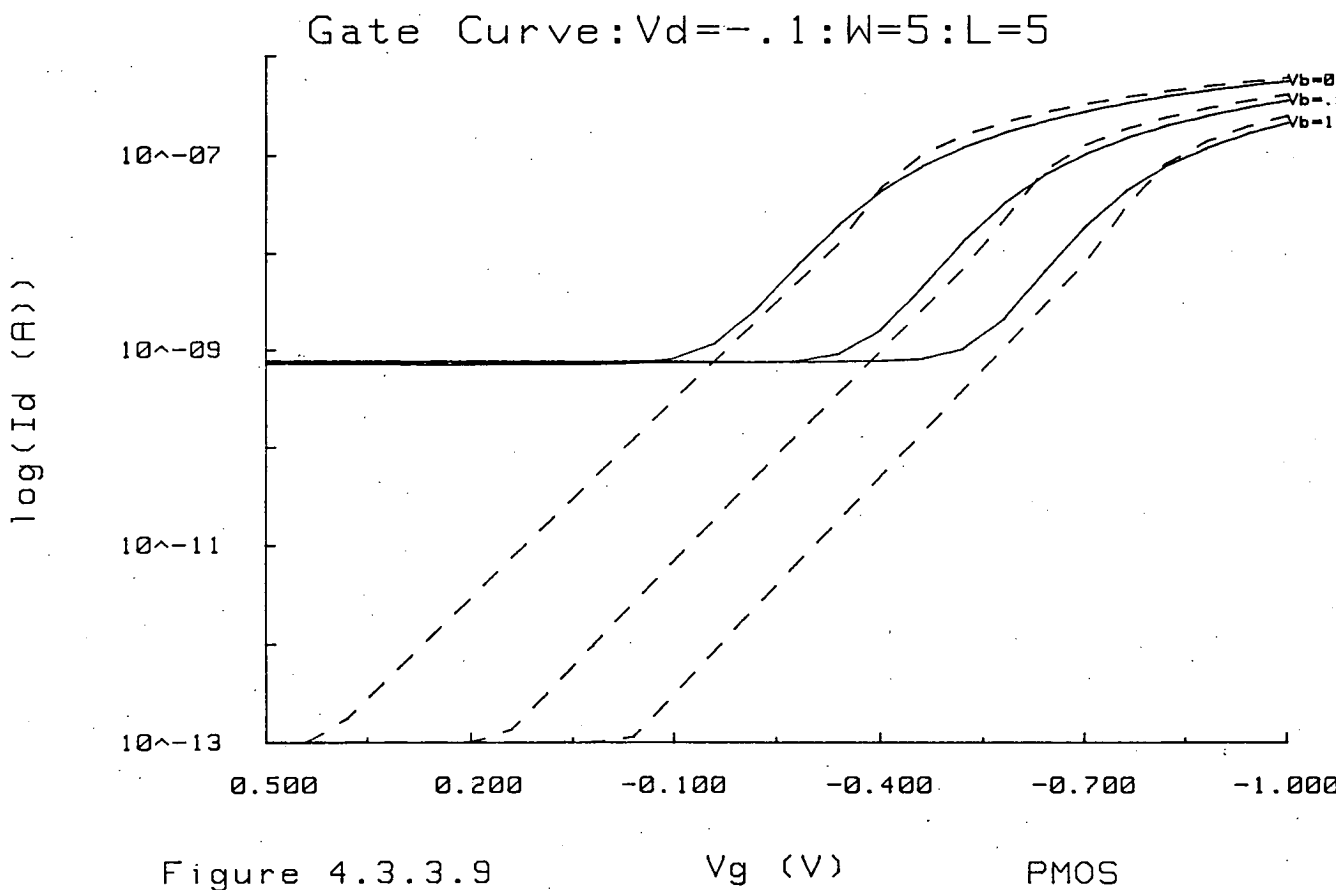


Table 4.3.3.2 Error figures for PMOS Enhancement Device(Figure 4.3.3.5)

V_g V	Average r.m.s. Error (X 10^{-6})	Max Absolute Error (X 10^{-6})	Average Percentage Error %
-1	0.13	0.25	6.6
-2	0.28	0.48	2.4
-3	0.44	0.83	1.6
-4	0.51	0.94	1.1
-5	0.79	1.40	1.3

Table 4.3.3.3 Error figures for PMOS Enhancement Device(Figure 4.3.3.6)

V_b V	Average r.m.s. Error (X 10^{-6})	Max Absolute Error (X 10^{-6})	Average Percentage Error %
0	0.03	0.05	3.1
1	0.06	0.12	4.7
2	0.10	0.19	9.1
3	0.12	0.23	10.4
4	0.13	0.26	11.6

shown in figure 4.3.3.9. As in the NMOS case, the slope is modelled well but the transition at threshold is inaccurate and the leakage current is not modelled.

4.4 Summary

A full set of parameter extraction algorithms has been described, in which the parameters are extracted one at a time from particular transistor characteristics. The techniques yield SPICE level 1 and level 3 parameters and are suitable for either NMOS or PMOS and for either enhancement or depletion transistors. In particular, in the evaluation of V_{to} , the method for finding the turn-on point by examining the second derivative of $I_d:V_g$ can be used for both enhancement devices and depletion devices. The more common method of looking for the steepest slope of $I_d:V_g$ is not suitable for depletion devices where the carrier mobility and hence the slope of $I_d:V_g$ increases above threshold. Since each parameter evaluated by these methods has a precise definition, parameters extracted from different devices, different wafers or even different processes can be compared.

The PARAMEX program, which consists of approximately 5000 lines of advanced HP BASIC, implements the extraction algorithms described in section 4.2. It is also capable of measuring the device characteristics required for extraction and of measuring and simulating any transistor characteristic. The errors between measurement and simulation can be computed in order to verify that the resulting parameters are reasonably accurate. Accurate characterisation has been achieved for NMOS and PMOS enhancement transistors, depletion NMOS transistors and NMOS enhancement transistors with drawn channel lengths of 1.5 μm . Several fabrication plants including Motorola, East Kilbride and Inmos, Newport have requested and received copies of PARAMEX. No commercially available software uses physical parameter extraction.

PARAMEX uses over a 1000 d.c. measurement points and therefore is too time-consuming to be incorporated in end of process parametric test. An M.Sc. project looked at minimising the number of measurements required to obtain a complete parameter set using the PARAMEX algorithms. This resulted in a program called FASTPARS, which uses only 27 measurements in order to produce values similar to those from PARAMEX. Full characterisation to determine parameter values and

normal variations with time should still be carried out using PARAMEX but then routine checking to ensure that production processing remains within specification can be done using FASTPARS. A project is underway to further reduce the measurement time by performing some of the measurements in parallel. This involves placing some simple circuitry behind each probe pin, to either force voltage or measure current, and also locating identical devices side by side on the testchip. Another project, aiming to reduce the number of probe pads on test chips, has investigated using PARAMEX to measure parameters from a device in series with another large device which functions as an on-chip switch. Hence a substantial quantity of research work has arisen from the original development of PARAMEX.

Some of the deficiencies of the SPICE level 3 model have become apparent through this study of parameter extraction and three are discussed below. Perhaps most significant is the change in carrier mobility μ with substrate bias V_b . In many circuits, devices operate with non-zero substrate bias and since the transconductance g_m increases as V_b increases, the drive current of a digital gate or the gain of a CMOS amplifier may be underestimated. To solve this a modification of θ with V_b would be required. The exact relationship could be devised after an experimental investigation of this variation. Secondly, the transition out of subthreshold is usually poorly modelled. The simulation in the linear region is good and the slope in subthreshold is correct but due to the transitional region between them, the simulated subthreshold current is often an order of magnitude away from the measured current. Perhaps in Wright's model (see Section 2.6), where this transition has been altered, a more accurate representation of this region might result. Finally the leakage current which depletion devices often exhibit when the substrate voltage is around zero, is not modelled. To account for this a leakage current which is substrate voltage (and possibly gate voltage) dependent may be included. If the current simulated using the normal level 3 model is below the leakage current, then the leakage current flows in the device.

*Chapter 5 : The Influence of Device Size, Manufacturing Variations
and Process Variations on Parameters.*

5.1 Parameters and Device Size

5.1.1 Introduction

The creators of the SPICE program and, in particular the MOS transistor models, intended a single set of parameters to be used for all devices produced using a particular fabrication process. These values therefore should be irrespective of the lengths and widths of the devices. Any geometrical aspect of device operation should be included in the analytical equations. However, the dependencies incorporated within the model are a poor representation of reality and so it is useful to investigate the variation of parameters for different sizes of devices. As was mentioned previously (section 2.6), Wright³⁹ overcomes this by using a preprocessor lookup table to provide the parametric values to be input to the model and the CASMOS developers, Oakley and Hocking²⁴ devised their own empirical relationships for length and width influences on device operation.

From a wafer manufactured using a version of the EMF VLSI NMOS process, parameters were measured on transistors which varied in size from $1\mu\text{m} \times 1\mu\text{m}$ to $30\mu\text{m} \times 30\mu\text{m}$. These results are tabulated in Table 5.1.1.1. The variations will be discussed, and the effects in the SPICE model and the corresponding relationship in CASMOS will be examined. Finally some guidelines on how to simulate devices of different sizes using SPICE are provided.

5.1.2 Threshold Voltage

Threshold voltage, V_{to} ranges from 0.72V for the $1\mu\text{m} \times 1\mu\text{m}$ device down to 0.58V for the $30\mu\text{m} \times 30\mu\text{m}$ device. The variation is plotted in figure 5.1.2.1. Threshold voltage increases as device width decreases below $4\mu\text{m}$ and decreases as device length decreases below $4\mu\text{m}$. The former effect is due to the fact that the fringing electric field around the gate takes up a larger proportion of the bulk charge required to turn the transistor on and the latter, the weaker effect, is due to the depletion regions around the drain and source junctions becoming a significant

Table 5.1.1.1 Variation of Parameters with Transistor Geometry					
Geometry (μm)	1X1	1.2X1.2	1.5X1.5	2X2	2.5X2.5
V_{io} (V)	0.72	0.65	0.65	0.62	0.62
γ ($V^{\frac{1}{2}}$)	0.90	0.80	0.75	0.73	0.68
μ_o ($m^2.Vs^{-1}$)	0.109	0.104	0.082	0.063	0.068
θ (V^{-1})	0.053	0.066	0.058	0.058	0.060
v_{max} ($m.s^{-1}$)	2.96×10^5	2.73×10^5	2.38×10^5	2.71×10^5	2.60×10^5
L_d (μm)	0.084	0.084	0.084	0.084	0.084
Δ_w (μm)	0.34	0.34	0.34	0.34	0.34
η	0.033	0.027	0.025	0.035	0.057
δ	0.117	0.197	0.273	0.321	0.452
κ	0.086	0.067	0.059	0.030	0.045
N_{fs} (m^{-2})	4.3×10^{15}	3.6×10^{15}	3.7×10^{15}	5.1×10^{15}	2.6×10^{15}

Table 5.1.1.1 Variation of Parameters with Transistor Geometry(cont)					
Geometry (μm)	3X3	4X4	5X5	7X7	30X30
V_{io} (V)	0.61	0.60	0.59	0.58	0.58
γ ($V^{\frac{1}{2}}$)	0.65	0.63	0.60	0.59	0.57
μ_o ($m^2.Vs^{-1}$)	0.063	0.063	0.062	0.062	0.061
θ (V^{-1})	0.060	0.063	0.062	0.062	0.057
v_{max} ($m.s^{-1}$)	3.06×10^5	3.39×10^5	4.44×10^5	6.47×10^5	1.95×10^5
L_d (μm)	0.084	0.084	0.084	0.084	0.084
Δ_w (μm)	0.34	0.34	0.34	0.34	0.34
η	0.050	0.127	0.209	0.445	-
δ	0.549	0.756	0.935	1.307	-
κ	0.047	0.099	0.118	0.271	0.865
N_{fs} (m^{-2})	3.1×10^{15}	3.0×10^{15}	3.1×10^{15}	3.2×10^{15}	4.5×10^{15}

proportion of the channel.

In the SPICE level 3 model, the influence of the depletion regions around the source and drain junctions is included by a complex expression which is partially empirical and partially analytical (see model equations in APPENDIX A). The threshold voltage is

$$V_{th} = V_{fb} + 2\phi_b - \sigma V_d + \gamma F_s (2\phi_b - V_b)^{\frac{1}{2}} + F_n (2\phi_b - V_b) \quad 5.1.2.1$$

where

$$F_s = 1 - \frac{x_j}{L} \left[\left[\frac{W_c}{x_j} + \frac{L_d}{x_j} \right] \left[1 - \left[\frac{W_{ps}}{x_j + W_{ps}} \right]^2 \right]^{\frac{1}{2}} - \frac{L_d}{x_j} \right] \quad 5.1.2.2$$

The F_n factor, which is inversely proportional to the width, models the effects of width on threshold through the parameter δ .

$$F_n = \frac{2\pi\epsilon_{si}}{4C_{ox}} \frac{\delta}{W} \quad 5.1.2.3$$

This produces large errors at non-zero substrate bias. Figure 5.1.2.2 shows the variation in threshold voltage which the model expects based on the parameters for the $1\mu\text{m} \times 1\mu\text{m}$ device. It suggests that the length effect dominates which is not in fact the case as demonstrated in figure 5.1.2.1.

CASMOS also recognises that the geometrical factors which influence threshold voltage are due to the variation in the bulk charge which is controlled by the gate. Consequently the effects are included as empirical factors, AK and BK, which modify the substrate bias coefficient which is denoted in CASMOS by K.

$$V_t = V_{to} + K \left[(2\phi_b - V_b)^{\frac{1}{2}} - V_b^{\frac{1}{2}} \right] - \alpha V_d \quad 5.1.2.4$$

$$K = K_{INF} \left[1 - \frac{AK}{LR} + \frac{BK}{WR} \right] \quad 5.1.2.5$$

In fact, the modification of the substrate bias coefficient is found to be inadequate for very short channels and so an extra offset, F is introduced

$$V_t = V_t + F \quad \text{when } L < L_{CRIT} \quad 5.1.2.6$$

Figure 5.1.2.1 - Measured Threshold Variation

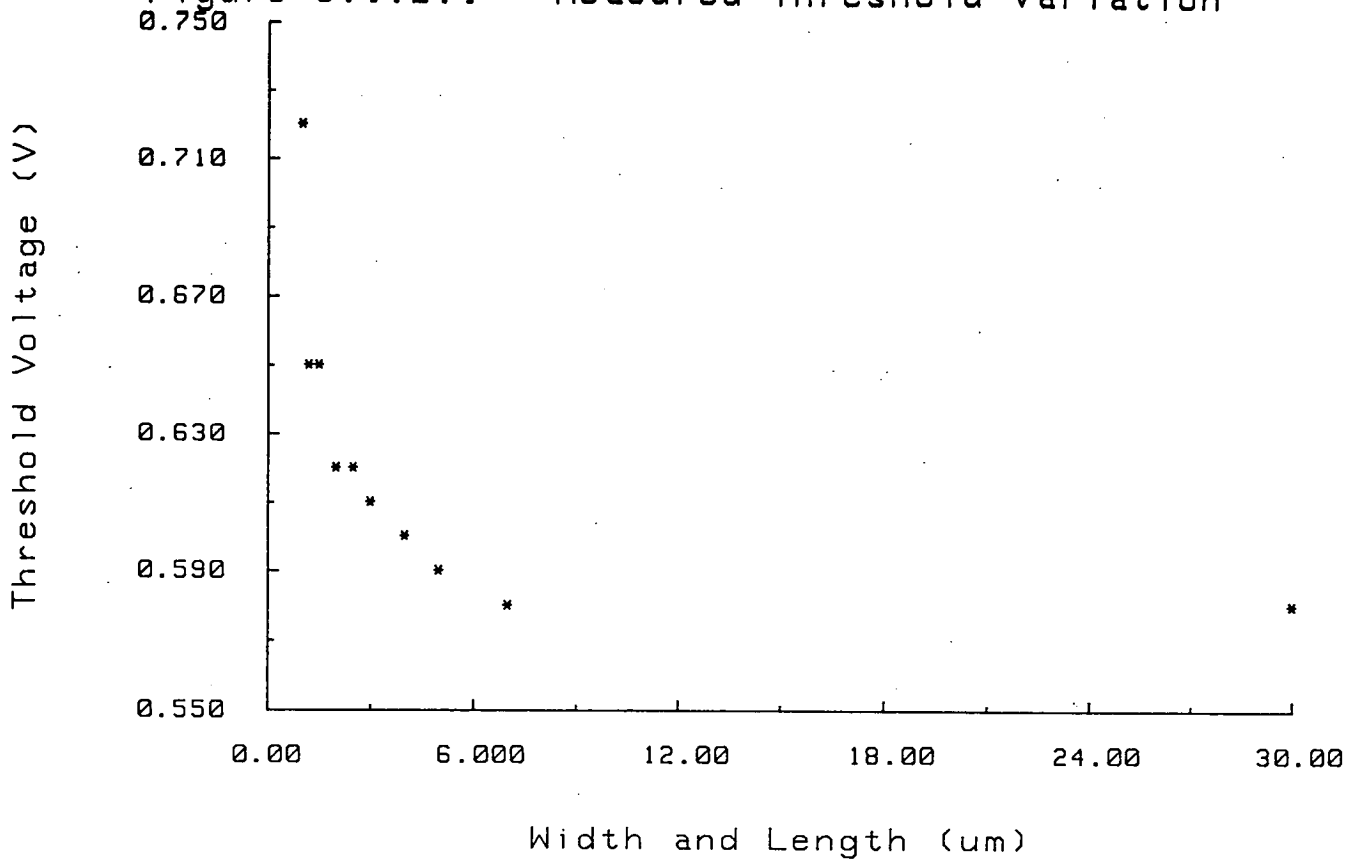
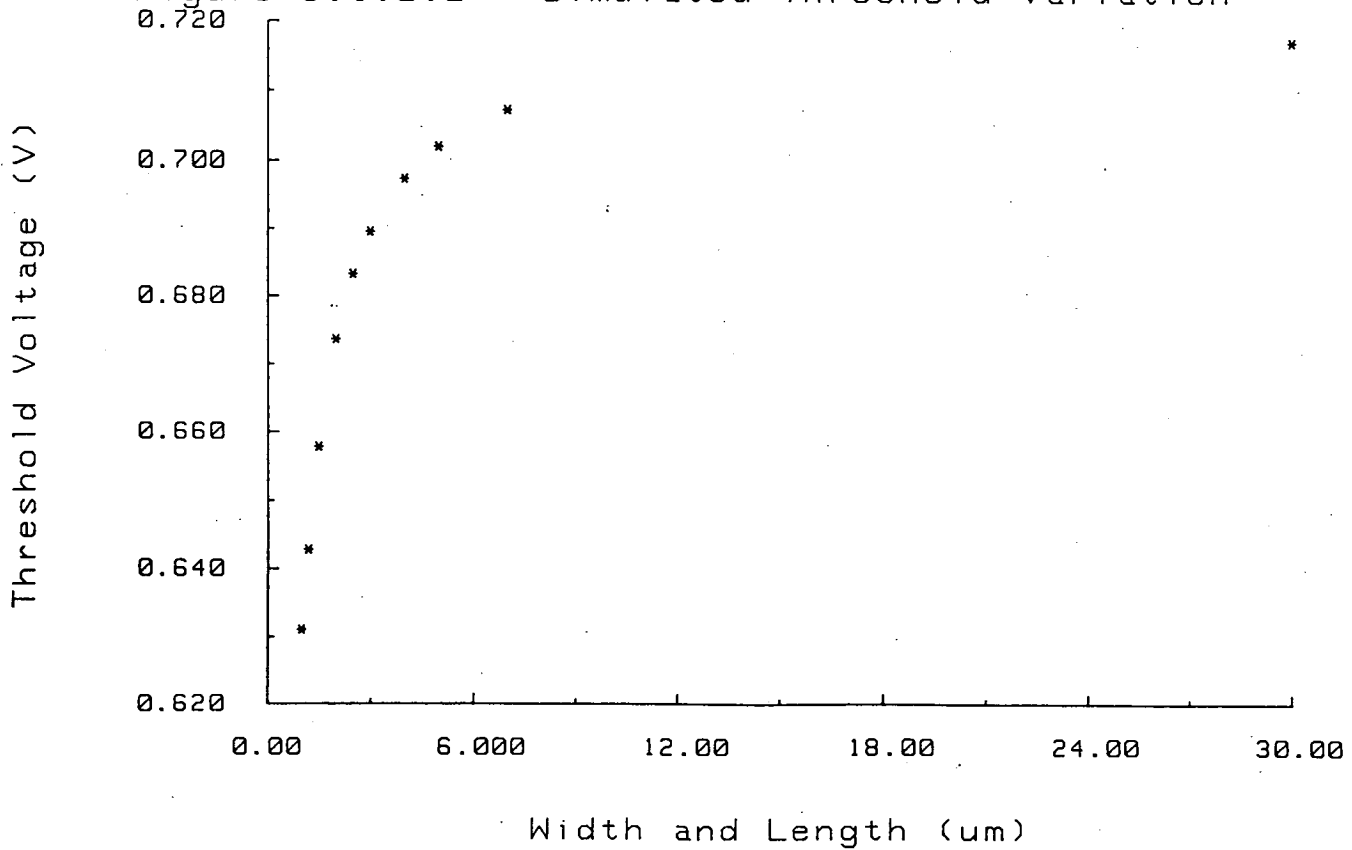


Figure 5.1.2.2 - Simulated Threshold Variation



5.1.3 Substrate Bias Coefficient, γ

In 5.1.2, the variations in threshold voltage were shown to be modelled by a modification of the substrate bias coefficient. It was stated that the stronger influence was the increase in threshold as width decreases and hence threshold increases when both dimensions are shrunk. In fact, the values of γ which have been extracted for SPICE and are displayed in Table 5.1.1.1, have had their device lengths taken into account during extraction (the short channel factor F_s is used in the extraction) and the variation in values results from the changing width. As was predicted above, there is an increase in γ as the width decreases. For the $30\mu\text{m} \times 30\mu\text{m}$ device, γ is 0.57 and there is an increase until 0.90 is reached for the $1\mu\text{m} \times 1\mu\text{m}$ device.

Assuming the values of γ are only width dependent, then the coefficient BK can be calculated for the CASMOS model. The value of K for the $30\mu\text{m} \times 30\mu\text{m}$ device is taken to be K_{INF} and the effective width and γ of the $1\mu\text{m} \times 1\mu\text{m}$ device are used to find BK

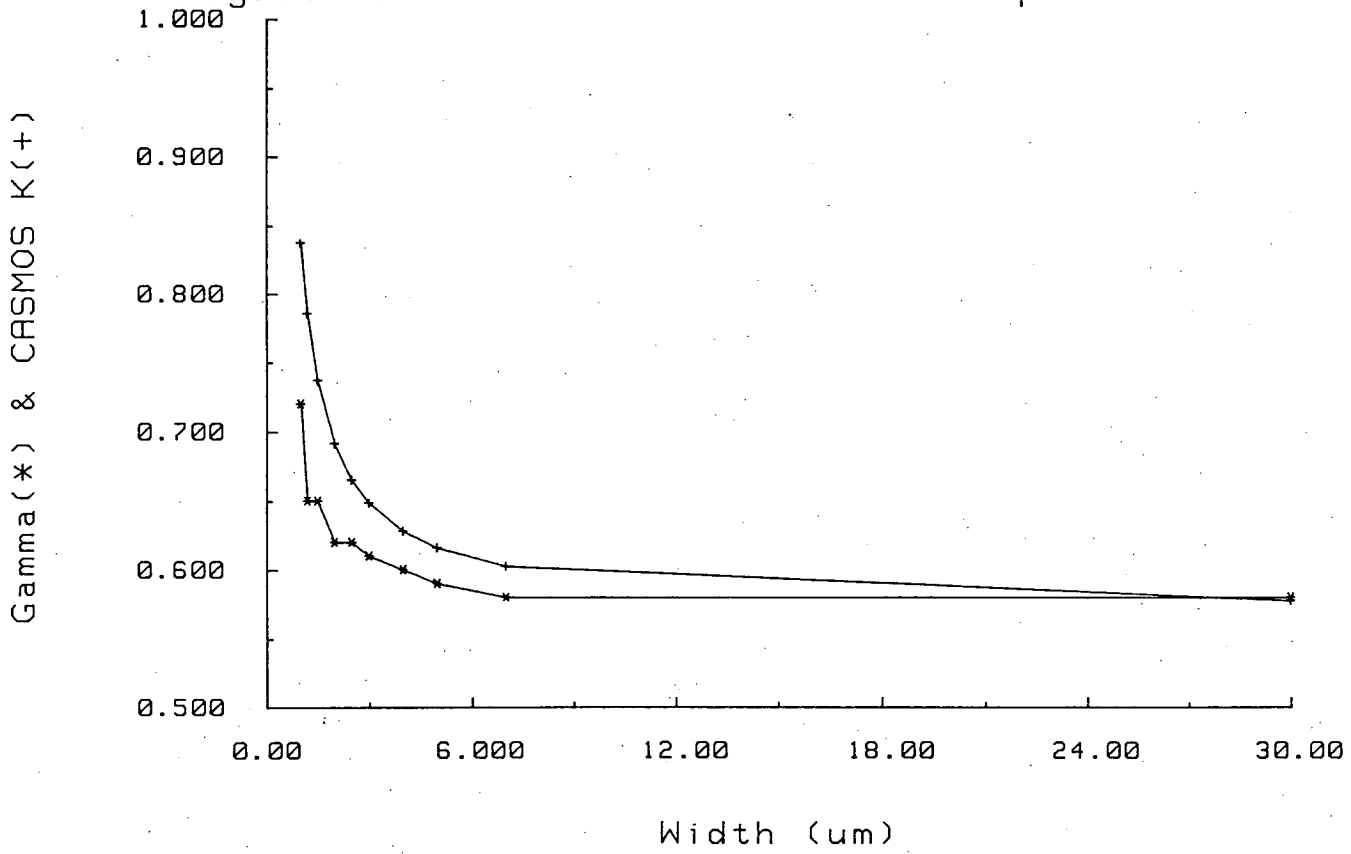
$$\begin{aligned} K &= K_{INF} \left[1 + \frac{BK}{WR} \right] & 5.1.3.1 \\ 0.90 &= 0.57 \left[1 + \frac{BK}{0.68 \times 10^{-6}} \right] \\ \Rightarrow BK &= 0.39 \times 10^{-6} \text{ m} \end{aligned}$$

Using this BK, K is plotted against width and compared with the values measured for γ (see figure 5.1.3.1). Both coefficients vary similarly except that K increases more rapidly at shorter channel lengths.

5.1.4 Mobility Variation with Gate Voltage, μ_o and θ

Carrier mobility is strongly influenced by channel length. Carrier mobility in the depletion regions around the source and drain is very high because of the high parallel electrical fields. In shorter channels, these regions become a significant portion of the channel and hence average carrier mobility increases. This is confirmed by the figures for μ_o in table 5.1.1. For a $30\mu\text{m} \times 30\mu\text{m}$ device it is $0.057 \text{ m}^2 \text{ Vs}^{-1}$; for a $3\mu\text{m} \times 3\mu\text{m}$ device it is $0.063 \text{ m}^2 \text{ Vs}^{-1}$ and for the $1\mu\text{m} \times 1\mu\text{m}$ device μ_o is

Figure 5.1.3.1 - Gamma/CASMOS K Comparison



0.109 $m^2 Vs^{-1}$. The extraction of μ_o and θ for the $1\mu m \times 1\mu m$, $1.5\mu m \times 1.5\mu m$, $2\mu m \times 2\mu m$ and $30\mu m \times 30\mu m$ devices are illustrated in figures 5.1.4.1, 5.1.4.2, 5.1.4.3 and 5.1.4.4 respectively.

θ is found to be approximately constant for the full range of device sizes. It varies between $0.053 V^{-1}$ and $0.066 V^{-1}$ with values varying by $\pm 12\%$.

The SPICE model does not provide for any increase in μ_o with decreasing channel length or for any change in θ . CASMOS, on the other hand, includes an expression for modifying θ , which is formulated in a similar fashion to the one for the substrate bias coefficient.

$$THETA = THINF \left[1 + \frac{ATHETA}{LR} - \frac{BTHETA}{WR} \right] \quad 5.1.4.1$$

This equation means that in CASMOS, θ increases as length decreases and θ decreases as width decreases.

5.1.5 Carrier Mobility and Drain Voltage, v_{max}

At normal operating voltages, drain voltage only degrades carrier mobility in shorter channels where the parallel electric field becomes high. The drain voltage versus mobility relationships for the $1\mu m \times 1\mu m$, $1.5\mu m \times 1.5\mu m$, $4\mu m \times 4\mu m$ and $30\mu m \times 30\mu m$ devices are shown in figures 5.1.5.1, 5.1.5.2, 5.1.5.3 and 5.1.5.4 respectively. For the $1\mu m \times 1\mu m$ device, mobility varies from $0.089 m^2 Vs^{-1}$ to $0.043 m^2 Vs^{-1}$ whereas for the $4\mu m \times 4\mu m$ device, mobility only changes from $0.049 m^2 Vs^{-1}$ to $0.043 m^2 Vs^{-1}$. The reduction for the $1\mu m \times 1\mu m$ device is over 50% compared with only about 12% for the $4\mu m \times 4\mu m$ device.

As the device dimensions increase from $2.5\mu m$ to $7\mu m$, v_{max} gradually increases from $2.6 \times 10^5 m s^{-1}$ to $6.5 \times 10^5 m s^{-1}$. The effect is not really significant for the $30\mu m \times 30\mu m$ as can be seen in figure 5.1.5.4 and performing the extraction for completeness results in a v_{max} of $1.9 \times 10^5 m s^{-1}$.

SPICE recognises the fact that this effect is more pronounced for shorter channels and mobility is modified according to

$$\mu_{eff} = \frac{\mu_s}{1 + \frac{V_d \mu_s}{Lv_{max}}} \quad 5.1.5.1$$

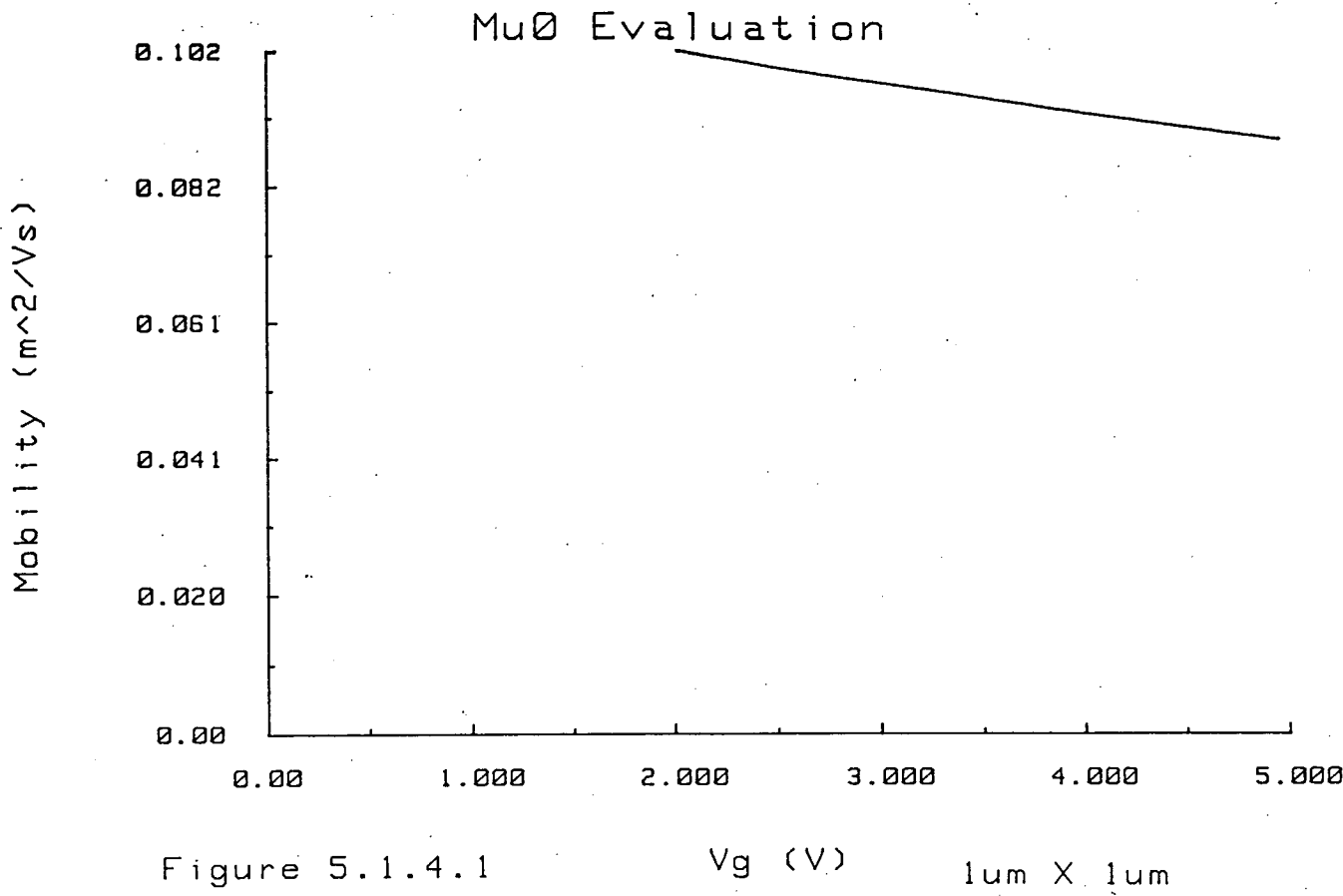


Figure 5.1.4.1

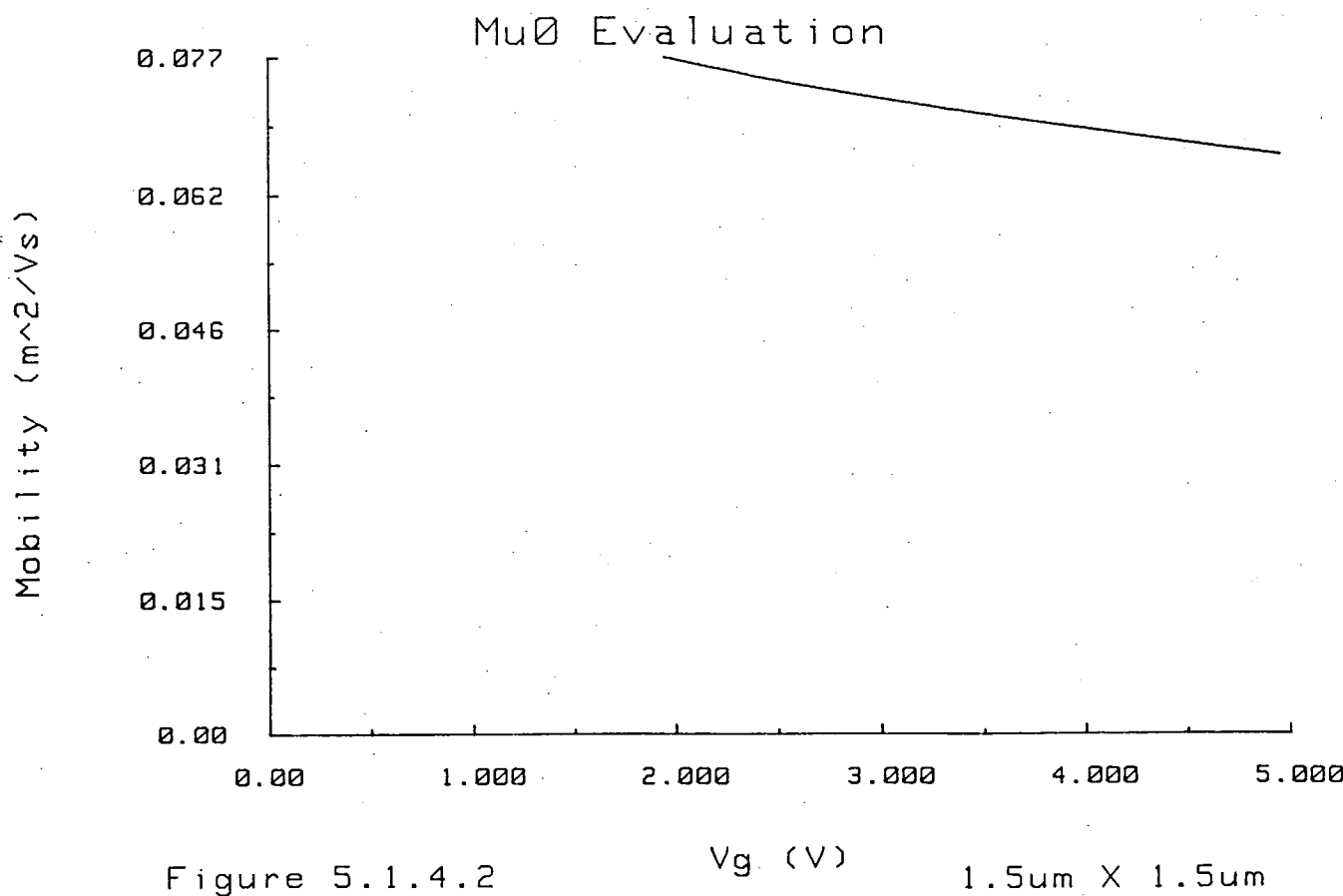


Figure 5.1.4.2

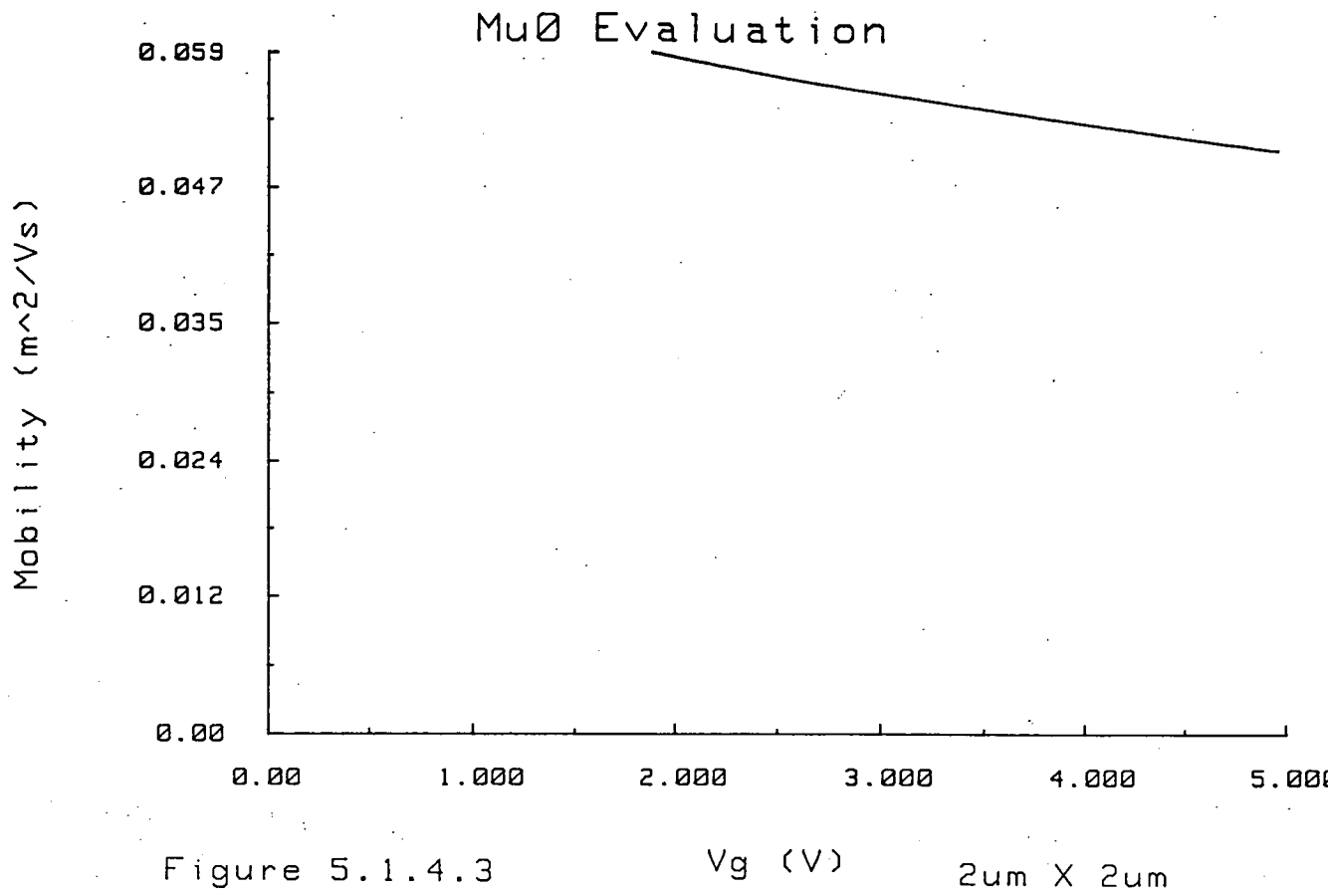


Figure 5.1.4.3

Vg (V)

2um X 2um

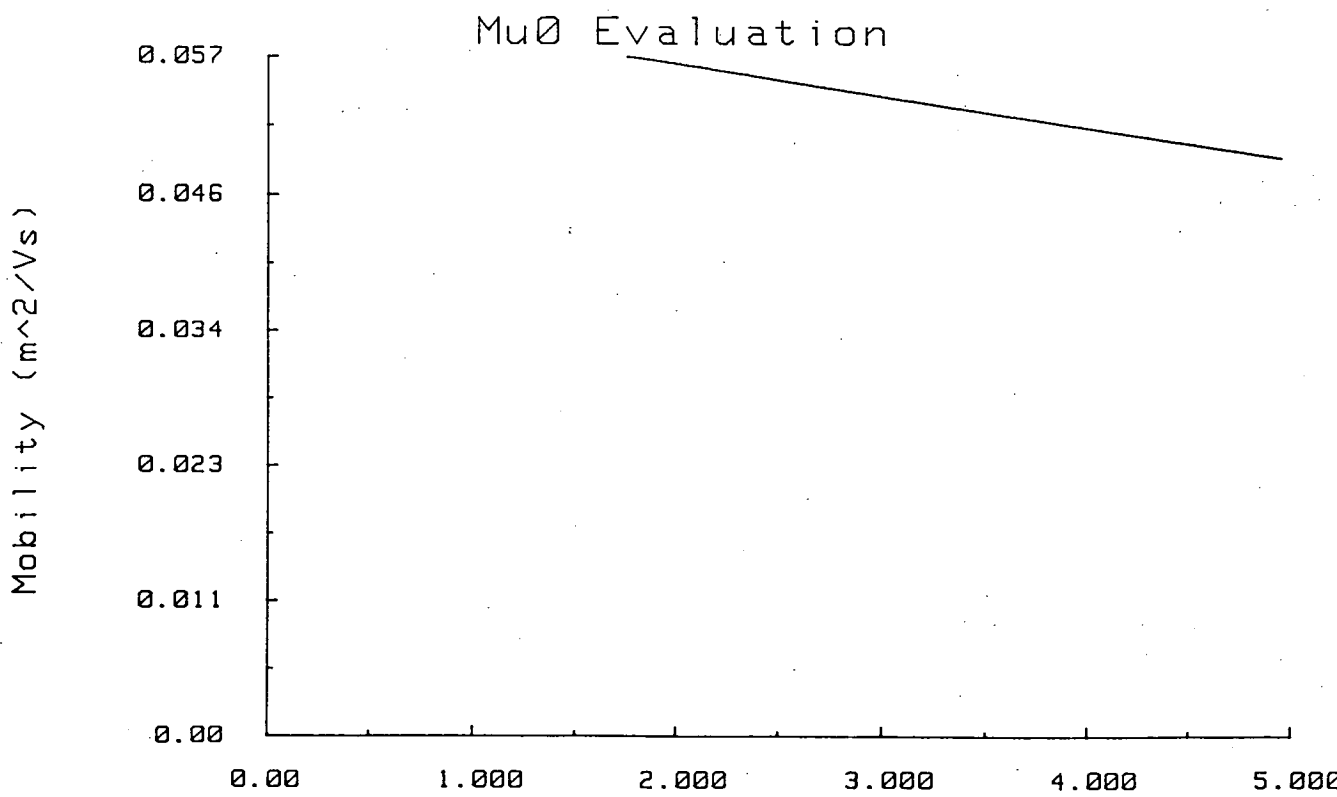


Figure 5.1.4.4

Vg (V)

30um X 30um

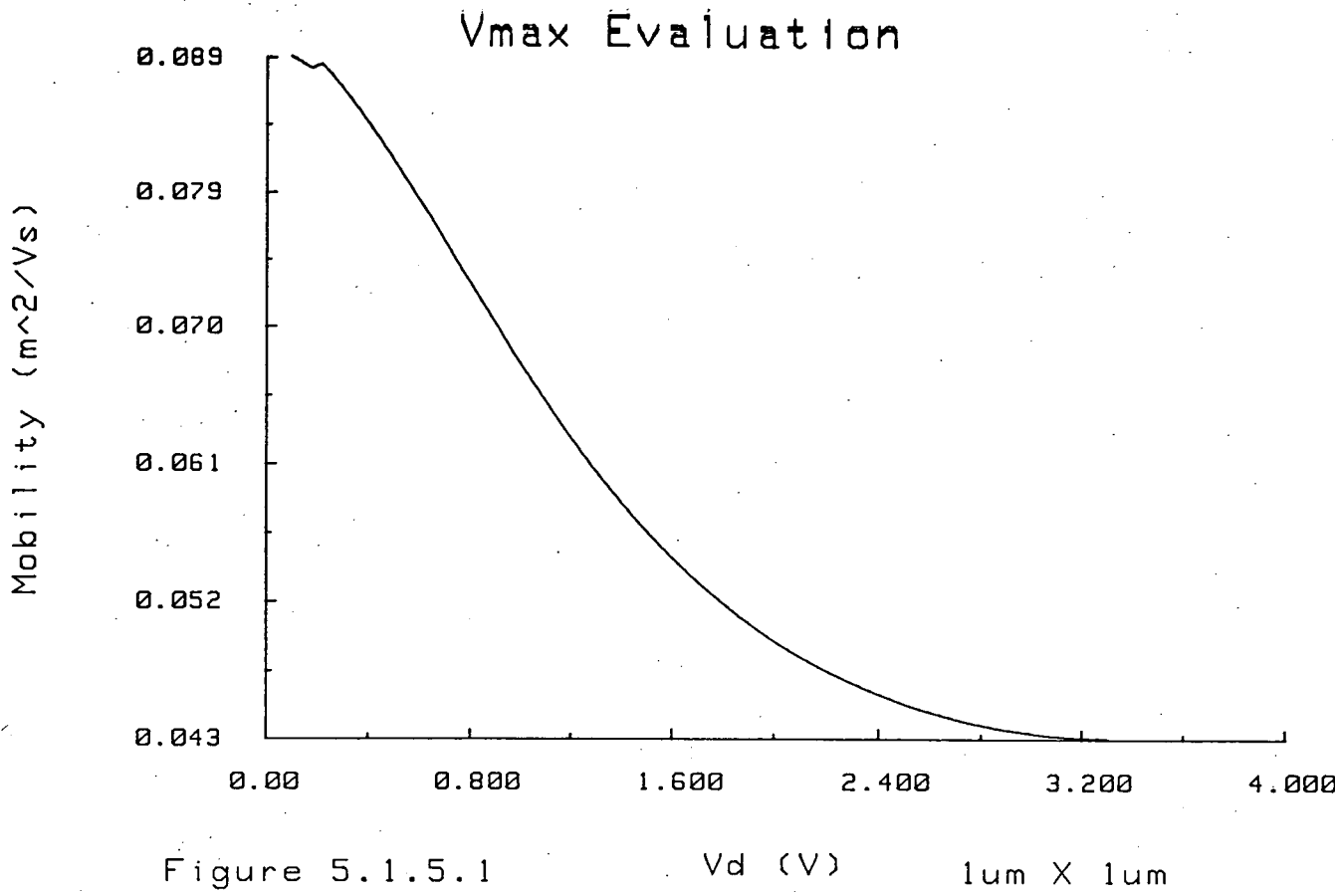


Figure 5.1.5.1

Vd (V)

1um X 1um

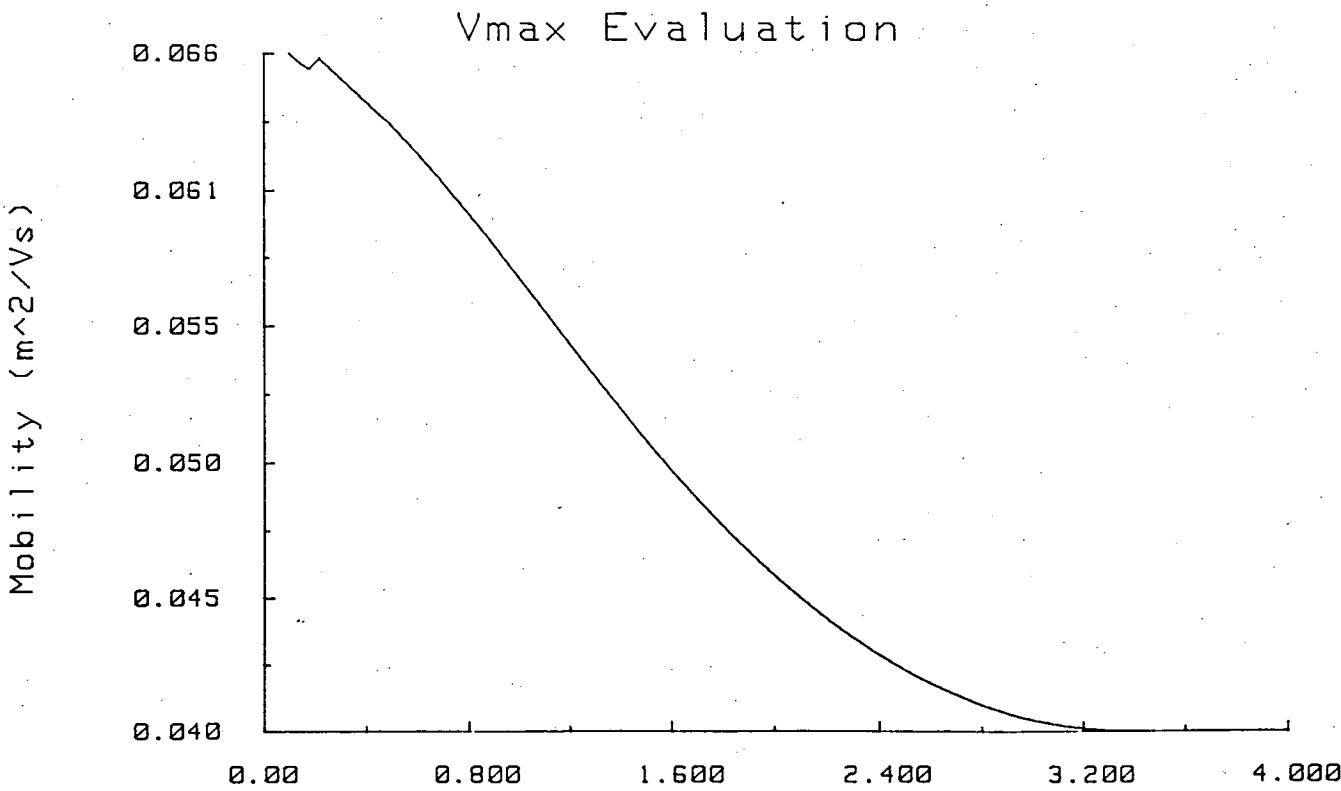


Figure 5.1.5.2

Vd (V)

1.5um X 1.5um

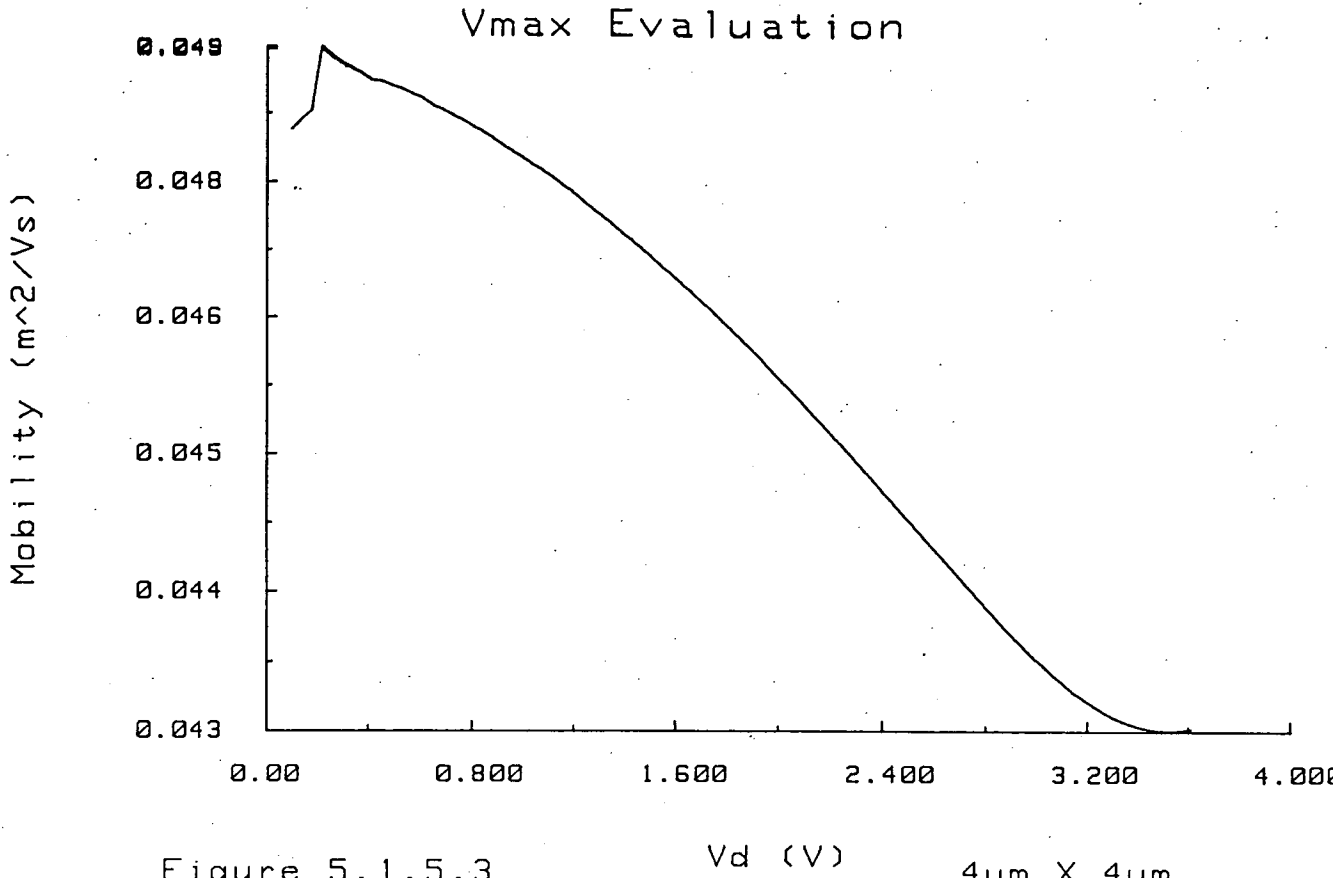


Figure 5.1.5.3 Vd (V) 4um X 4um

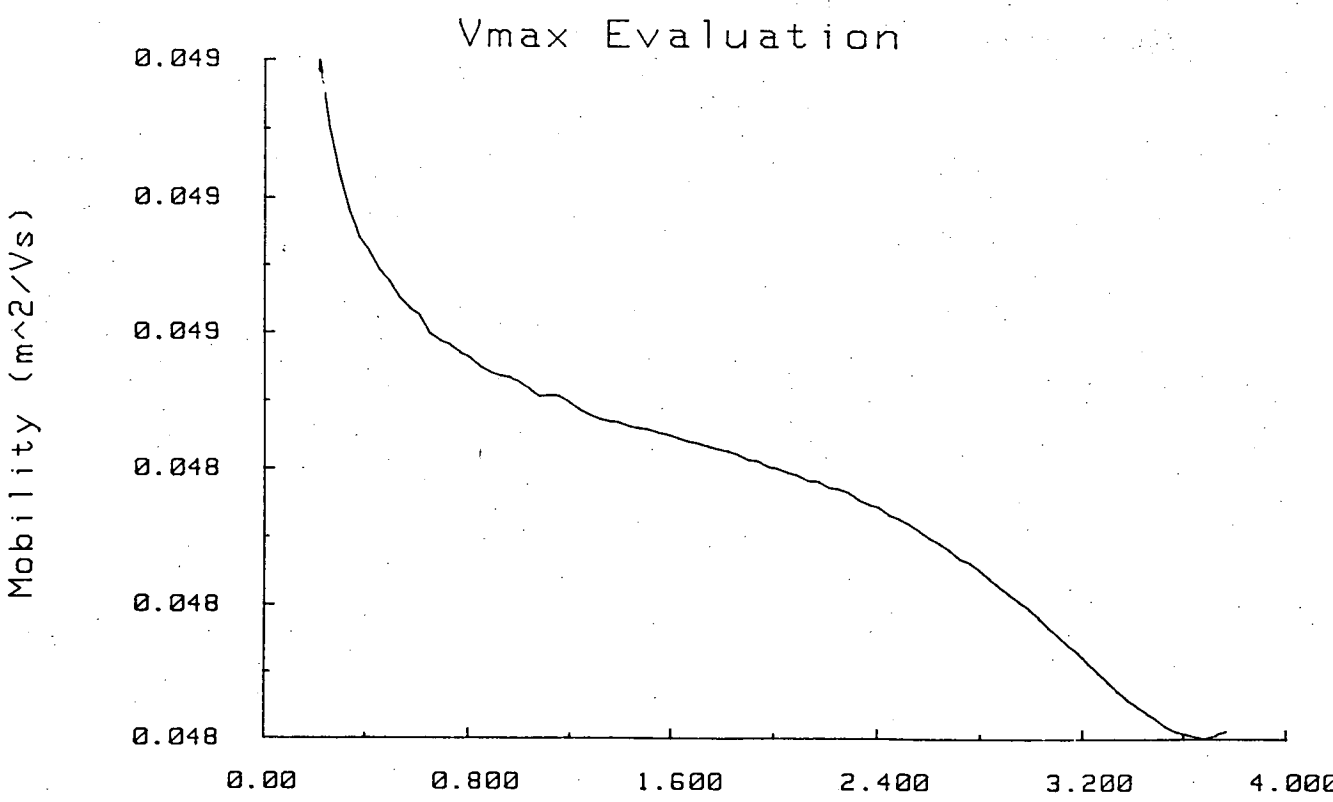


Figure 5.1.5.4 Vd (V) 30um X 30um

Figure 5.1.5.5 shows how mobility varies according to this relationship for different channel lengths. The parameters in table 5.1.1.1 were used. If a single set of parameters are intended to represent devices of all geometries then v_{max} should be measured on a short channel device where it has maximum effect.

CASMOS agrees that mobility modulation by drain voltage is dominated by velocity saturation in short channel devices. It is modelled by

$$\mu_{eff} = \frac{\mu_o}{1 + \theta(V_g - V_t) + \alpha V_d} \quad 5.1.5.2$$

where

$$\alpha = \frac{\alpha_n}{L^2} \quad 5.1.5.3$$

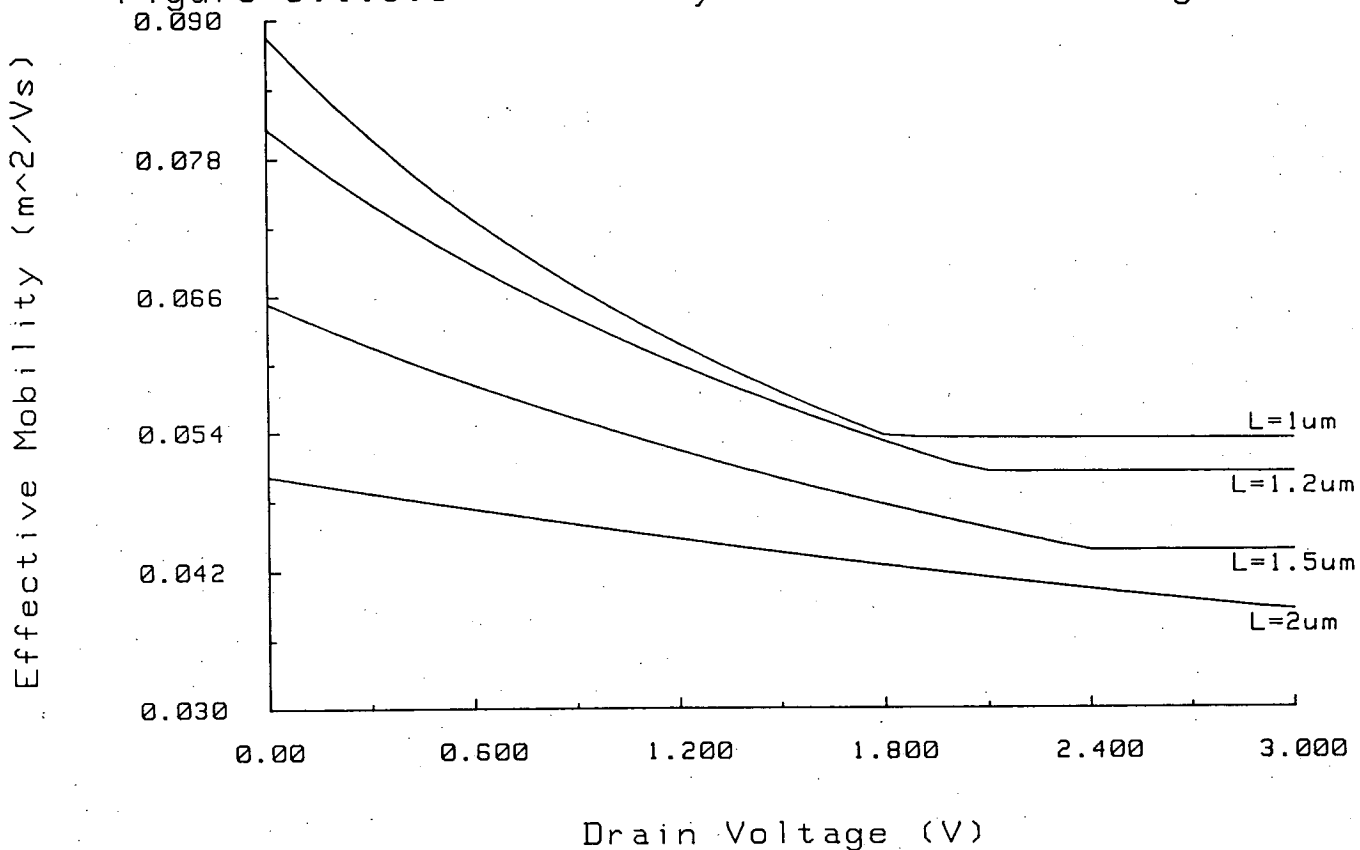
The value of α for a particular length is calculated from a normalised value α_n using the effective channel length. The effect is proportional to L^{-2} .

5.1.6 Threshold Modulation by Drain Voltage, η

Another effect which becomes important when considering short channels is the modulation of threshold voltage by the drain voltage. For short channels, the depletion region around the drain junction becomes an appreciable part of the bulk charge required to turn the device on and there is a noticeable decrease in threshold as drain voltage increases. This can be seen in the subthreshold characteristics for a $1\mu\text{m} \times 1\mu\text{m}$, a $1.2\mu\text{m} \times 1.2\mu\text{m}$, a $1.5\mu\text{m} \times 1.5\mu\text{m}$ and a $5\mu\text{m} \times 5\mu\text{m}$ which are shown in figures 5.1.6.1, 5.1.6.2, 5.1.6.3 and 5.1.6.4. These were measured with drain voltages of 0.1V, 1V, 2V, 3V, 4V and 5V with the substrate held at 0V. From figures 5.1.6.5 and 5.1.6.6 which show the extraction of η for the $1\mu\text{m} \times 1\mu\text{m}$ and the $1.5\mu\text{m} \times 1.5\mu\text{m}$ respectively, the actual variation in threshold can be seen. For the $1\mu\text{m} \times 1\mu\text{m}$, there is a 134 mV change in threshold as the drain voltage increases from 1V to 5V whereas for the $1.5\mu\text{m} \times 1.5\mu\text{m}$ device, it is only 25 mV. The CASMOS parameter η_{CAS} is equivalent to the SPICE variable σ calculated from η having taken into account the device length. η_{CAS} is calculated from a normalised value η_{CASn} having taken into account the device length. The equations for SPICE are

$$\sigma = \frac{\eta 8.15 \times 10^{-22}}{C_{ox} L^3} \quad 5.1.6.1$$

Figure 5.1.5.5 - Velocity Saturation and Length



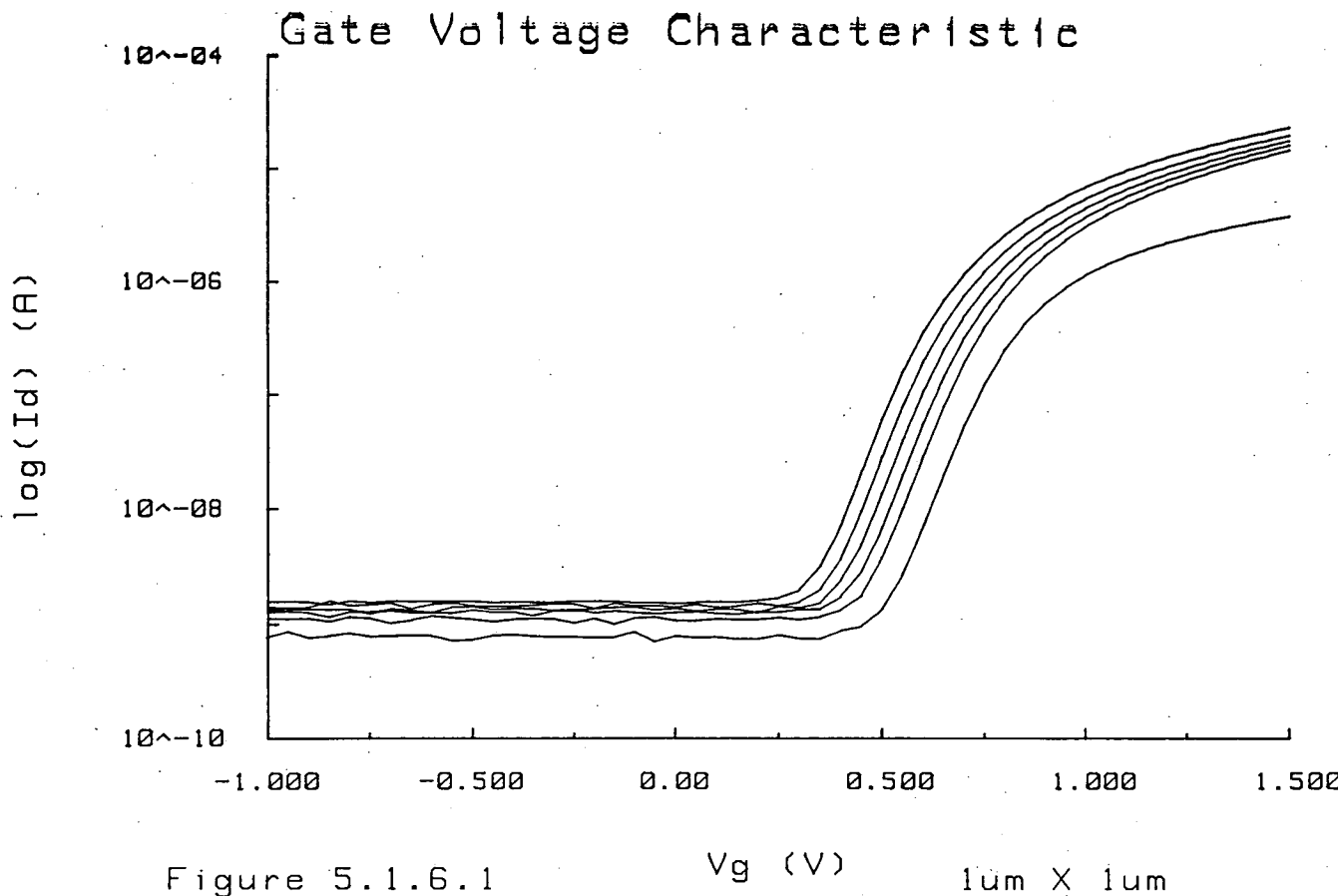


Figure 5.1.6.1 V_g (V) $1\mu\text{m} \times 1\mu\text{m}$

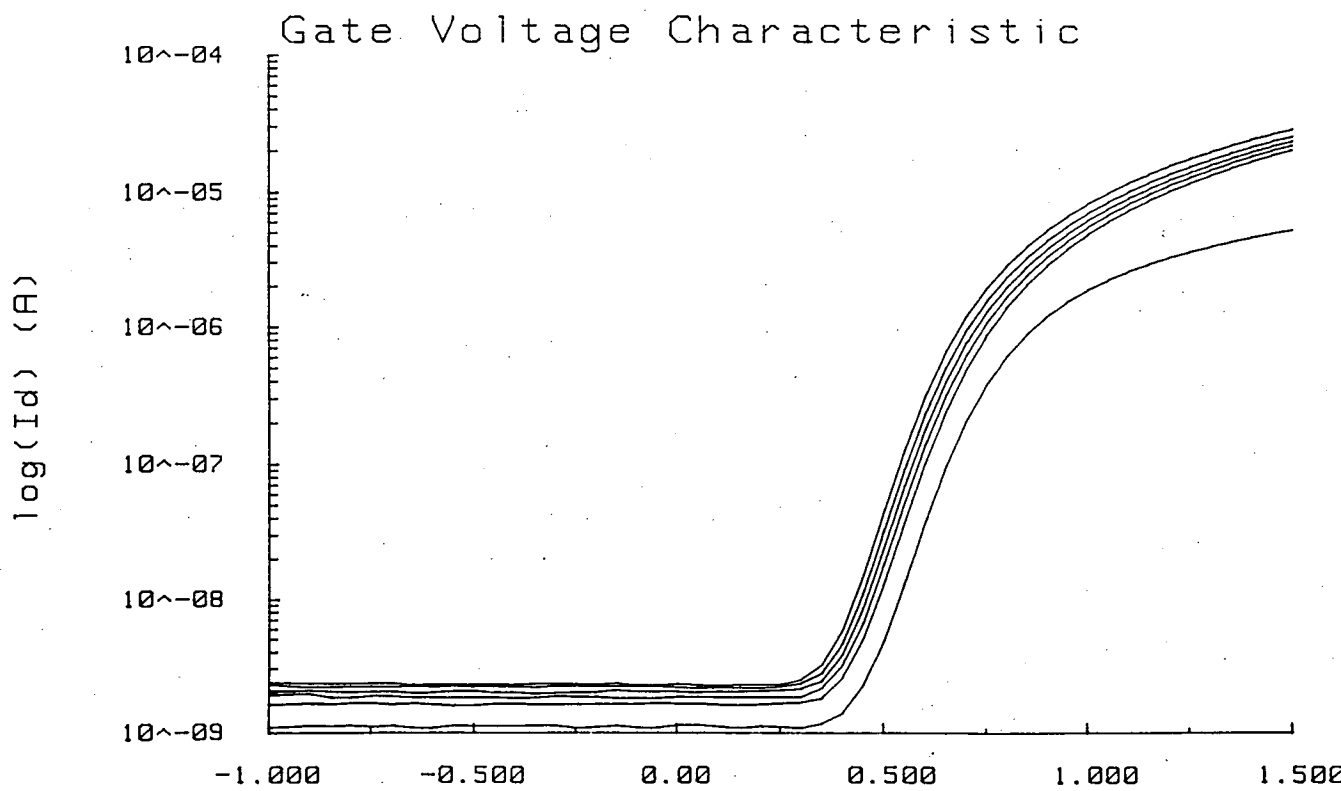


Figure 5.1.6.2 V_g (V) $1.2\mu\text{m} \times 1.2\mu\text{m}$

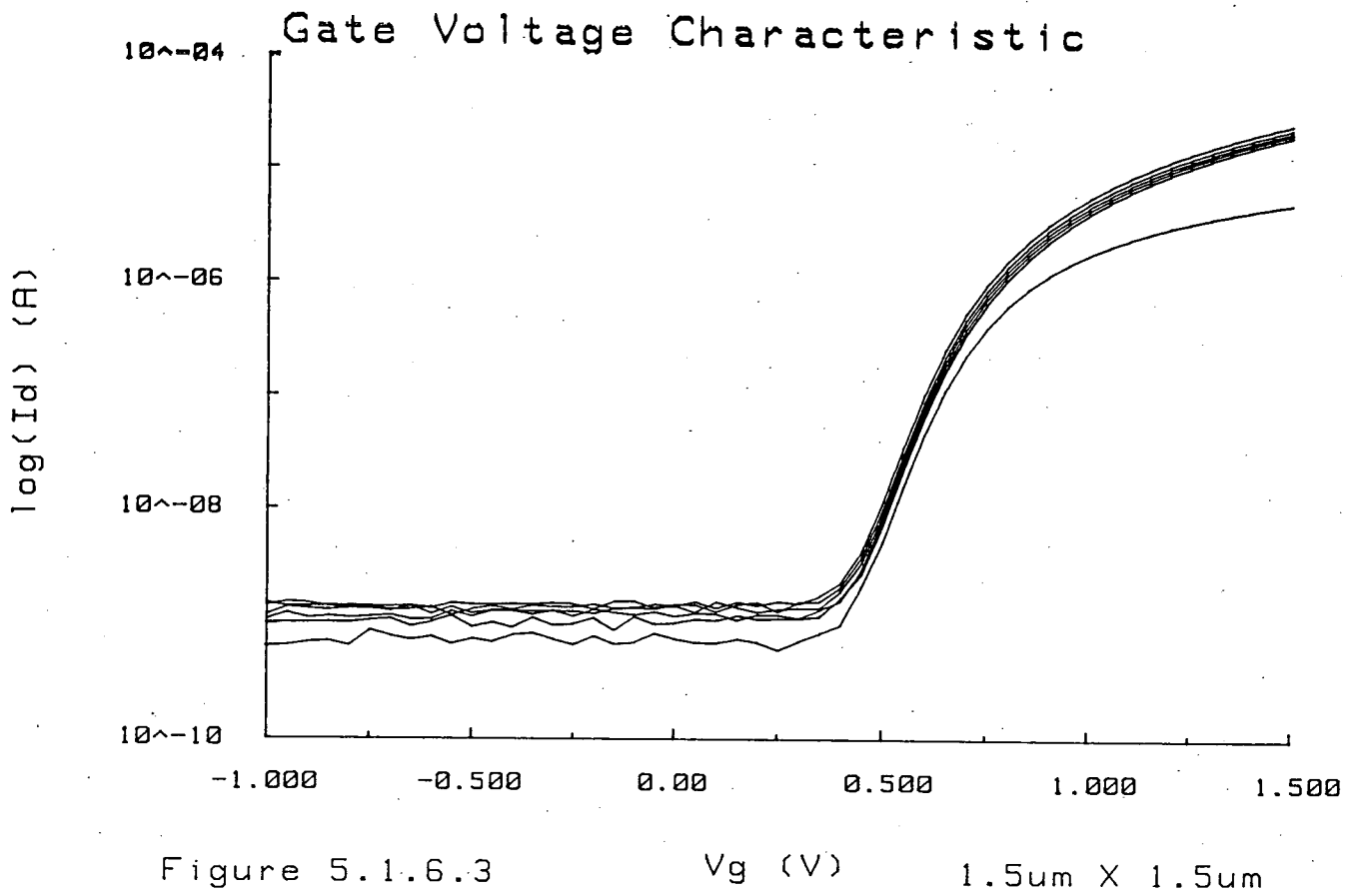


Figure 5.1.6.3

V_g (V)

$1.5\mu\text{m} \times 1.5\mu\text{m}$

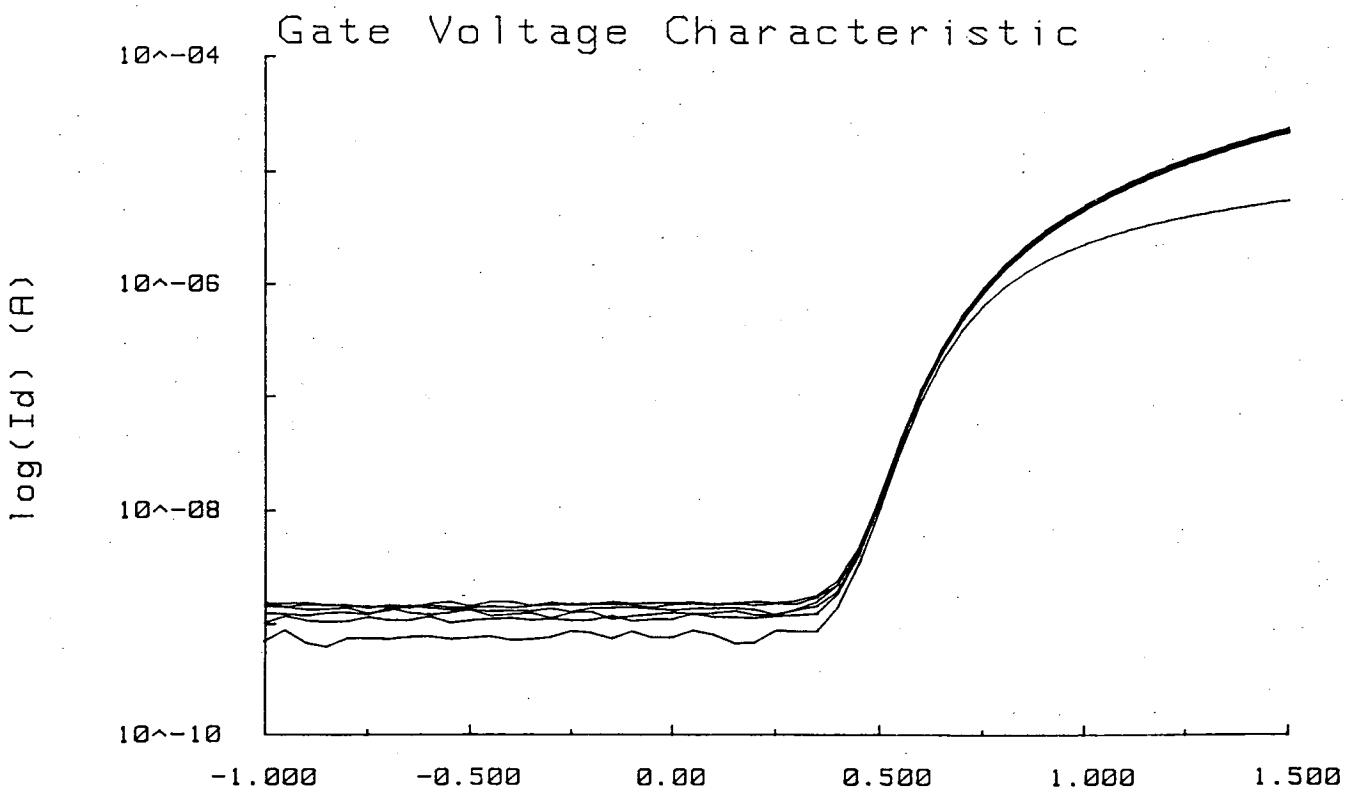


Figure 5.1.6.4

V_g (V)

$5\mu\text{m} \times 5\mu\text{m}$

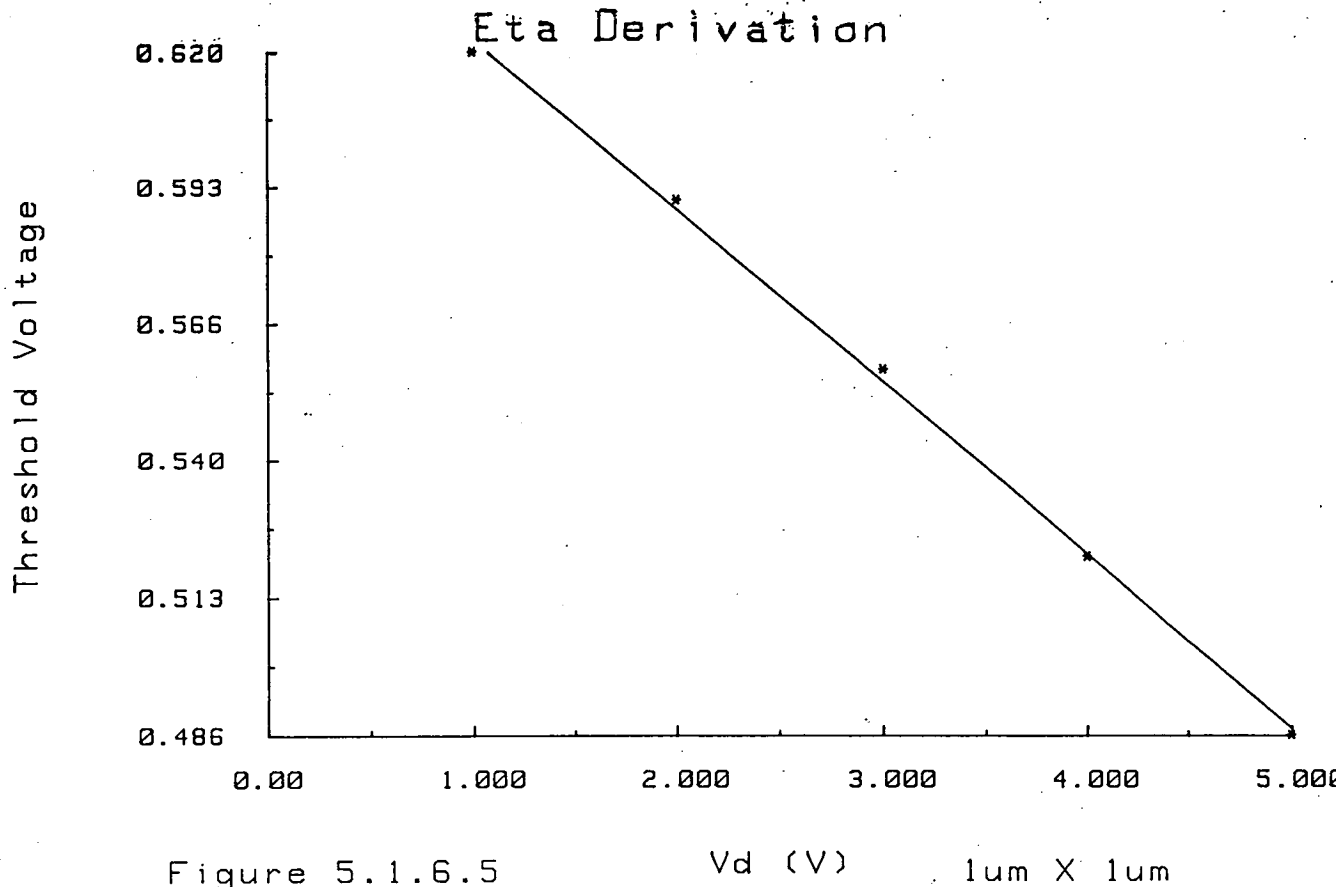


Figure 5.1.6.5

Vd (V)

1um X 1um

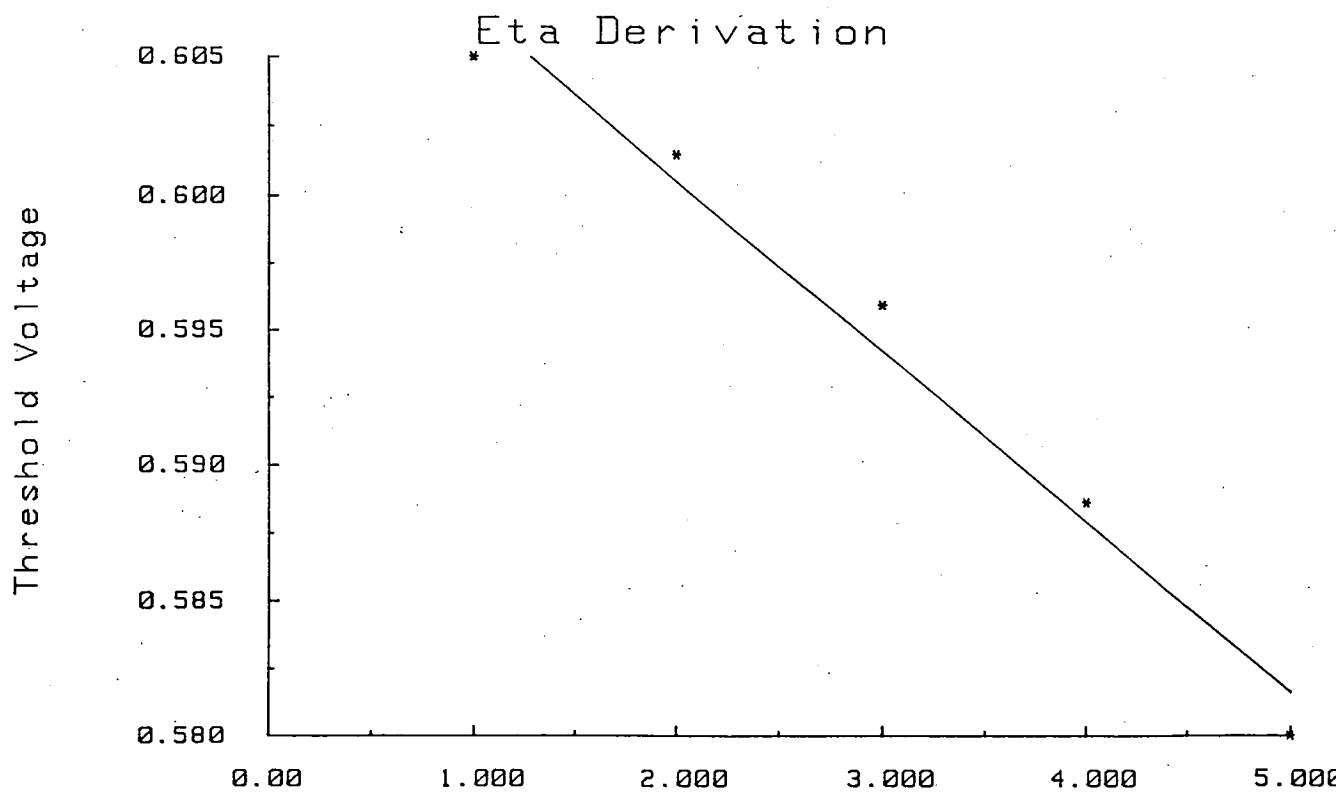


Figure 5.1.6.6

Vd (V)

1.5um X 1.5um

$$V_{th} = V_{th} + \dots - \sigma V_d \quad 5.1.6.2$$

and for CASMOS are

$$\eta_{CAS} = \frac{\eta_{CASn}}{L^{\frac{3}{2}}} \quad 5.1.6.3$$

$$V_i = V_{io} + \dots - \eta_{CAS} V_d \quad 5.1.6.4$$

From the equations it can be seen that CASMOS uses a $L^{-\frac{3}{2}}$ dependence for η_{CAS} whereas SPICE uses L^{-3} .

5.1.7 Saturation Slope Coefficient, κ

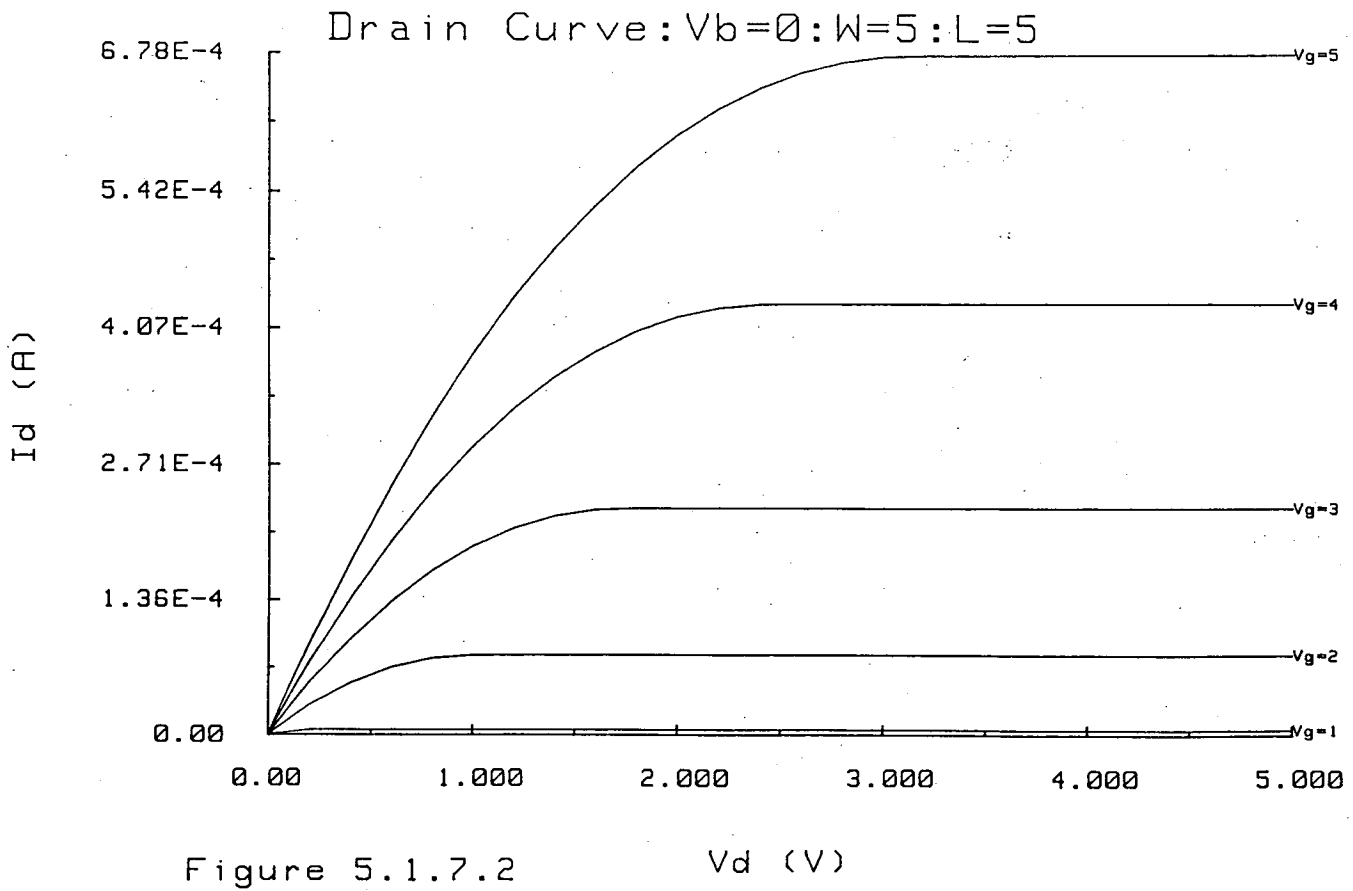
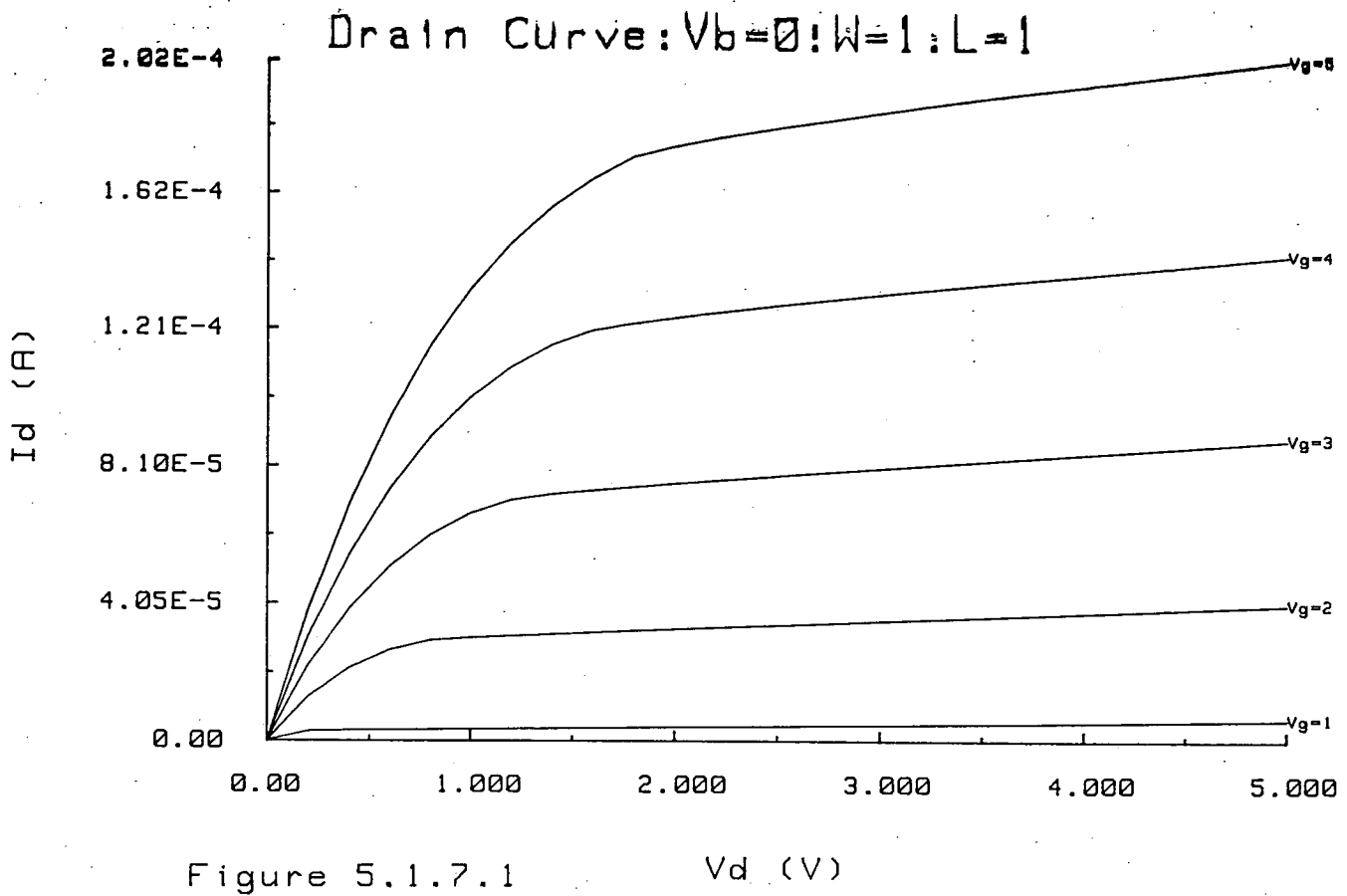
The saturation slope coefficient is also a strong function of channel length. This can be seen in the drain voltage:drain current characteristics of $1\mu\text{m} \times 1\mu\text{m}$ device and a $5\mu\text{m} \times 5\mu\text{m}$ device (figures 5.1.7.1 and 5.1.7.2). The effect is due to the high drain voltage leading to a proportion of the channel length no longer being under gate control. This region only becomes a significant proportion of the channel for transistors which are less than $3\mu\text{m}$ long. The κ values which result for the $1\mu\text{m} \times 1\mu\text{m}$, $1.2\mu\text{m} \times 1.2\mu\text{m}$, $1.5\mu\text{m} \times 1.5\mu\text{m}$ and $2\mu\text{m} \times 2\mu\text{m}$ are 0.086, 0.067, 0.059 and 0.030 respectively. The effect is small at larger dimensions and so the values for κ which are obtained are not significant. If a single parameter is to be used for all channel lengths then the value for the minimum permissible length should be selected.

In CASMOS, the base saturation slope coefficient C_{SAT} is modified to provide λ which corresponds to the parameter κ in SPICE. It is found to be a constant for channel lengths of less than about $5\mu\text{m}$ and above $5\mu\text{m}$, a L^{-1} dependence is used so that

$$\lambda = \frac{C_{SAT}}{L_{SAT}} \quad LR < L_{SAT} \quad 5.1.7.1$$

or else

$$\lambda = \frac{C_{SAT}}{LR} \quad 5.1.7.2$$



5.1.8 Formulation of Parameter Sets for Different Geometries

Most parameters show some sort of dependence on the geometry of the devices from which they are extracted. The majority of parameters have a greater dependence on channel length rather than channel width and these dependences generally become more significant at lengths of less than $3\ \mu\text{m}$. The parameters V_{to} and γ show moderate increases as device geometry is reduced. On the other hand, the carrier mobility μ_o increases dramatically from $0.063\ m^2\ V_s^{-1}$ for the $2\ \mu\text{m} \times 2\ \mu\text{m}$ device to $0.109\ m^2\ V_s^{-1}$ for the $1\ \mu\text{m} \times 1\ \mu\text{m}$ device. This is a 73% increase. Drain modulation of mobility, represented by v_{max} , is especially important in lengths of less than $3\ \mu\text{m}$ but in NMOS, where carrier mobility is higher, its effect is also significant in longer channels. Static feedback and saturation slope (modelled by η and κ respectively) are both phenomena which are relatively insignificant at lengths above $3\ \mu\text{m}$ but are very important in short channels.

In practice, the vast majority of devices in digital circuits are of minimum channel length. Devices with less than unity aspect ratios are very weak and are therefore rarely required, and to obtain greater drive current, the width is scaled up. However in many analog applications better matching and noise performance can be obtained with larger devices and consequently both dimensions tend to be scaled. Hence, the parameters for simulation have to be derived with all channel lengths (assuming that length variation is the dominant geometrical influence on performance) in mind. If for example, there is a requirement that a single set of parameters should be used for a $2\ \mu\text{m}$ process then the following is proposed. Most parameters (V_{to} , γ , N_{fs} , δ , μ_o and θ) should be extracted from a device whose length is such that it exhibits average behaviour; possibly $3\ \mu\text{m}$ for a $2\ \mu\text{m}$ process. The significance of the small geometry effects modelled by η , v_{max} and κ increase greatly as length is reduced and therefore these should be the values found for the $2\ \mu\text{m}$ device. Another option, where greater accuracy is needed, is to use a version of the preprocessor lookup table. A suggestion would be to have four sets of parameters: one for $2\ \mu\text{m}$ long devices, one for $3\ \mu\text{m}$ devices, one for $4/5\ \mu\text{m}$ devices and one for $6\ \mu\text{m}$ and above.

In principle the idea behind CASMOS, to empirically model the geometrical variation in parameters, is good. However in the example above for a $2\ \mu\text{m}$ process, given that only four sets of parameters are required to provide fairly accurate

simulation over all geometries, the CASMOS approach would seem to require excessive effort. Possibly with the advent of a $1\mu\text{m}$ process, where a lookup table would be required to hold parameters for six or seven lengths, modelling the geometrical variation of parameters would be worthwhile. The accuracy of the CASMOS equations would have to be investigated but it may lead to more accurate simulation of the longer devices

5.2 Manufacturing Variations and Parameters

5.2.1 The Importance of Manufacturing Variations

In Chapter 1, it was established that for reasons of both time and cost, accurate circuit simulation is essential in the design of integrated circuits. A set of techniques for physical parameter extraction are described in Chapter 4 and the factors which decide how to simulate devices of different sizes are discussed in Section 5.1. A further consideration which must be addressed when attempting to design circuits with optimum performance are random variations in the manufacturing process and their impact on device parameters and hence circuit operation.

To allow for manufacturing variations, most designers use a set of worst case parameters. Successful simulation with a reasonable margin for error should ensure that the circuit operates within the required specifications. In extreme cases, where optimum performance is desired, fully operational circuits may be picked out at functional test. Chips not meeting the tight specifications of speed or power dissipation for example, can be sold more cheaply and hence need not be wasted.

Designers rarely know what actually constitutes a worst case set of parameters. The set is usually derived from an average set of values by altering each parameter in a direction which would seem undesirable and correlation between different parameters is not taken into account. In conjunction with altering parameters, parasitic elements such as resistors and capacitors are overestimated and a reasonable margin for error is required in the simulation. The result is that circuits are designed with more redundancy than is strictly necessary and hence performance is restricted.

By measuring parameters at sites spread across a wafer, an investigation into manufacturing variations on a single wafer was made. Other researchers have made a similar analysis by looking at yield as a function of wafer position.⁶⁴ Their conclusion was that yield fell off very markedly towards the periphery of the wafer and hence random defects did not account for this loss. N-channel parameters were measured for $3.5\mu\text{m}$ devices across two four inch wafers which were fabricated using two different CMOS silicon gate processes. The processes are variations of a p-well CMOS process used for fabricating $5\mu\text{m}$ digital standard parts. P-channel parameters were measured

for 3.5 μ m devices on two other similar wafers. The pattern of chips which were measured is shown in figure 5.2.1.1. With the flat side of the wafer at the bottom, every fourth chip horizontally contained suitable discrete devices for obtaining a complete set of parameters. A margin of two chips was ignored around the periphery of the wafer and there were three drop-in chips placed in a vertical column in the middle of the wafer. Hence if all the devices were functional, 170 sites would be measured on each wafer. The HP Basic Statistics and Data Manipulation package was then used to calculate the means, maximums and minimums and the correlations between parameters. Wafer maps were also plotted using further HP software available with the HP4062B Semiconductor Test System.

5.2.2 Best and Worst Case Parameter Sets

NMOS parameters were measured across two wafers and the resulting statistics for the variations in parameters are listed in Tables 5.2.2.1 and 5.2.2.2. Similarly two sets of PMOS parameters were measured and the statistics for their variations are provided in Tables 5.2.2.3 and 5.2.2.4. From these figures, circuit designers would like to select a set of worst and best case parameters but the correct method for this selection is difficult to determine. If consideration is restricted to digital circuits then the current carrying capability of the device is most important. It can be deduced whether increasing a parameter will decrease or increase the current. Thus, if a parameter leads to a decrease in current when its value is increased, the mean value $+3\sigma$ can be taken as the worst case parameter and the mean value -3σ can be taken as the best case parameter. Correspondingly, if a parameter leads to an increase in current when its value increases, the mean value $+3\sigma$ can be taken as the best case parameter and the mean value -3σ can be taken as the worst case parameter. The contribution each parameter makes to device current and the best and worst case parameter sets which result are listed in Table 5.2.2.5. This method of deducing parameter sets for circuit design is probably the most widely used.

To help assess the accuracy of these techniques in modelling the manufacturing variations in device operation, it is useful to measure the drive current in the devices. Here I_{drive} has been defined as the current flowing in the device with 5V applied to gate and drain and the source and substrate are grounded. The I_{drive} definition for NMOS and PMOS is illustrated in figure 5.2.2.1. The parameter set

***** WAFER PATTERN *****

Figure 5.2.1.1

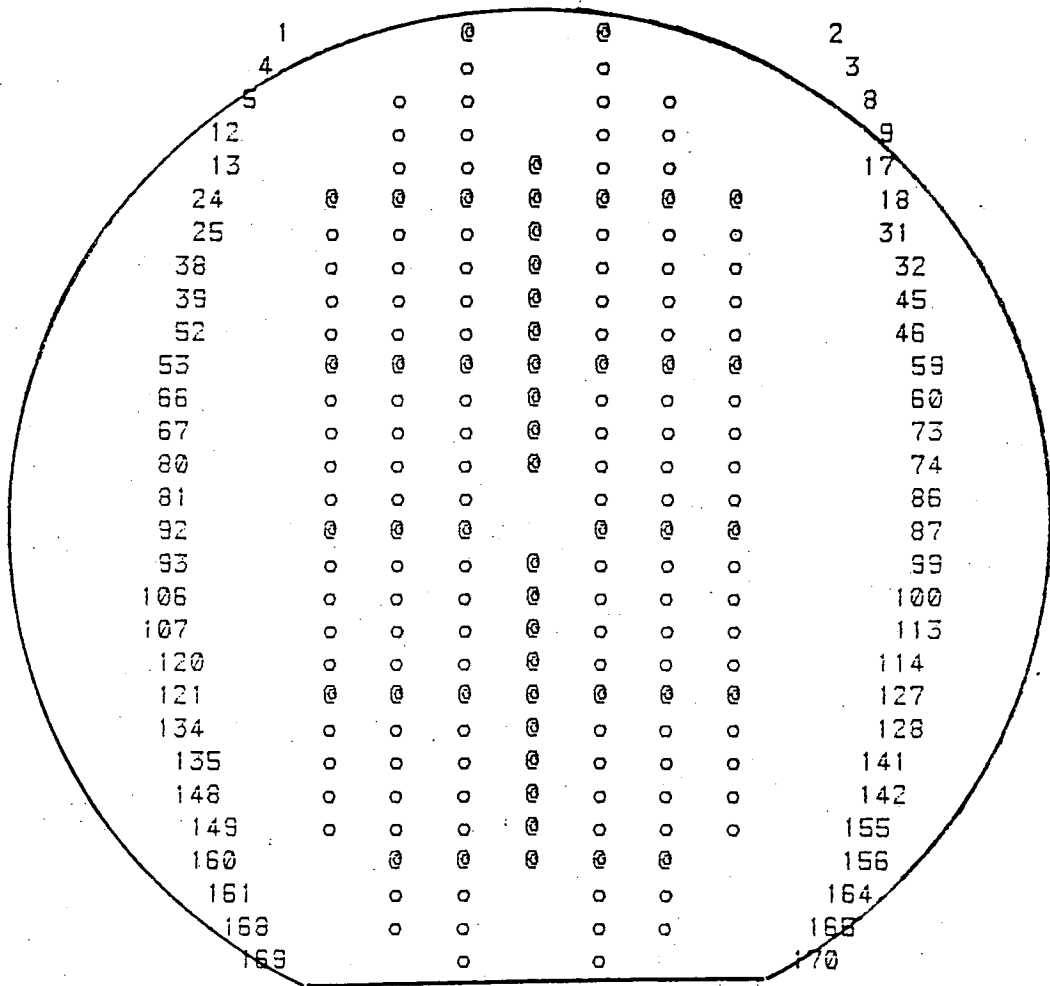


Table 5.2.2.1 NMOS Parameter Statistics I (163 Measurements)

Parameter	Mean	σ	Minimum	Maximum	-3σ	$+3\sigma$
t_{ox} (Å)	625	6.5	613	646	606	645
V_{to} (V)	0.51	0.023	0.45	0.57	0.44	0.58
γ ($V^{\frac{1}{2}}$)	1.39	0.033	1.25	1.46	1.29	1.49
μ_o ($cm^2.Vs^{-1}$)	533	24.5	481	594	460	607
θ (V^{-1})	0.027	0.002	0.021	0.034	0.020	0.034
v_{max} ($m.s^{-1}$)	1.80E5	6.0E3	1.60E5	1.96E5	1.62E5	1.98E5
N_{fs} (cm^{-2})	1.7E15	0.35E15	1.1E15	2.6E15	0.65E15	2.75E15
L_d (μm)	1.06	0.042	0.94	1.20	0.93	1.19
Δ_w (μm)	0.85	0.041	0.75	0.98	0.73	0.97
δ	0.22	0.025	0.16	0.28	0.15	0.30
η	0.041	0.006	0.028	0.068	0.023	0.059
κ	0.30	0.02	0.24	0.36	0.24	0.36

Table 5.2.2.2 NMOS Parameter Statistics II (152 Measurements)

Parameter	Mean	σ	Minimum	Maximum	-3σ	$+3\sigma$
t_{ox} (Å)	760	5.7	752	779	743	777
V_{to} (V)	0.82	0.026	0.76	0.88	0.74	0.90
γ ($V^{\frac{1}{2}}$)	2.30	0.097	2.08	2.85	2.04	2.62
μ_o ($cm^2.Vs^{-1}$)	543	51	350	683	390	696
θ (V^{-1})	-0.006	0.005	-0.031	0.003	-0.021	0.009
v_{max} ($m.s^{-1}$)	3.1E5	0.25E5	2.6E5	3.8E5	2.3E5	3.9E5
N_{fs} (cm^{-2})	1.3E15	0.3E15	6.5E14	1.9E15	4E14	2.2E15
L_d (μm)	1.27	0.051	1.15	1.45	1.12	1.42
Δ_w (μm)	1.22	0.063	1.09	1.37	1.03	1.41
δ	0.201	0.032	0.124	0.340	0.105	0.297
η	0.011	0.002	0.003	0.019	0.005	0.017
κ	0.278	0.023	0.195	0.336	0.209	0.347

Table 5.2.2.3 PMOS Parameter Statistics I (144 Measurements)						
Parameter	Mean	σ	Minimum	Maximum	-3σ	$+3\sigma$
t_{ox} (Å)	693	10.0	676	721	663	723
V_{to} (V)	-0.63	0.018	-0.58	-0.68	-0.58	-0.68
γ ($V^{\frac{1}{2}}$)	1.28	0.023	1.22	1.34	1.21	1.35
μ_o ($cm^2.Vs^{-1}$)	164	8.2	143	185	139	189
θ (V^{-1})	0.024	0.0037	0.017	0.055	0.013	0.035
v_{max} ($m.s^{-1}$)	-	-	-	-	-	-
N_{fs} (cm^{-2})	3.6E15	0.55E15	2.8E15	5.1E15	1.9E15	5.3E15
L_d (μm)	1.05	0.055	0.94	1.20	0.89	1.22
Δ_w (μm)	0.85	0.040	0.75	0.95	0.73	0.97
δ	0.31	0.025	0.25	0.38	0.24	0.39
η	0.044	0.0056	0.032	0.066	0.027	0.061
κ	0.04	0.036	0.0014	0.42	-0.068	0.15

Table 5.2.2.4 PMOS Parameter Statistics II (115 Measurements)						
Parameter	Mean	σ	Minimum	Maximum	-3σ	$+3\sigma$
t_{ox} (Å)	866	9.3	822	894	838	894
V_{to} (V)	-1.29	0.025	-1.24	-1.38	-1.22	-1.37
γ ($V^{\frac{1}{2}}$)	2.05	0.047	1.97	2.18	1.91	2.19
μ_o ($cm^2.Vs^{-1}$)	262	25	216	331	187	337
θ (V^{-1})	-0.002	0.004	-0.014	0.004	-0.014	0.010
v_{max} ($m.s^{-1}$)	-	-	-	-	-	-
N_{fs} (cm^{-2})	3.2E15	0.44E15	2.5E15	4.2E15	1.9E15	4.5E15
L_d (μm)	1.10	0.072	0.95	1.29	0.88	1.32
Δ_w (μm)	1.28	0.046	1.15	1.39	1.14	1.42
δ	0.184	0.024	0.129	0.243	0.112	0.256
η	0.024	0.004	0.014	0.036	0.012	0.036
κ	0.127	0.028	0.057	0.183	0.043	0.211

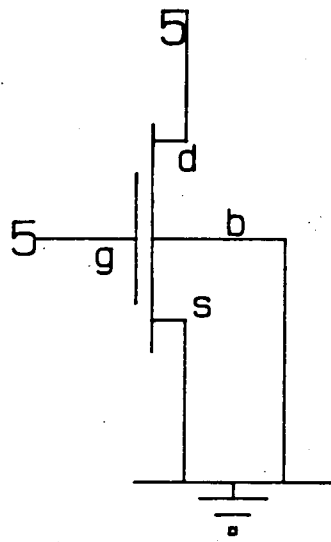
Table 5.2.2.5 Best and Worst Case Parameters Using +/-3 σ

Parameter	Contribution to Current	NMOS I		NMOS II	
		Worst	Best	Worst	Best
t_{ox} (Å)	-	645	606	777	743
V_{to} (V)	-	0.58	0.44	0.90	0.74
γ ($V^{\frac{1}{2}}$)	-	1.49	1.29	2.62	2.04
μ_o ($cm^2.Vs^{-1}$)	+	460	607	390	696
θ (V^{-1})	-	0.034	0.020	0.009	-0.021
v_{max} ($m.s^{-1}$)	+	1.62E5	1.98E5	2.3E5	3.9E5
N_{fs} (cm^{-2})	+	0.65E15	2.75E15	4E14	2.2E15
L_d (μm)	+	0.93	1.19	1.12	1.42
Δ_w (μm)	-	0.97	0.73	1.41	1.03
δ	-	0.30	0.15	0.297	0.105
η	+	0.023	0.059	0.005	0.017
κ	+	0.24	0.36	0.209	0.347

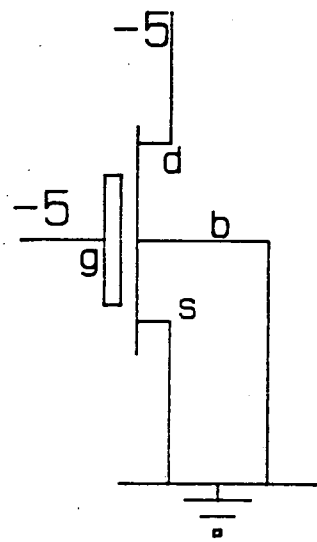
Table 5.2.2.5 Best and Worst Case Parameters Using +/-3 σ (cont)

Parameter	Contribution to Current	PMOS I		PMOS II	
		Worst	Best	Worst	Best
t_{ox} (Å)	-	723	663	894	838
V_{to} (V)	-	-0.68	-0.58	-1.37	-1.22
γ ($V^{\frac{1}{2}}$)	-	1.35	1.21	2.19	1.91
μ_o ($cm^2.Vs^{-1}$)	+	139	189	187	337
θ (V^{-1})	-	0.035	0.013	0.010	-0.014
v_{max} ($m.s^{-1}$)	+	-	-	-	-
N_{fs} (cm^{-2})	+	1.9E15	5.3E15	1.9E15	4.5E15
L_d (μm)	+	0.89	1.22	0.88	1.32
Δ_w (μm)	-	0.97	0.73	1.42	1.14
δ	-	0.39	0.24	0.256	0.112
η	+	0.027	0.061	0.012	0.036
κ	+	-0.068	0.15	0.043	0.211

Figure 5.2.2.1
Idrive Definition



(a) NMOS



(b) PMOS

D=

from the chip with least I_{drive} were taken to be worst case and those with greatest I_{drive} were taken to be best case. These parameter sets are shown in table 5.2.2.6.

The drain voltage characteristics resulting from both of these sets of best and worst parameters are plotted in figures 5.2.2.2, 5.2.2.3, 5.2.2.4 and 5.2.2.5. The key to these figures is

—————	best case using mean $\pm 3\sigma$
— — — —	best case using I_{drive}
- - - - -	worst case using I_{drive}
- - - - -	worst case using mean $\pm 3\sigma$

For each of the four wafer spreads measured in this experiment, altering each parameter separately leads to a much greater spread in transistor currents than is realistically the case. Figure 5.2.2.2 shows that the best or worst cases when using $\pm 3\sigma$ indicate that there is a possible 60% difference in current whereas there is only a 10% difference in measured drive currents. A similar pattern is found for each of the other parameter sets as shown in the figures.

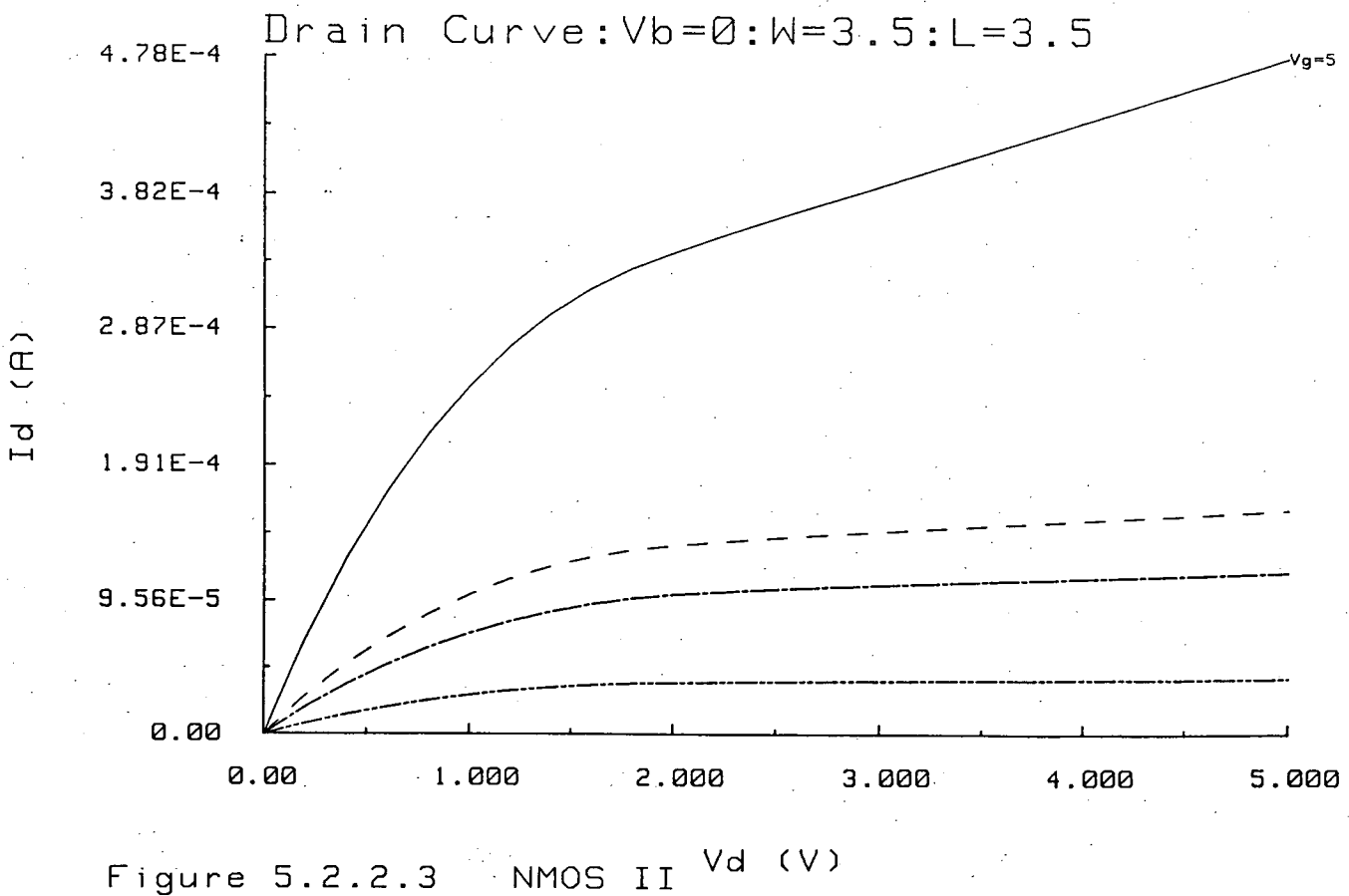
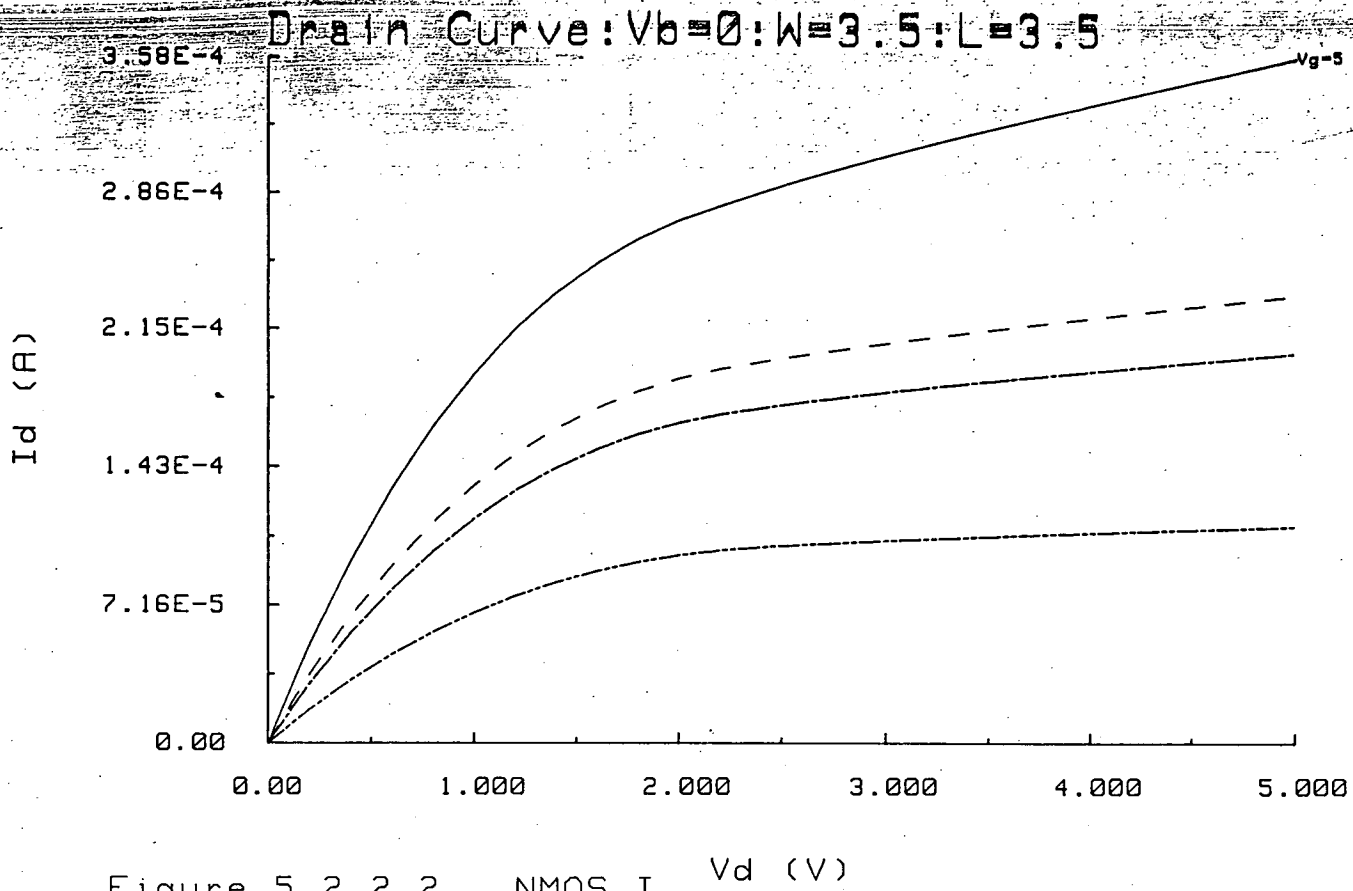
5.2.3 Correlation Between Parameters

The reason for the discrepancy between the two methods for determining best and worst cases mentioned above is that the parameters have some correlation between them. This is not taken into account when choosing mean $\pm 3\sigma$ values. Some of the correlations between parameters have physical bases. For instance, a higher channel implant will increase threshold voltage V_{to} but will also cause an increase in the substrate bias coefficient γ so a reasonable correlation can be expected between them. This should be further pronounced because both V_{to} and γ increase when t_{ox} increases.

The correlation between parameters for the four sets measured are shown in tables 5.2.3.1, 5.2.3.2, 5.2.3.3 and 5.2.3.4. A value of 1 indicates perfect correlation, -1 indicates perfect negative correlation and 0 indicates that there is no correlation.

Table 5.2.2.6 Best and Worst Case Parameters Using Idrive					
Parameter	Contribution to Current	NMOS I		NMOS II	
		Worst	Best	Worst	Best
t_{ox} (Å)	-	624	632	757	758
V_{to} (V)	-	0.51	0.45	0.83	0.80
γ ($V^{1/2}$)	-	1.39	1.35	2.21	2.23
μ_o ($cm^2.Vs^{-1}$)	+	536	558	577	511
θ (V^{-1})	-	0.027	0.030	-0.003	-0.004
v_{max} ($m.s^{-1}$)	+	1.84E5	1.77E5	3.07E5	2.61E5
N_{fs} (cm^{-2})	+	1.72E15	1.89E15	1.94E15	1.27E15
L_d (μm)	+	1.11	1.16	1.16	1.29
Δ_w (μm)	-	0.90	0.86	1.24	1.12
δ	-	0.242	0.270	0.124	0.275
η	+	0.039	0.048	0.010	0.013
κ	+	0.282	0.248	0.313	0.262

Table 5.2.2.6 Best and Worst Case Parameters Using Idrive (cont)					
Parameter	Contribution to Current	PMOS I		PMOS II	
		Worst	Best	Worst	Best
t_{ox} (Å)	-	676	686	862	864
V_{to} (V)	-	-0.65	-0.62	-1.29	-1.30
γ ($V^{1/2}$)	-	1.29	1.26	2.02	2.18
μ_o ($cm^2.Vs^{-1}$)	+	152	167	300	232
θ (V^{-1})	-	0.020	0.023	0.001	-0.014
v_{max} ($m.s^{-1}$)	+	-	-	-	-
N_{fs} (cm^{-2})	+	5.10E15	4.32E15	2.91E15	3.30E15
L_d (μm)	+	1.01	1.11	0.95	1.29
Δ_w (μm)	-	0.89	0.88	1.32	1.37
δ	-	0.265	0.318	0.133	0.209
η	+	0.037	0.046	0.027	0.017
κ	+	0.041	0.014	0.173	0.063



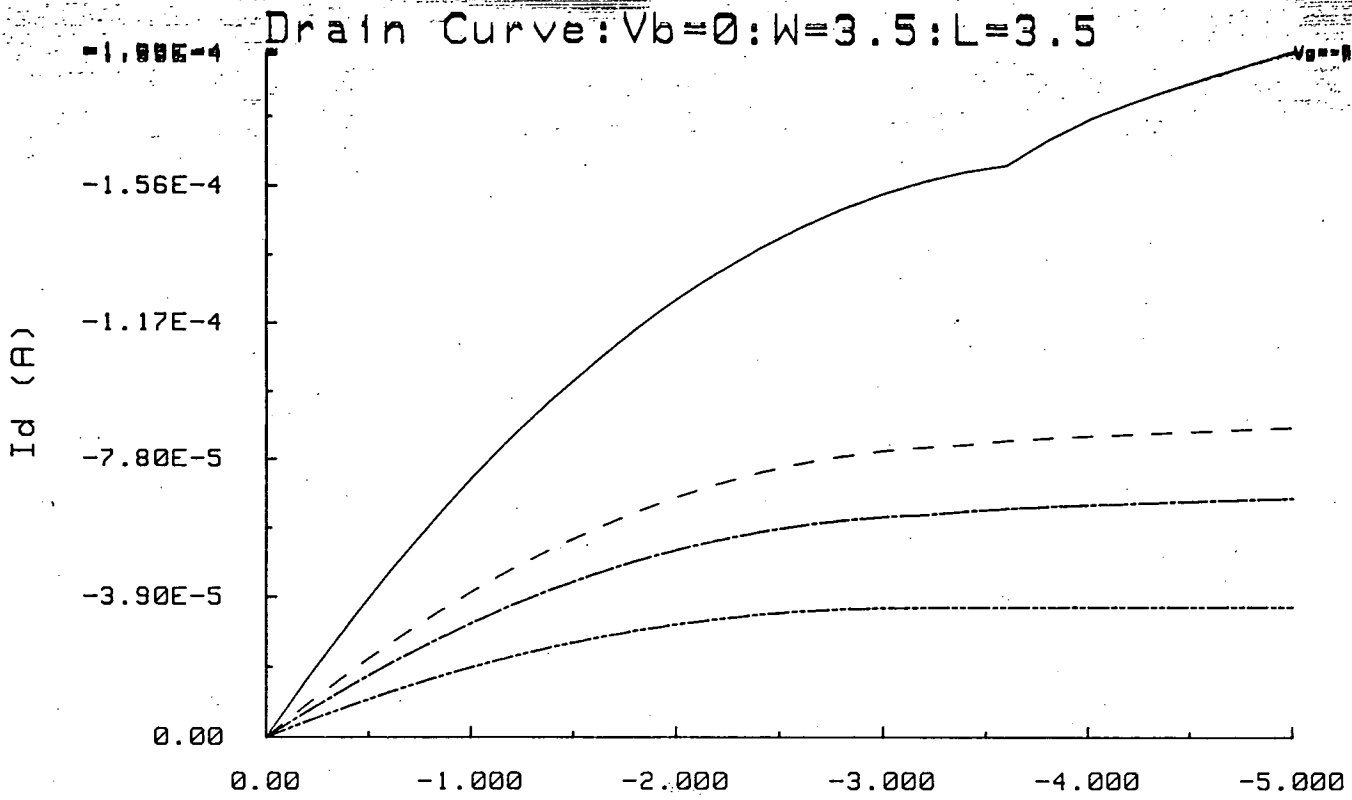


Figure 5.2.2.4 PMOS I V_d (V)

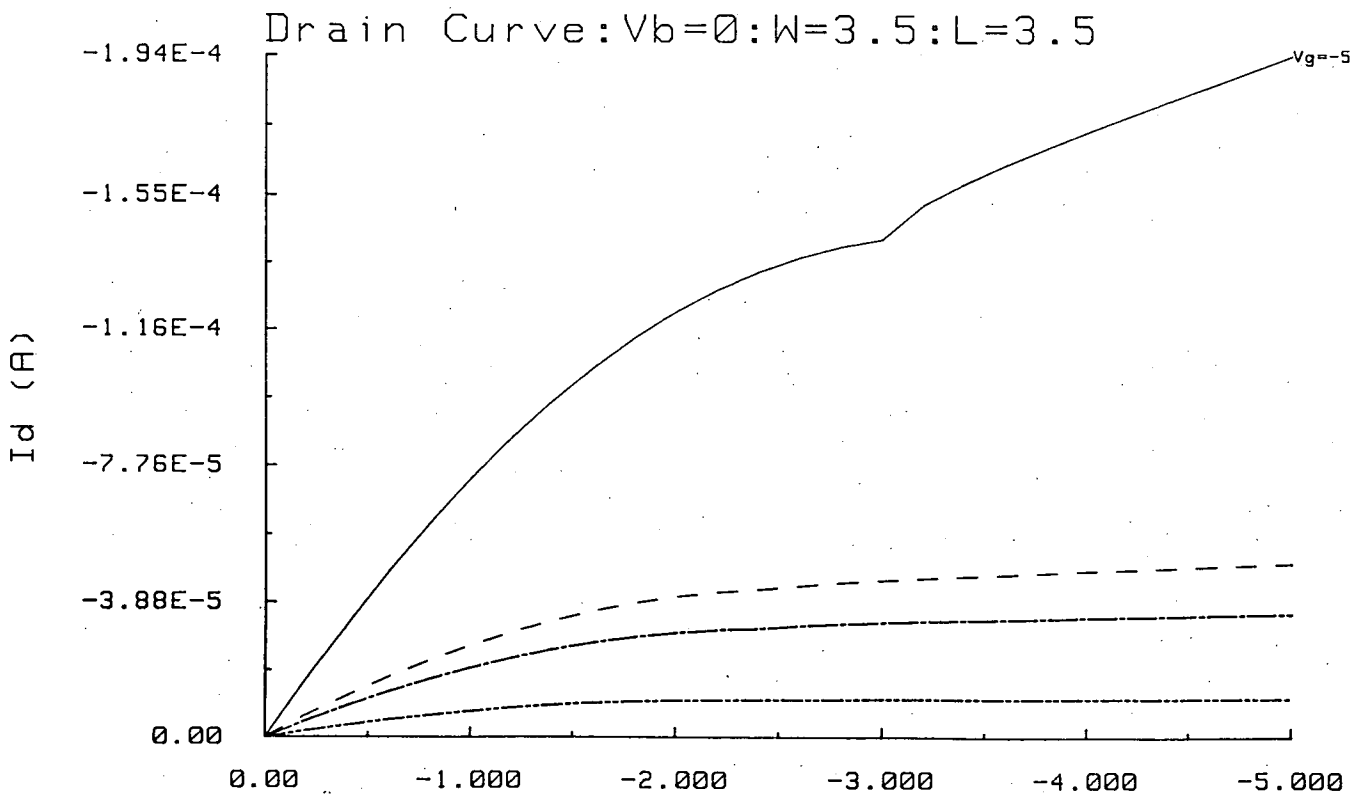


Figure 5.2.2.5 PMOS II V_d (V)

Any value of magnitude less than 0.15 has been ignored as being insignificant. From these figures, it was difficult to establish where the major correlations lay and so the average correlation for first order parameters for all four parameter sets are shown in table 5.2.3.5. It should be noted that the PMOS V_{to} s are negative and so the correlations with V_{to} were made positive if they were negative and vice versa in order to provide the values for average correlation.

The significant correlations where all four parameter sets agree are

Parameters	Average Correlation
V_{to}, γ	0.40
γ, μ_o	-0.44
γ, θ	-0.67
γ, Δ_w	0.46
μ_o, θ	0.45

The parameters V_{to} and γ vary together as predicted. The second combination, γ and μ_o can again be explained by a higher implant dose resulting in a higher substrate bias coefficient and lower mobility. The mobility modulation coefficient shows a high negative correlation with γ and a high positive correlation with μ_o . Both of these cases highlight the problems associated with choosing mean values $\pm 3\sigma$ individually since the parameters which vary together have opposite effects on device currents. A high μ_o tends to occur in the same device as a high θ and so μ_o would lead to an increase in current but θ would lead to decrease. This implies that the extremes of operation predicted by choosing each worst case parameter individually do not occur in practice. The final major correlation between first order parameters is between γ and Δ_w . A possible explanation for this is the diffusion of the field implant into the channel which would increase both the reduction in channel width and the average channel impurity concentration.

One other statistic which stands out is the average correlation of -0.79 between L_d and κ . This is probably a consequence of the fact that the extraction was carried out on devices with short effective lengths of around $1.5\mu\text{m}$. As shown in Section 5.1, the absolute value of κ is dependent on the device length. It is not

Table 5.2.3.5 Average Correlation of Major Parameters

Parameter	V_{io}	γ	μ_o	θ	L_d	Δ_w	κ
t_{ox}	0.14	0.12	0.22	-	0.15	-	-
V_{io}		0.40	-	-0.11	-0.23	0.24	-
γ			-0.44	-0.67	0.36	0.46	-
μ_o				0.45	-0.36	0.29	-
θ					-0.51	-0.47	-
L_d						0.32	-0.79

geometry independent and so a small perturbation in an effective length of $1.5\mu\text{m}$ leads to a significant variation in κ .

Wafer maps were plotted for each parameter for each of the four parameter sets which were measured. The wafer maps for V_{io} , γ , μ_o , θ , L_d and κ for the NMOS I data are included in figures 5.2.3.1 to 5.2.3.6. Each wafer map was examined and areas where a particular parameter was generally above or below its average value were marked.

When the maps for V_{io} and γ were compared, the positive correlation can be seen in the high valued areas at the bottom of the wafer and the low valued areas at the top right of the wafer. The opposite case for μ_o upholds the negative correlation and θ shows a similar pattern to μ_o leading to positive correlation. The very high negative correlation between L_d and κ is evident from their wafer maps in figures 5.2.3.5 and 5.2.3.6.

5.2.4 Worst Case Parameters and Circuits

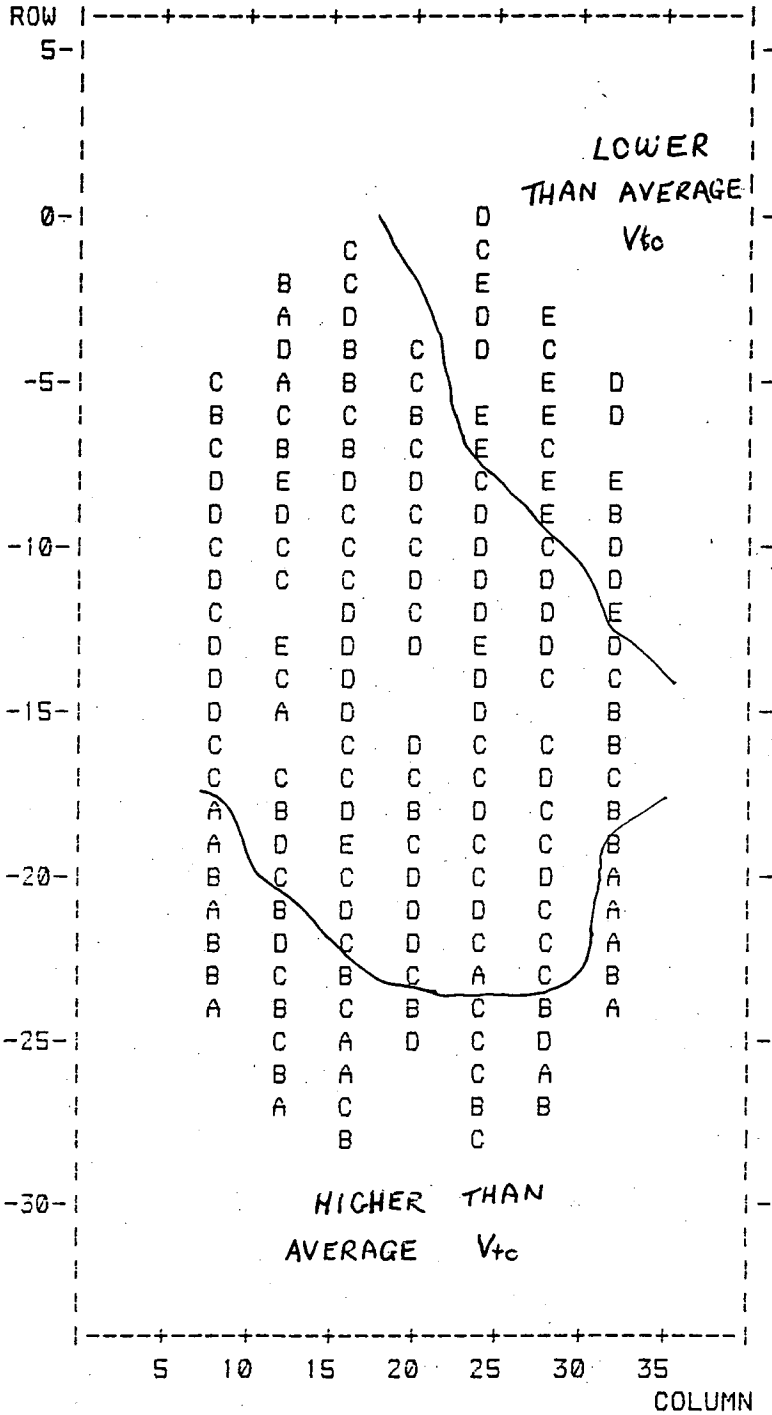
The simplest digital CMOS circuit is the inverter. Other gates can easily be condensed to look like an inverter in order to analyse the switching point and output rise and fall times. The switching voltage of an inverter is derived in APPENDIX D^{65,66} and is given by

$$V_{IN}(switch) = \frac{\left(\frac{Beta_p}{Beta_n}\right)^{\frac{1}{2}} (V_{DD} - V_{TP}) + V_{TN}}{1 + \left(\frac{Beta_p}{Beta_n}\right)^{\frac{1}{2}}} \quad 5.2.4.1$$

Figure 5.2.4.1 shows a schematic diagram of an inverter with a capacitor connected to its output to simulate a fairly large output load. The inverter, without a capacitor on its output was driven with a pulse and its response for various combinations of NMOS and PMOS parameters is shown in figures 5.2.4.2 to 5.2.4.5. The parameters were chosen to give the extreme of operation i.e. best case NMOS and worst case PMOS or vice versa. Both the realistic and 3σ sets were used. The figure numbers, parameters used and predicted switching points according to equation 5.2.4.1 are listed below

Parameters across a Wafer

DATE 14 Aug 1986



VAR NAME IVt0 1

CHIPS/WAFER = 163

**** LIMIT ****

A >= 0.540E-00

B >= 0.520E-00

C >= 0.500E-00

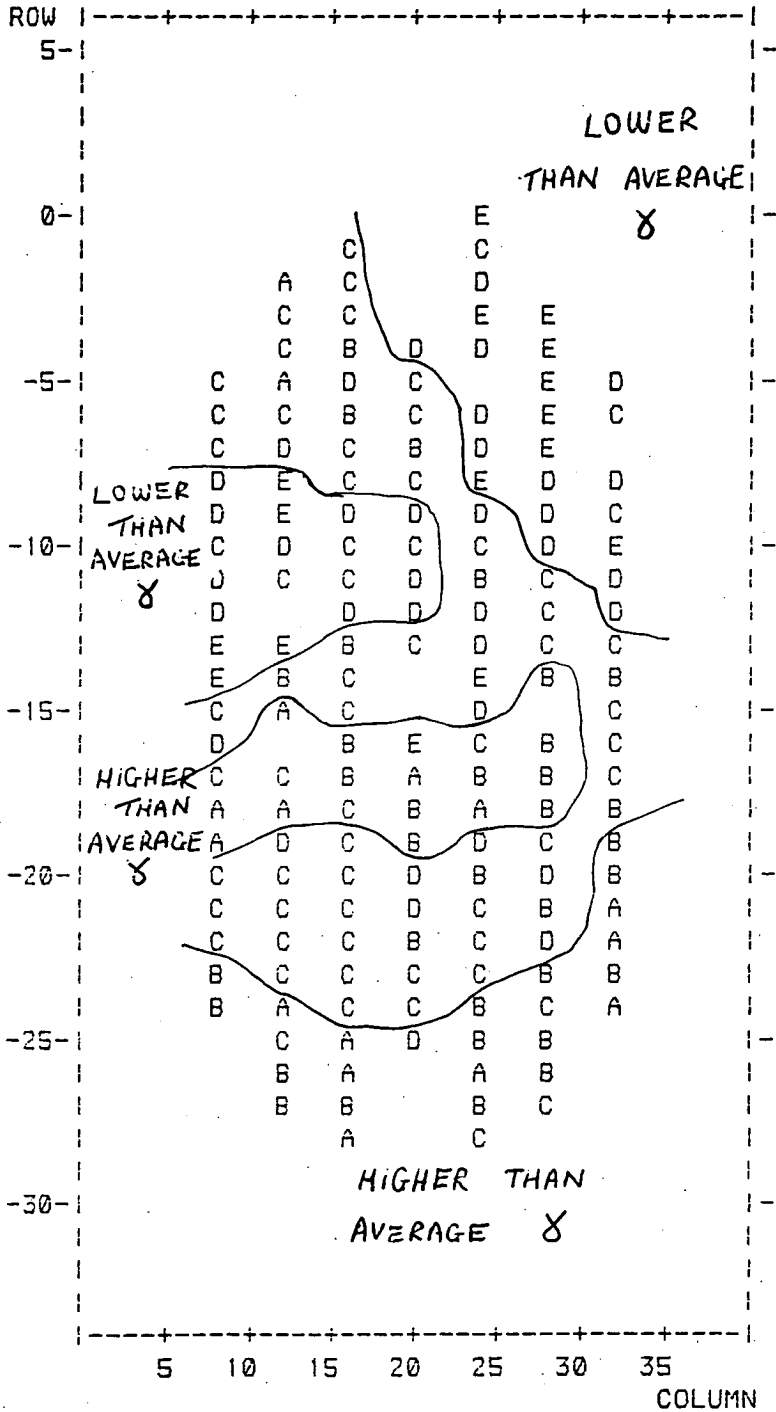
D >= 0.480E-00

E < 0.480E-00

Figure 5.2.3.1

Parameters across a Wafer

DATE | 4 Aug 1966 |



VAR NAME |Gamma |

CHIPS/WAFER = 163

**** LIMIT ****

A >= 1.440E-00

B >= 1.410E-00

C >= 1.380E-00

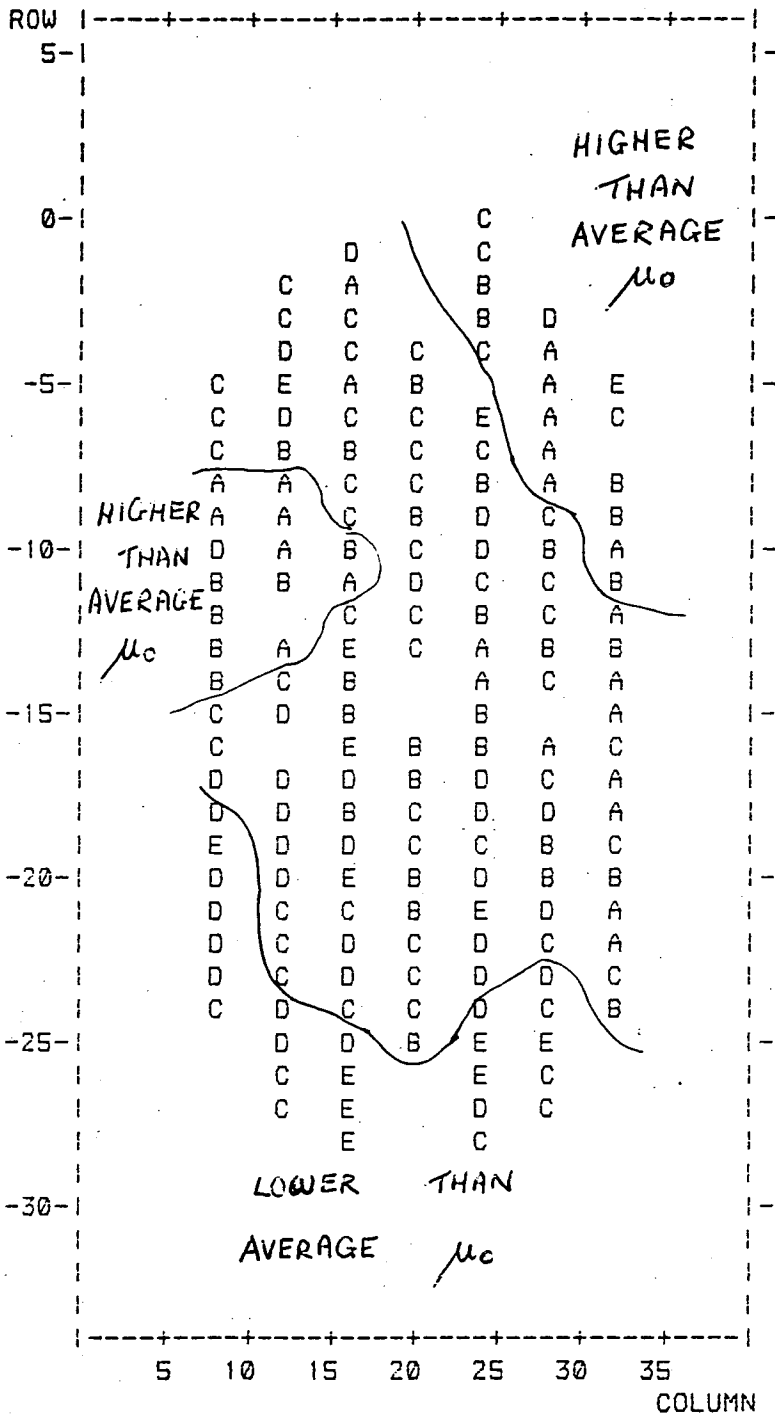
D >= 1.350E-00

E < 1.350E-00

Figure 5.2.3.2

Parameters across a Wafer

DATE | 4 Aug 1986 |

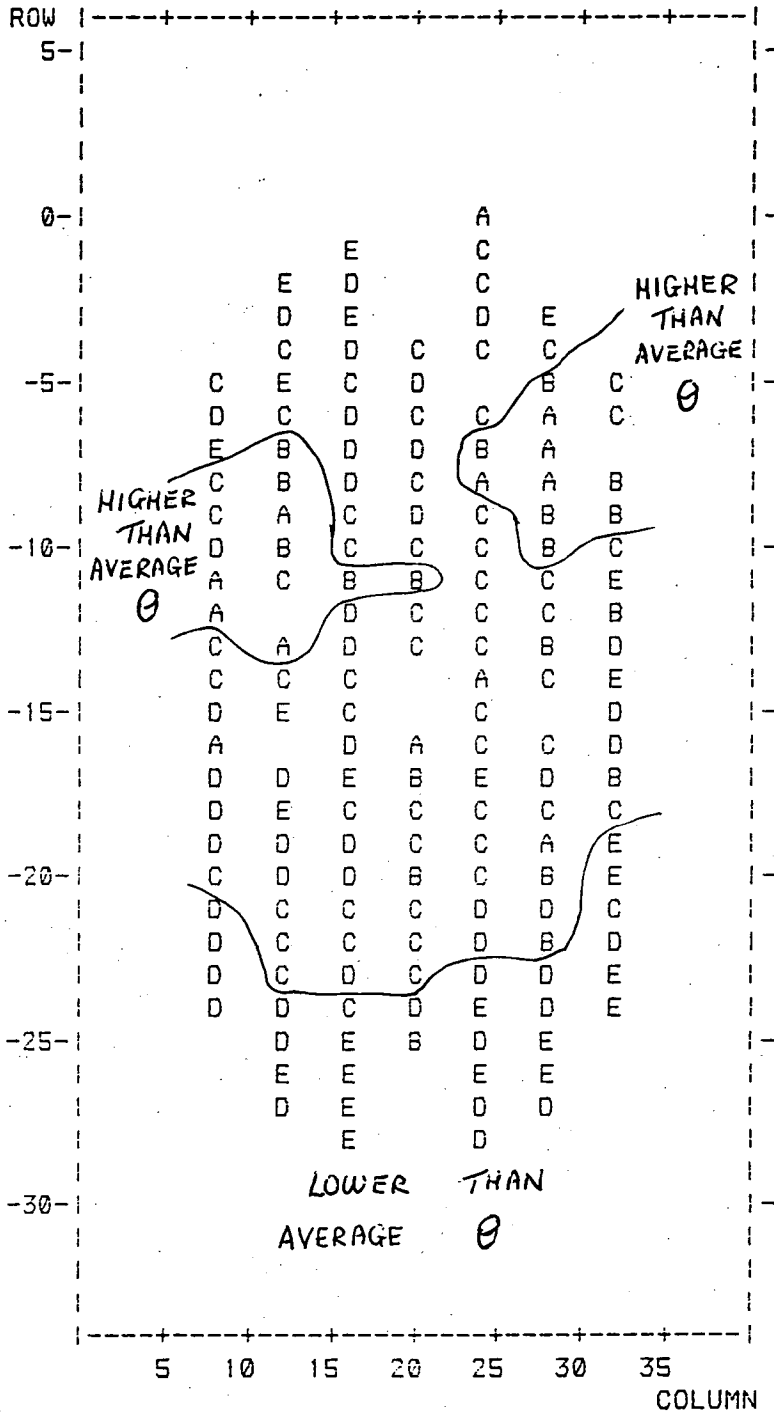


VAR NAME 1μ₀ 1
 CHIPS/WAFER = 163
 **** LIMIT ****
 A >= 0.056E-00
 B >= 0.054E-00
 C >= 0.052E-00
 D >= 0.050E-00
 E < 0.050E-00

Figure 5.2.3.3

Parameters across a Wafer

DATE | 4 Aug 1986 |



VAR NAME | Theta |

CHIPS/WAFER = 163

**** LIMIT ****

A >= 0.310E-01

B >= 0.290E-01

C >= 0.270E-01

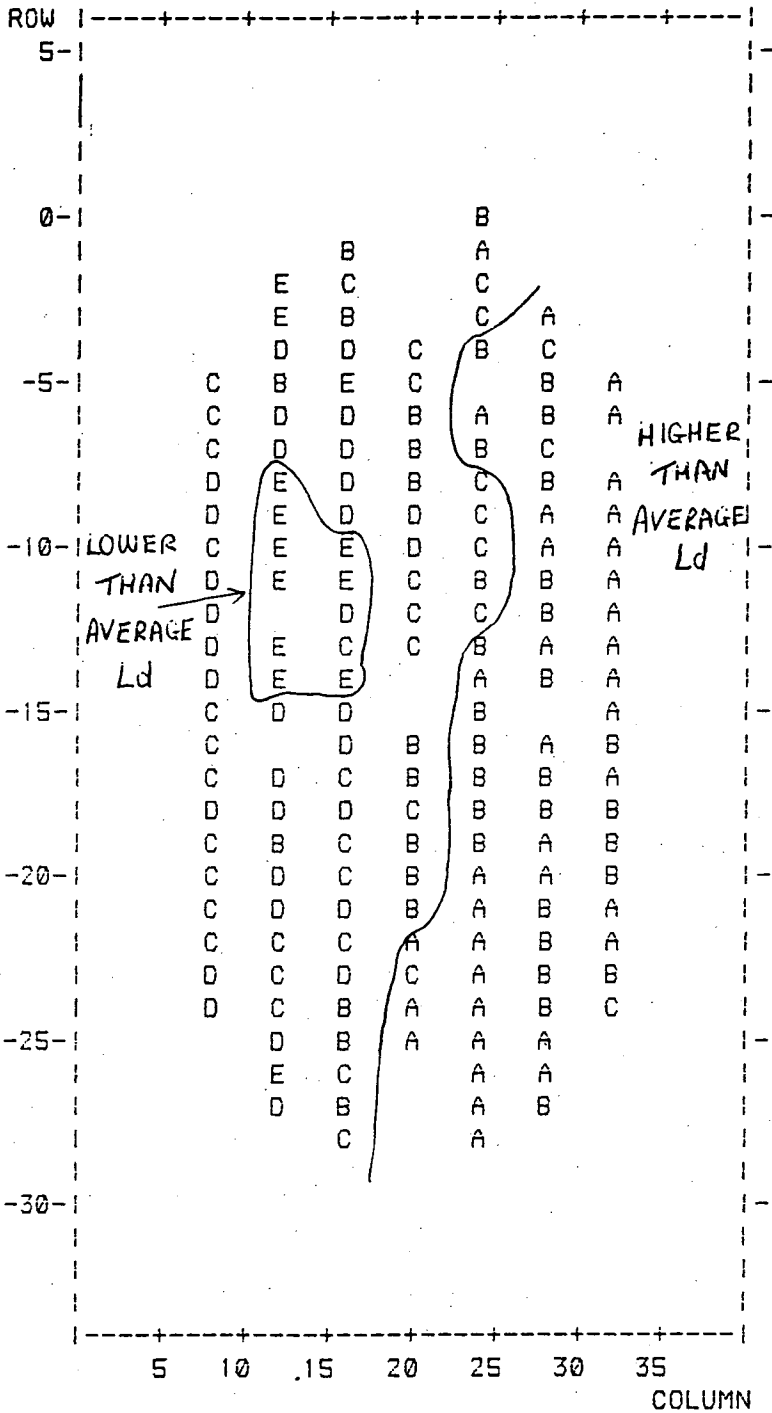
D >= 0.250E-01

E < 0.250E-01

Figure 5.2.3.4

Parameters across a Wafer

DATE | 4 Aug 1986 |



VAR NAME | Ld |

CHIPS/WAFER = 163

**** LIMIT ****

A >= 1.090E-06

B >= 1.060E-06

C >= 1.030E-06

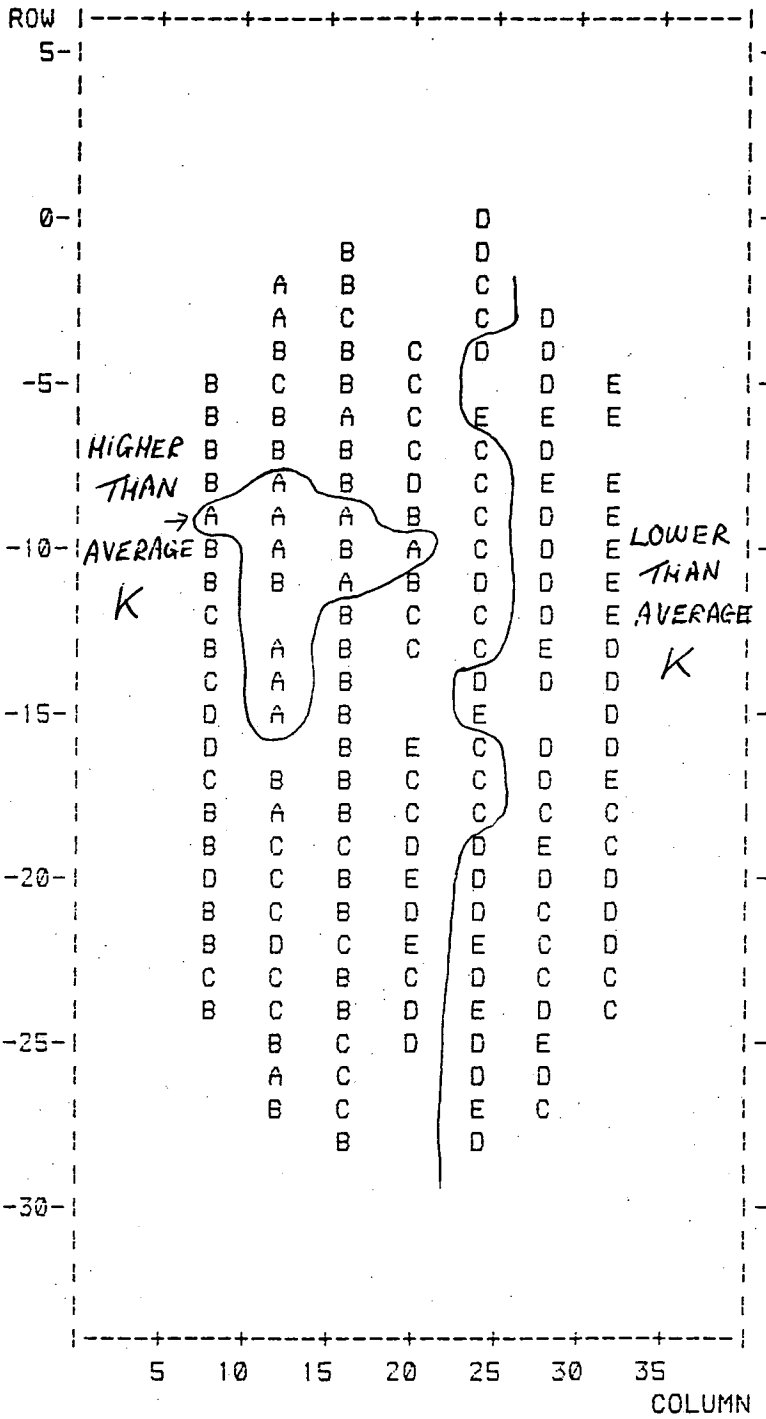
D >= 1.000E-06

E < 1.000E-06

Figure 5.2.3.5

Parameters across a Wafer

DATE | 4 Aug 1986 |



VAR NAME |Kappa|

CHIPS/WAFER = 163

**** LIMIT ****

A >= 0.325E-00

B >= 0.310E-00

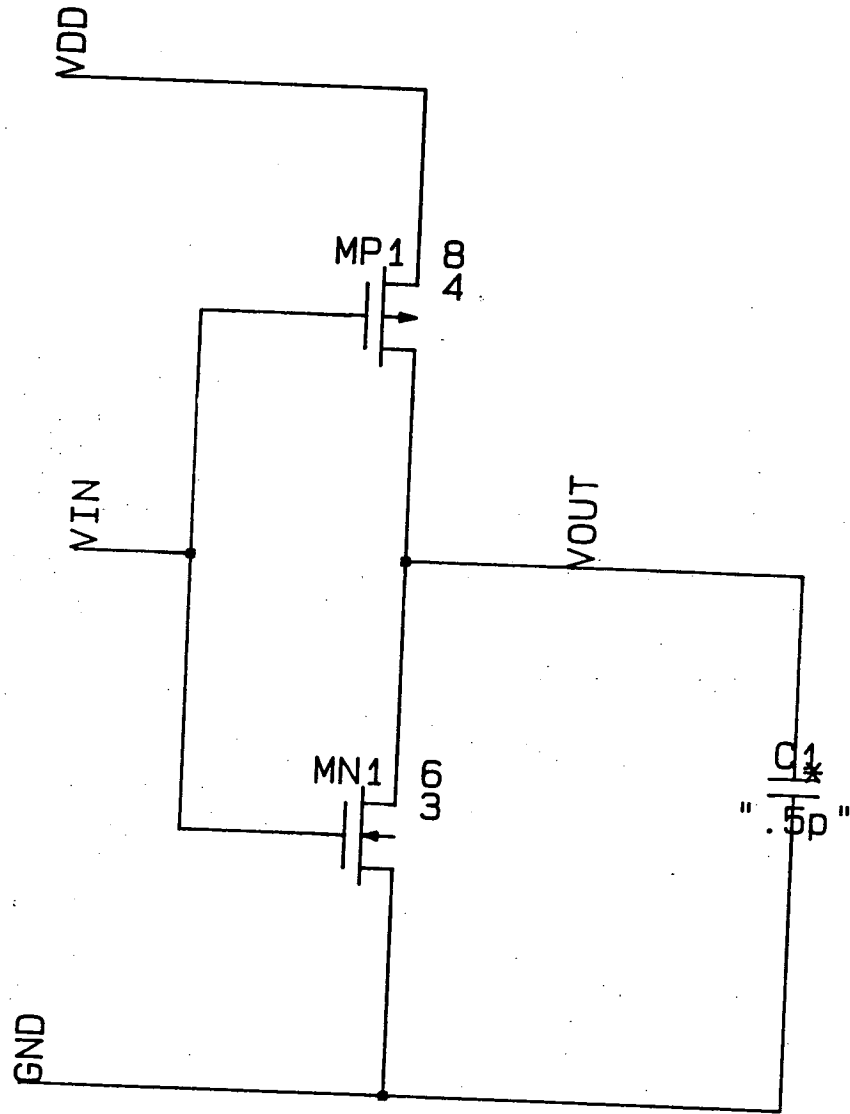
C >= 0.295E-00

D >= 0.280E-00

E < 0.280E-00

Figure 5.2.3.6

Figure 5.2.4.1



CMOS Inverter
with Output Capacitor

Figure	NMOS	PMOS	Predicted Transition
5.2.4.2	realistic best	realistic worst	1.5
5.2.4.3	realistic worst	realistic best	1.7
5.2.4.4	3 σ best	3 σ worst	1.2
5.2.4.5	3 σ worst	3 σ best	2.2

From the figures which plot the variations in V_{IN} and V_{OUT} the predicted switching points using the first order equation 5.2.4.1 agree with the simulation using the level 3 MOS model. In line with the conclusions made by looking at the device characteristics, it is again seen that the 3 σ parameters lead to a much wider spread in inverter transition voltages than is seen by using the drive current deduced parameters. One other aspect of CMOS logic gate operation has been tested: the rise time when driving a fairly large load simulated by a 0.5pF capacitor. The results when using the same combinations of parameters as were used for the inverter switching characteristics are shown in figures 5.2.4.6 to 5.2.4.9. The input voltage fell linearly from 5V to 0V during the test time period 50ns to 100ns. The figure numbers, parameters used and times when 4.5V is reached are

Figure	NMOS	PMOS	Time ($V_{OUT}=4.5V$)
5.2.4.6	realistic best	realistic worst	108 ns
5.2.4.7	realistic worst	realistic best	105 ns
5.2.4.8	3 σ best	3 σ worst	113 ns
5.2.4.9	3 σ worst	3 σ best	98 ns

The total time in figure 5.2.4.6 for V_{OUT} to rise from 0.5V to 4.5V is 20ns and so the 15ns variation in rise time for the 3 σ parameters indicates approximately a 75% variation in rise time. This variation is only 15% for realistic best and worst case parameters.

Figure 5.2.4.2

USER: GRIBBEN

28-June-87 03: 35 P

user\$disk: [gribben.bspace.pwm] INVERT.CKT

ICAP BLOCK: INVERT

VERSION: 2

Temp. = 25 Degrees C

SWITCHING CHARACTERISTIC OF CMOS INVERTER USING REALISTIC BEST CASE NMOS
AND WORST CASE PMOS

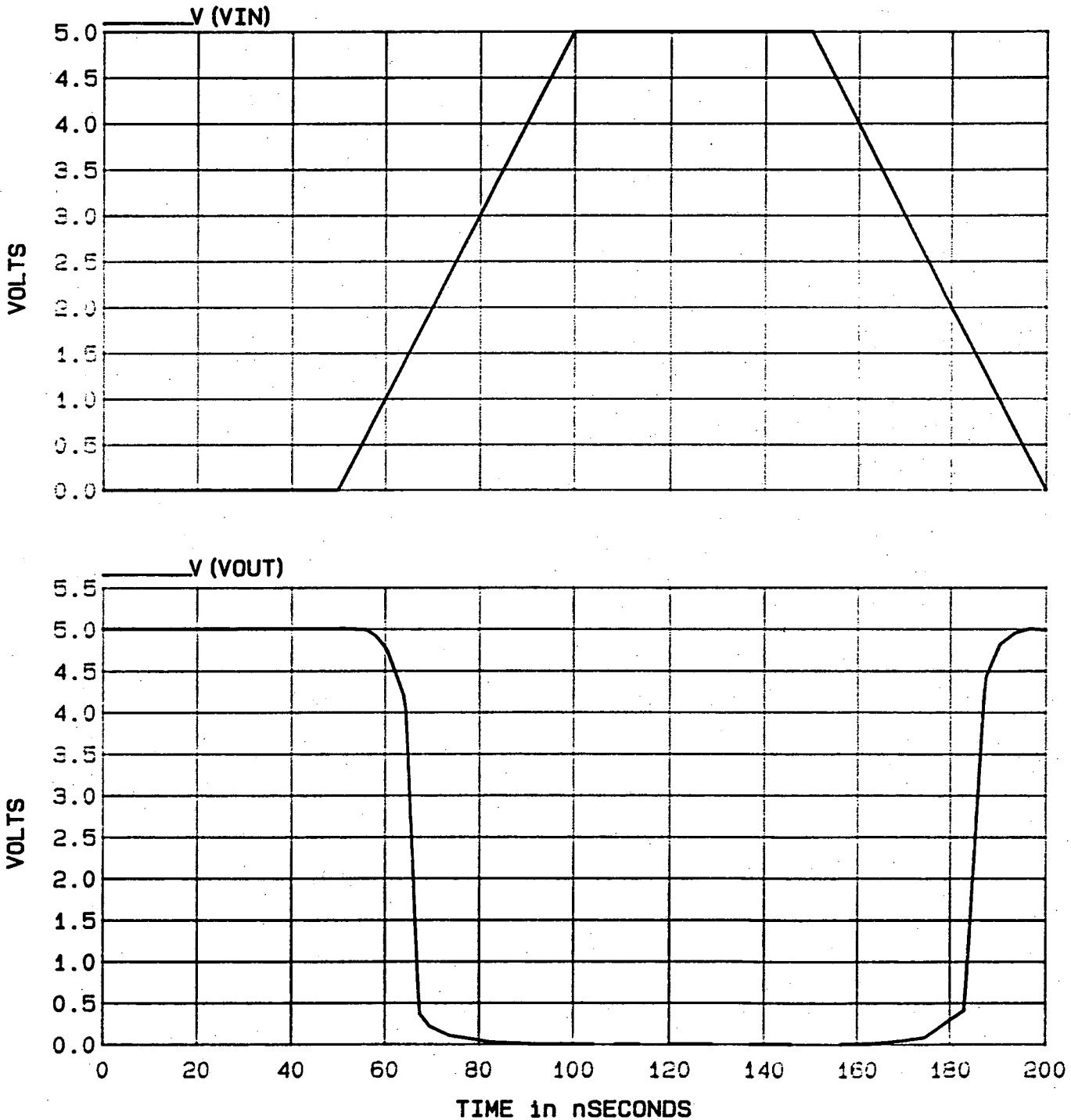


Figure 5.2.4.3

USER: GRIBBEN

26-June-87 03: 12 PM

user\$disk: [gribben.bspace.pwm] INVERT.CKT

ICAP BLOCK: INVERT

VERSION: 2

Temp. = 25 Degrees C

SWITCHING CHARACTERISTIC OF CMOS INVERTER USING REALISTIC WORST CASE NMOS
AND BEST CASE PMOS PARAMETERS

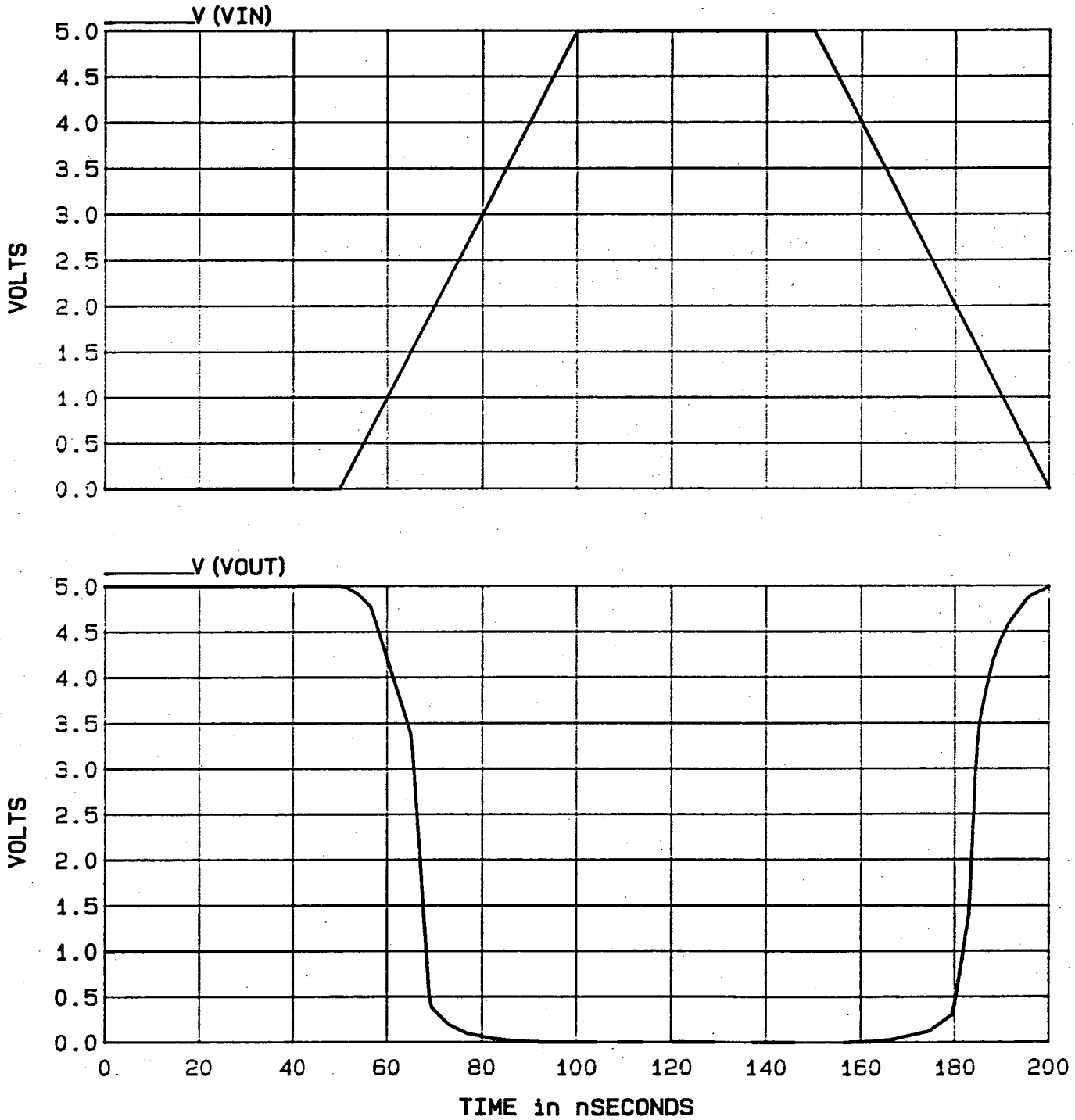


Figure 5.2.4.4

USER: GRIBBEN

28-June-87 03:30

user\$disk: [gribben.bspace.pwm] INVERT.CKT

ICAP BLOCK: INVERT

VERSION: 2

Temp. = 25 Degrees C

SWITCHING CHARACTERISTIC OF CMOS INVERTER USING BEST CASE NMOS (3SIGMA) AND WORST CASE PMOS (3SIGMA)

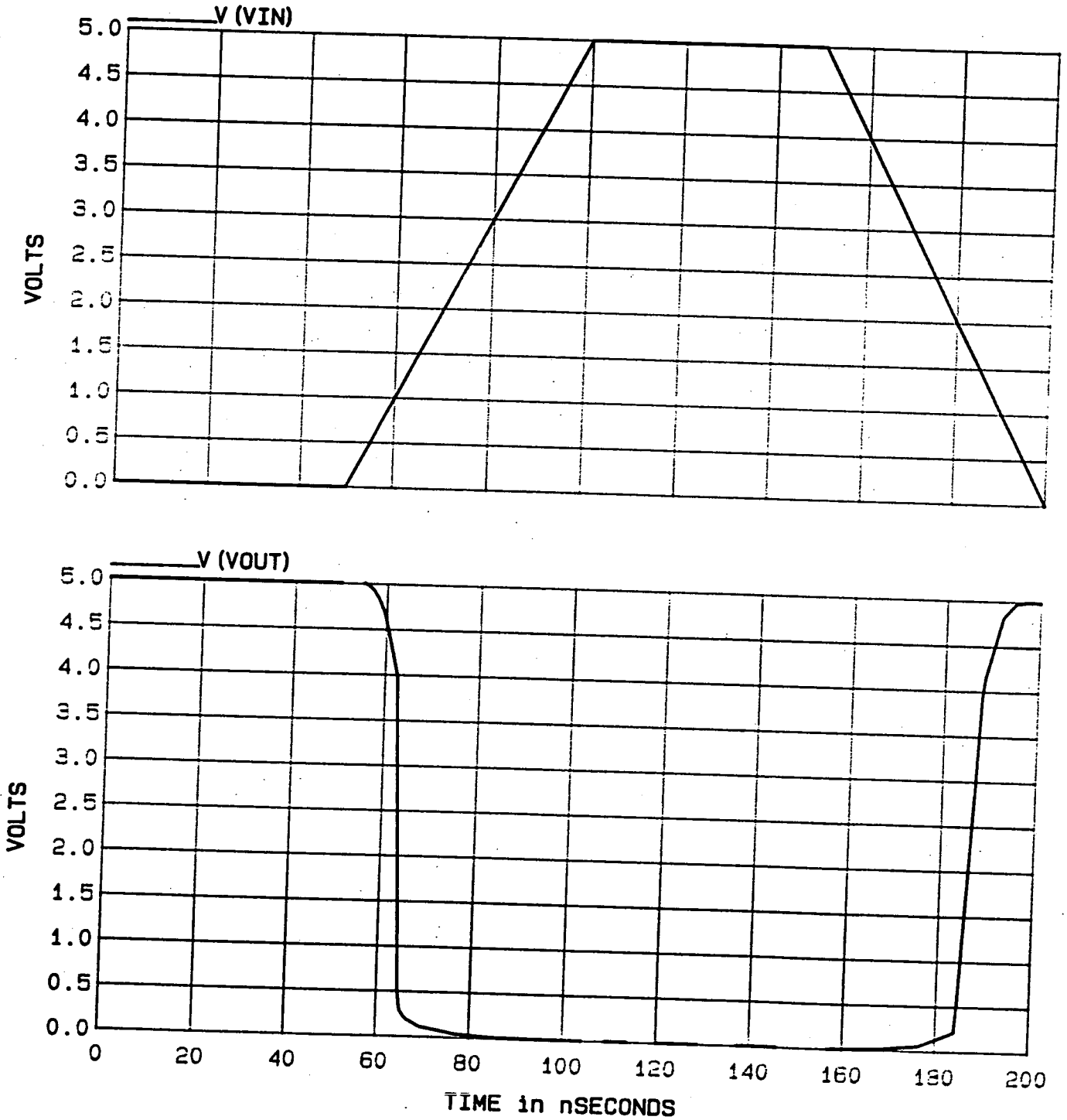


Figure 5.2.4.5

USER: GRIBBEN

28-June-87 01:55 P

user\$disk: [gribben.bspace.pwm] INVERT.CKT

ICAP BLOCK: INVERT

VERSION: 2

Temp. = 25 Degrees C

SWITCHING CHARACTERISTIC OF CMOS INVERTER USING WORST CASE NMOS (3SIGMA)
AND BEST CASE PMOS (3SIGMA)

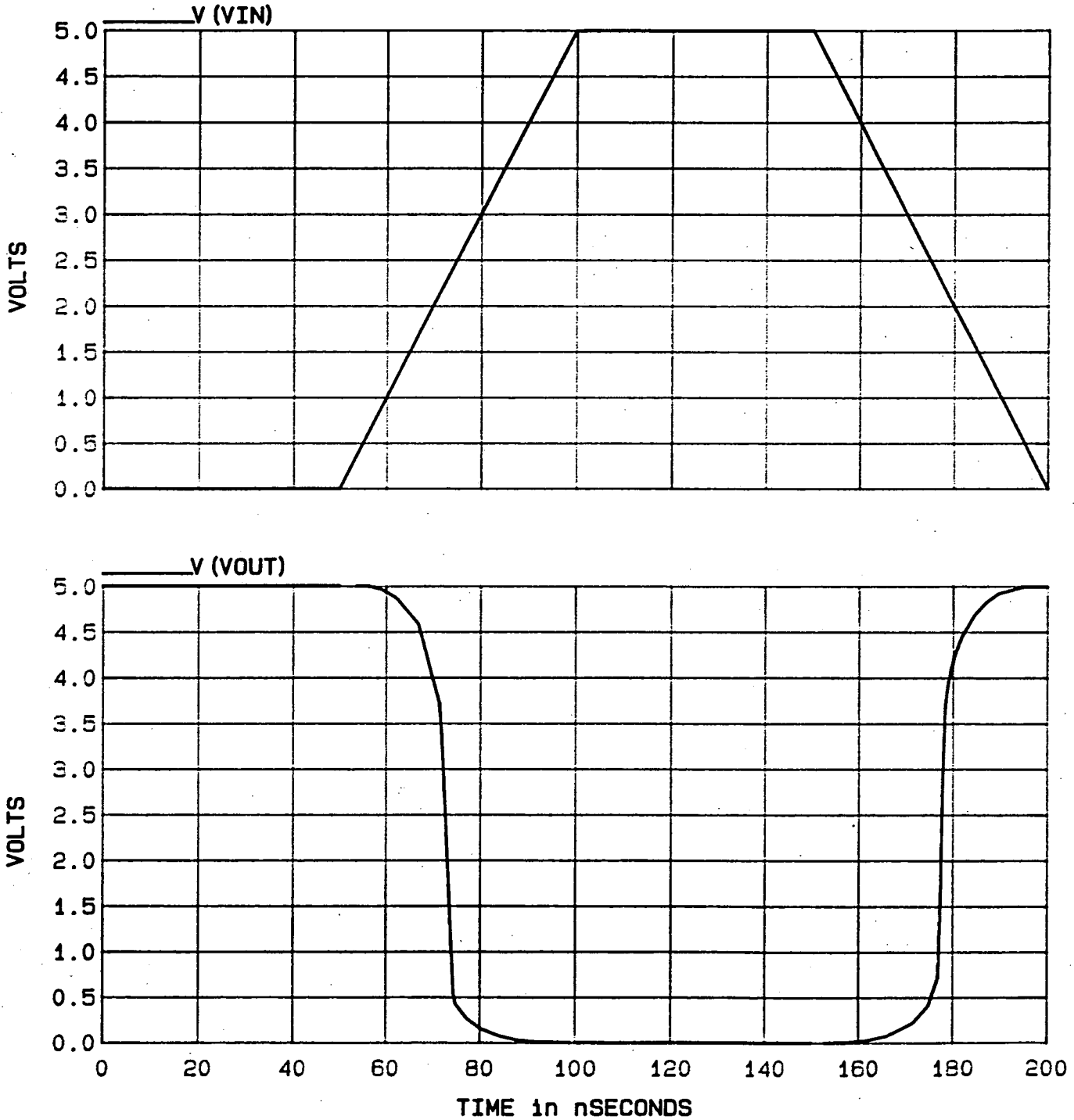


Figure 5.2.4.6

USER: GRIBBEN

28-June-87 04: 21

user\$disk: [gribben.bspace.pwm] INVERT.CKT

ICAP BLOCK: INVERT

VERSION: 2

Temp. = 25 Degrees C

OUTPUT RISE TIME INTO 0.5PF CAPACITOR WITH REALISTIC BEST CASE NMOS AND WORST CASES PMOS PARAMETERS

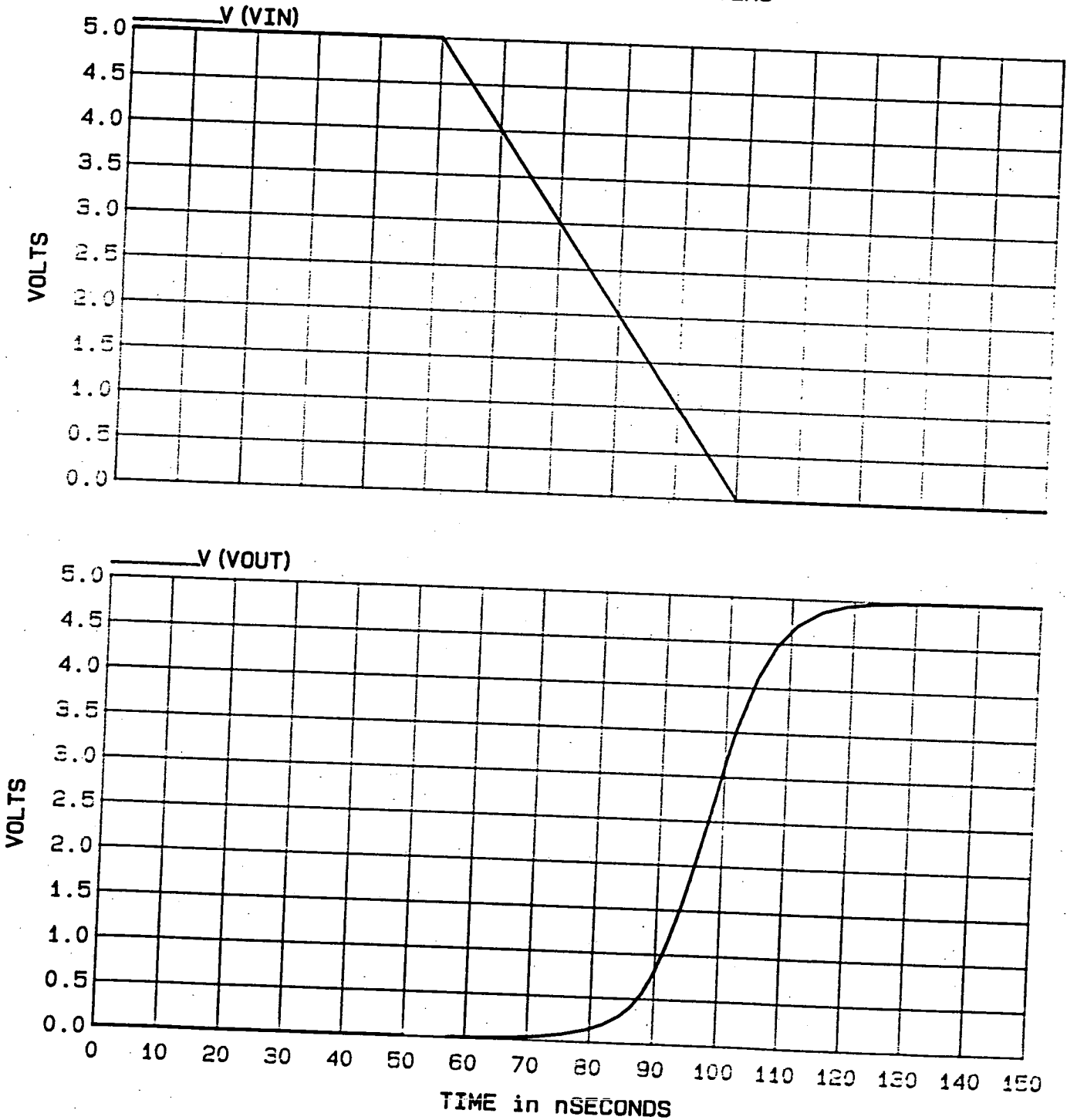


Figure 5.2.4.7

USER: GRIBBEN

26-June-87 04:30 P

user\$disk: [gribben.bspace.pwm] INVERT.CKT

ICAP BLOCK: INVERT

VERSION: 2

Temp. = 25 Degrees C

OUTPUT RISE TIME INTO .5PF CAPACITOR WITH WORST CASE NMOS AND BEST CASE PMOS PARAMETERS

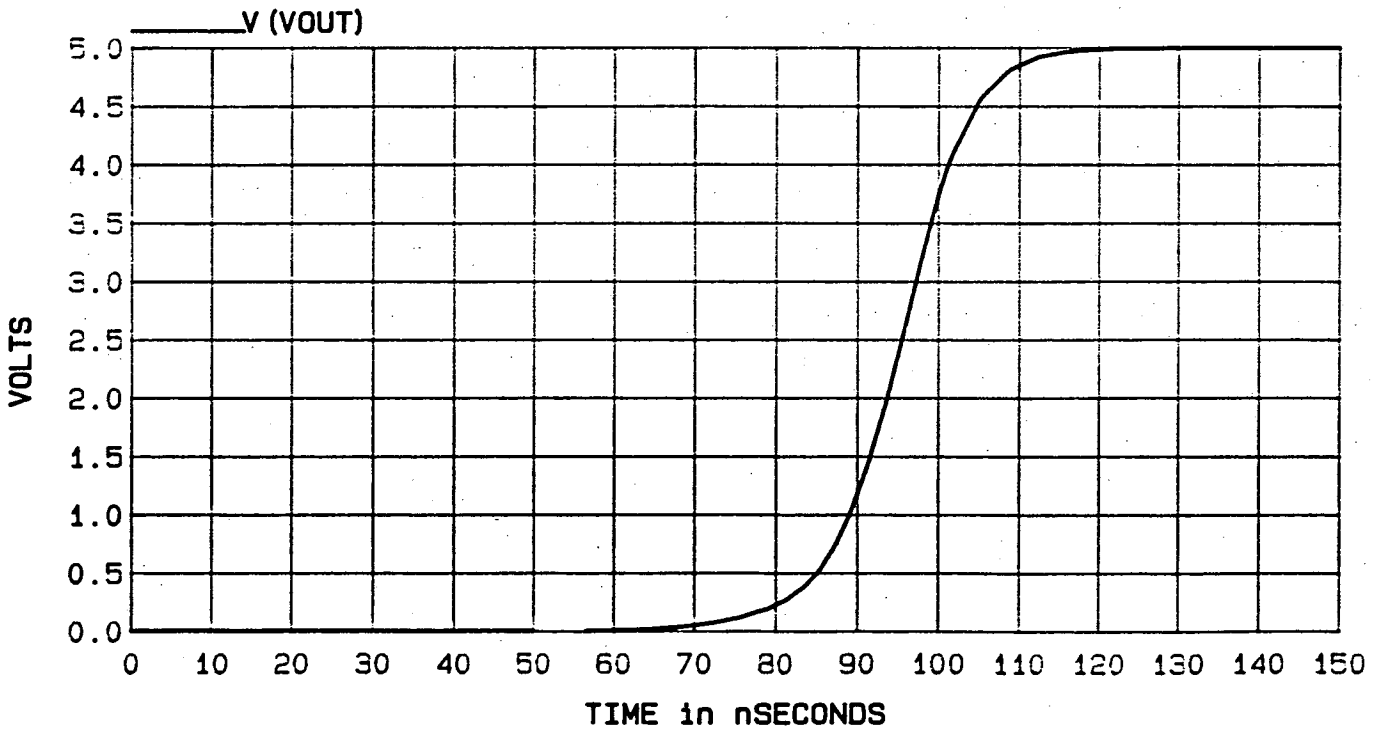
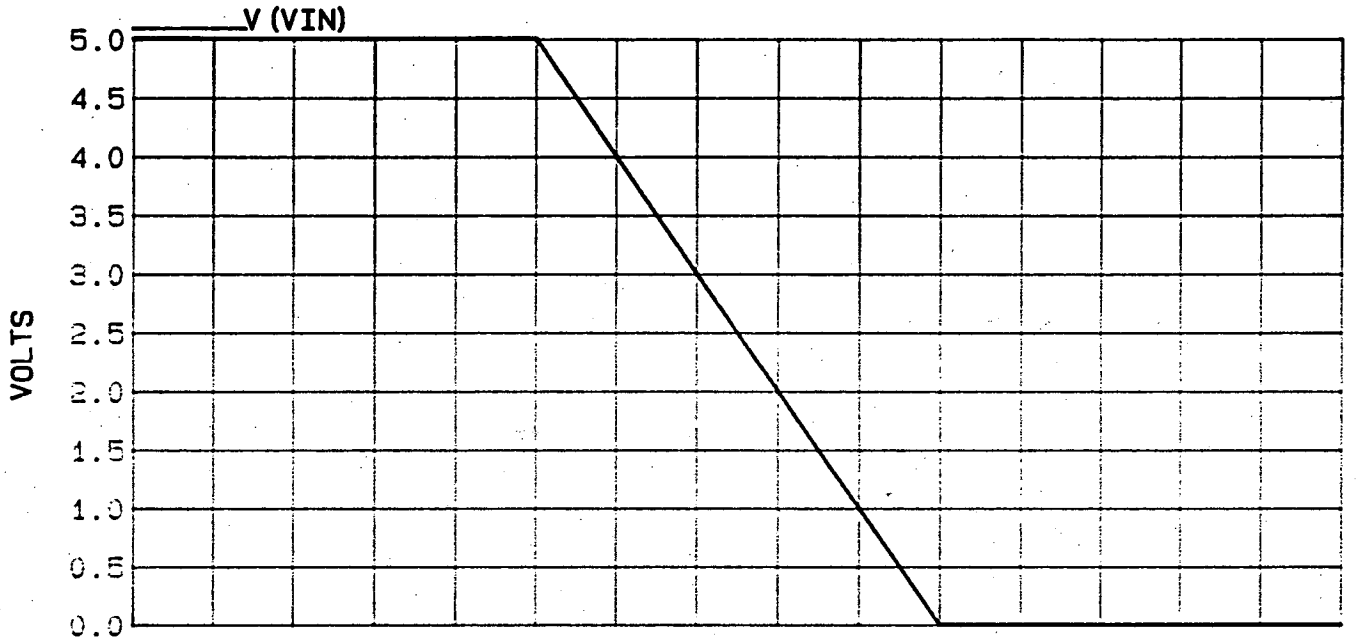


Figure 5.2.4.8

USER: GRIBBEN

26-June-87 05: 25

user\$disk: [gribben.bspace.pwm] INVERT.CKT

ICAP BLOCK: INVERT

VERSION: 2

Temp. = 25 Degrees C

OUTPUT RISE TIME INTO .5PF CAPACITOR WITH BEST CASE (3SIGMA) NMOS AND WORST CASE (3SIGMA) PMOS PARAMETERS

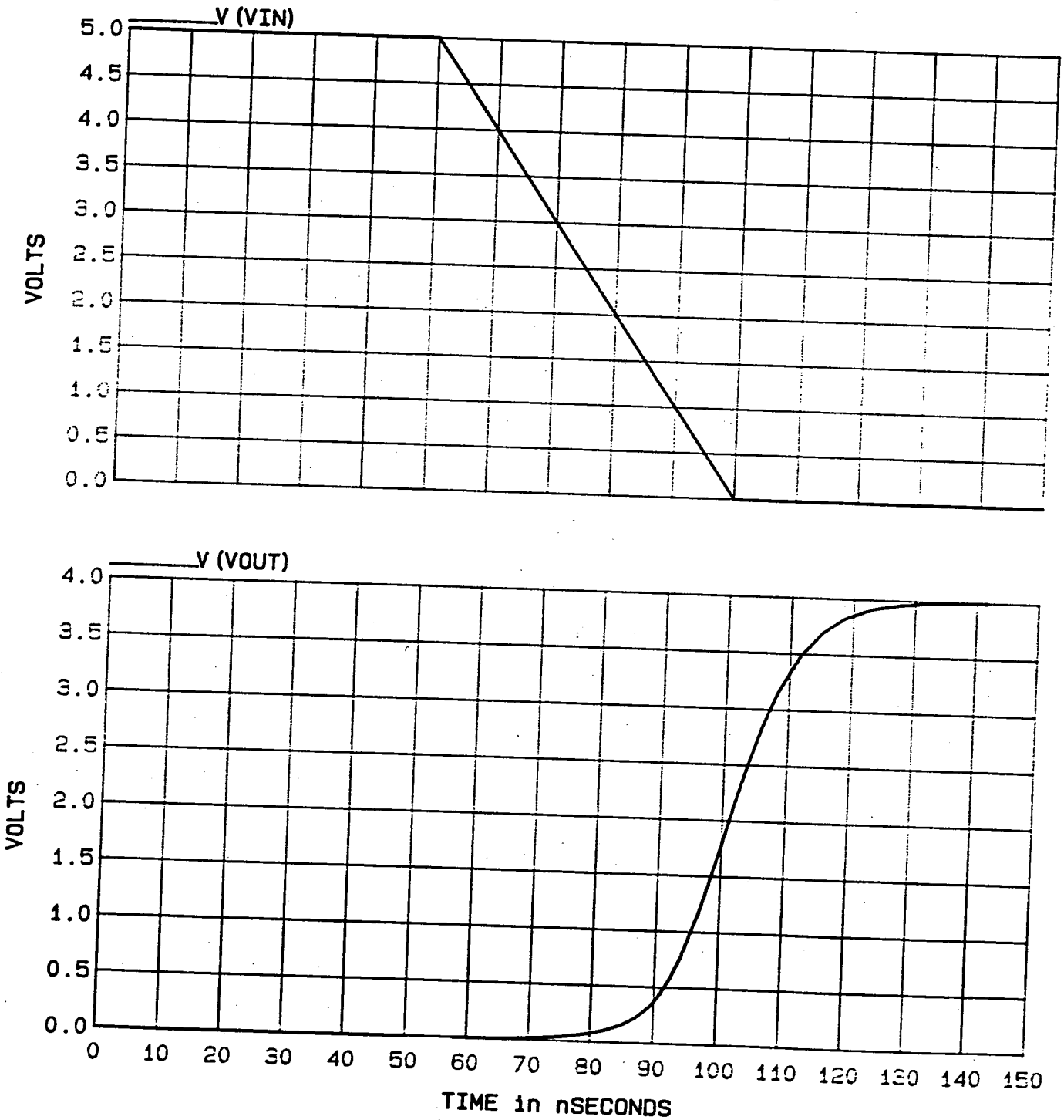


Figure 5.2.4.9

USER: GRIBBEN

26-June-87 05: 20 P

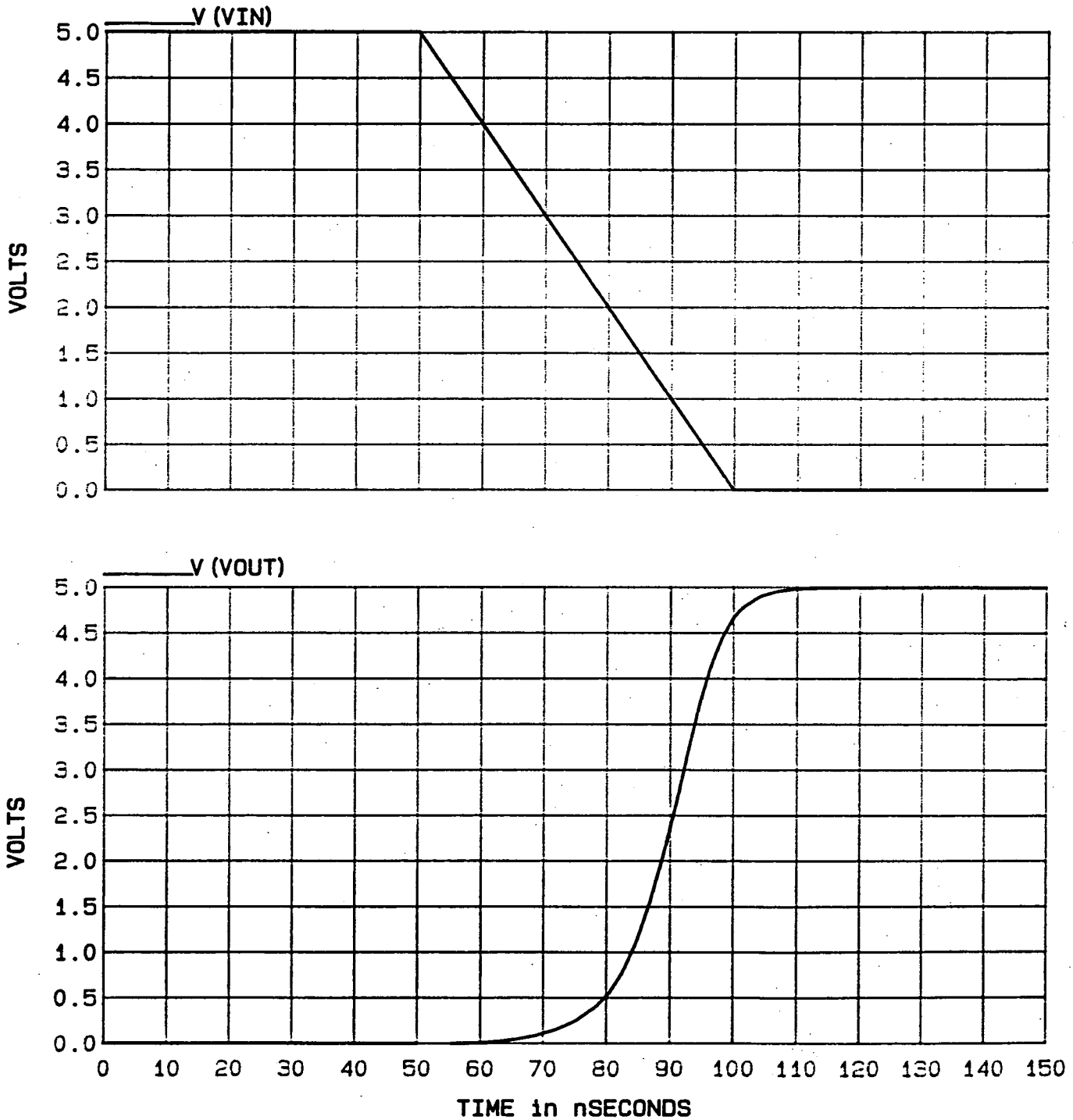
user\$disk: [gribben.bspace.pwm] INVERT.CKT

ICAP BLOCK: INVERT

VERSION: 2

Temp. = 25 Degrees C

OUTPUT RISE TIME INTO .5PF CAPACITOR WITH WORST CASE (3SIGMA) NMOS AND BES
CASE (3SIGMA) PMOS PARAMETERS



5.2.5 Derivation of a Worst Case Parameter Set

In this experiment all parameters were measured at every site on a wafer. Clearly in order to characterise not just a single wafer but a complete process, measurements will have to be carried out using several wafers from numerous batches. This would require a very large resource in terms of time, measurement equipment, extraction and numerical computation. Routine monitoring would also be more difficult than it would be ideally. In an associated study⁶⁷, it was discovered that three simple measurements can give a good indication of device performance. These measurements are I_{drive} , V_t and $Beta_0$. I_{drive} is as defined in figure 5.2.2.1, V_t can be derived from a simple two point measurement in saturation⁶⁸ and $Beta_0$ can be calculated from the first order equation

$$Beta_0 = \frac{I_d}{v_d (V_g - V_t) - \frac{V_d^2}{2}} \quad 5.2.5.1$$

The derivation of a worst case parameter set can then be achieved as follows:

- (i) measure I_{drive} , V_t and $Beta_0$ at all sites.
 - (ii) exclude those sites where I_{drive} is less than $mean-3\sigma$ or greater than $mean+3\sigma$.
 - (iii) find the maximum and minimum I_{drive} from the remaining sites.
 - (iv) check that the $Beta_0$ values for these sites is within $mean\pm 5\%$ and if not choose next worst/best I_{drive} . This is to ensure that the devices are operating in a more or less normal mode (typical devices).
 - (v) extract full parameter sets for these sites.
 - (vi) make a small adjustment to μ_o so that I_{drive} is exactly $\pm 3\sigma$.
- and (vii) adjust V_{to} to $V_t \pm 3\sigma$. This is done because of the special significance of V_{to} in digital circuit operation as seen in Section 5.2.4.

Routine parameter extraction can then be reduced to only involve measurement of I_{drive} , V_t and $Beta_0$. These are parameters which are probably measured at parameter check at the end of fabrication already. In circuit design, speed is usually the most critical aspect of performance and so designers will normally work with the $I_{drive}-3\sigma$ parameter set. When the circuit is complete, the $I_{drive}+3\sigma$ set

could be used to test power consumption.

Worst and best case parameters are more difficult to deduce for analogue circuits. In these circuits, flat saturation characteristics, pair matching and low noise are the important considerations. Of these three, only the saturation slope is easily incorporated in the SPICE simulations.

5.3 Parameters and the Fabrication Process

5.3.1 Introduction

At the end of the fabrication process, parametric test is carried out in order to determine whether the process has been successful. This usually entails measuring quantities such as contact resistance, sheet resistance and linewidth. If these values are within what is often a rather lenient specification, then functional circuits are expected. The only parameters likely to have been measured which provide more information about the precise operation of the circuit (e.g. speed and power dissipation) are gain and threshold voltage. Routine monitoring of SPICE parameters, which is practically feasible using the fast measurement scheme described in Chapter 4, could be used for both process control and to provide toleranced parameters for circuit design. In order to provide process control by SPICE parameter measurement, it is necessary to quantify the relationships between device parameters and process variables.

The ability to predict parameters is still a dream for the future in the silicon fabrication industry. Some of the top circuit design houses in Europe are still coming to terms with the measurement of SPICE parameters and an assessment of the tolerances on these is not available. The fabrication industry is loath to include SPICE parameter measurement in parametric test since they feel it would be time consuming and the parameters would not be as directly relevant to them as what is currently measured. However, as the push continues towards achieving more complex functionality and improved speed/power performance, designers will need to know precisely what the parameters are and what the tolerances on them are.

The link between device parameters and process variables would encourage silicon circuit manufacturers to incorporate SPICE parameter measurement in parametric test. In order to discover exactly what this link is, a knowledge of how all the parameters vary with process changes such as a slightly higher implant dose or a slightly thinner gate oxide is required. This philosophy has implications for processing, design and simulation.

Much tighter process controls will be required in order to accurately predict eventual device parameters. Precise implant doses will have to be measured and

rigorous temperature monitoring must take place in order to regulate the amount of impurity redistribution. This might entail the use of flash annealing techniques used in small geometry processing to produce high impurity gradients. A computerised process management facility will probably be an essential part of such a system. In large scale production, measurement of parameters could be used for process control and the parameters would not only show the effect on circuit operation but would be able to pinpoint where any process variation was taking place.

The design phase too, will be affected by a knowledge of the relationships between process variables and parameters. The process can be chosen to produce the desired circuit operation and so the SPICE parameters become another variable to be defined during the design phase. With an intelligent computerised process monitoring system, an infinite number of process variations would be workable but practically, to avoid confusion which might lead to wasted silicon, only a limited number of variations would probably be allowed.

If SPICE parameters are to become a vital link between process and design, then they should be calculated by process simulators. Currently process simulation packages result in a one-dimensional(1-D) or possibly 2-D device profile and in some cases a value of threshold voltage or diffusion length. Whether they are obtained from an empirical or analytical equation or from a look-up table, the simulator should yield a complete parameter set. The combined use of SUPREM to provide doping profiles, MINIMOS to simulate device characteristics and a parameter extraction package would achieve this goal. Unfortunately this approach is rather time-consuming and awkward.

In this research, an experiment was carried out in order to examine the relationships between process variables and device parameters and to investigate the possibility of applying some empirical equations to these relationships. Two batches of wafers, where each wafer had undergone a different set of changes in fabrication, were measured. The first batch of wafers were CMOS and both NMOS and PMOS parameters were measured at ten sites on all 24 wafers. This resulted in 480 sets of parameters from which the average NMOS and PMOS values were calculated for each wafer. The second batch consisted of 18 NMOS wafers where enhancement parameters were measured at 5 sites and averaged to obtain average parameter sets for each wafer. The results were statistically analysed in order to discover any relationships between

process variables and device parameters.

5.3.2 The Experimental Batches

Two experimental batches of wafers were used in the investigation of parameters and process. The first was a $5\mu\text{m}$ CMOS batch produced by Motorola. The changes are set out in Figure 5.3.2.1. The second set of wafers were the experimental wafers from a development batch of $1.5\mu\text{m}$ NMOS. Various small geometry fabrication techniques were tested and so some wafers did not contain working MOSFETs. The variations in the fabrication process for the working wafers are detailed in Figure 5.3.2.2. These wafers were used in order to look especially for any process-parameter correlations at small geometry.

Three steps were varied in the wafers produced using Motorola's $5\mu\text{m}$ CMOS process: the blanket substrate implant (n-type); the p-type tub implant and the gate oxide thickness. The substrate implant which was varied from $1 \times 10^{12} \text{ cm}^{-2}$ to $3 \times 10^{12} \text{ cm}^{-2}$ and the gate oxide thickness which was varied from approximately 700A to 1000A will affect both p- and n-channel devices. On the other hand the p-tub implant of between $8 \times 10^{12} \text{ cm}^{-2}$ and $12 \times 10^{12} \text{ cm}^{-2}$ should affect only n-channel parameters.

The NMOS process changes need a little more explanation since various experiments were being carried out in order to derive a standard small geometry process. Firstly some wafers underwent the deposition of polysilicon before the growth of the thick oxide. This is the SEPOX process described in Chapter 3 for reducing the bird's beak. Secondly the integrity of the gate oxide has been found to be crucial to producing stable devices and to achieve this a dummy gate oxide is grown and etched before the final oxide. Here a second sacrificial oxide has been used on some wafers. Several different depletion implants were tried in order to obtain the correct threshold voltage in the depletion devices. The gate oxide was 250A and in some cases was grown in steam. Nitridation was used to produce a mixture of silicon nitride and silicon dioxide on some wafers. Finally the shallow and deep boron implants which control the threshold of the enhancement transistors and prevent punchthrough were changed. In particular cases, a heavy dose of shallow implant was used without a deep implant. The high diffusivity of boron was intended to produce the deeper

Figure 5.3.2.1 Process Changes in Batch 1011

Wafer	Substrate Implant cm^{-2}			P-Tub Implant cm^{-2}		Gate Oxide Thickness (Å)			
	1E12	2E12	3E12	8E12	12E12	700	800	900	1000
1	X			X		X			
2	X			X			X		
3	X			X				X	
4	X			X					X
5	X				X	X			
6	X				X		X		
7	X				X			X	
8	X				X				X
9		X		X		X			
10		X		X			X		
11		X		X				X	
12		X		X					X
13		X			X	X			
14		X			X		X		
15		X			X			X	
16		X			X				X
17			X	X		X			
18			X	X			X		
19			X	X				X	
20			X	X					X
21			X		X	X			
22			X		X		X		
23			X		X			X	
24			X		X				X

Figure 5.3.2.2 Implant Changes in Batch 611											
Wafer	Arsenic Implant ($\times 10^{12} \text{ cm}^{-2}$)					Boron Implant ($\times 10^{11} \text{ cm}^{-2}$)					
	3.3	3.5	3.7	4.0	4.3	Shallow			Deep		
						5	6	9	5	10	20
1			X			X					X
2	X						X		X		
3			X			X					X
4	X						X		X		
5					X	X					X
6			X				X		X		
8			X				X		X		
11					X		X			X	
12	X							X		NONE	
13				X		X					X
14		X					X		X		
15				X		X					X
17		X						X		NONE	
19		X						X		NONE	
20				X			X			X	
21			X				X			X	
23			X				X			X	
24			X					X		NONE	

Wafers undergoing polysilicon deposition before LOCOS:

13,14,15,17,19 and 20

Wafers receiving an extra sacrificial gate oxide:

11,12,13,14,15,17,19,20,21,23 and 24

Wafers on which dry gate oxidation was used:

5,6,8,11,12,13,14 and 15

Wafers where nitridation took place after gate oxidation:

3,4,5,6,11,12,13,14,19,20 and 21

concentration necessary to avoid punchthrough.

5.3.3 Process-Parameter Relationships

Using a fully automated version of the parameter extraction program, PARAMEX, PMOS and NMOS parameters were measured at ten sites on each of the CMOS wafers. The sites were randomly selected from different areas of the wafer. The resulting values were averaged to provide the parameters for each wafer. The process variables: oxide thickness, n-substrate implant dose and p-tub implant dose were entered into the database and a statistical analysis was performed to determine the correlation between these variables and the SPICE parameters. The correlation statistics are listed in figures 5.3.3.1 and 5.3.3.2 for NMOS and PMOS respectively.

Most parameters are influenced by the gate oxide thickness. For both NMOS and PMOS, threshold voltage V_{to} and substrate bias coefficient γ correlate significantly with oxide thickness. Threshold voltage V_{to} shows a correlation of 0.44 and -0.70 for NMOS and PMOS respectively. The minus sign arises due to the fact that PMOS thresholds are negative. γ and t_{ox} correlate with values of 0.82 and 0.74 in NMOS and PMOS respectively. Both of these results are expected since the thicker gate oxide diminishes the control of the gate voltage. A higher gate voltage is required to produce the surface potential needed to turn the device on and also to overcome the influence of substrate bias on the surface potential.

Figure 5.3.3.3 illustrates the values of PMOS γ for each wafer and the correlation between oxide thickness and γ can be seen. Referring to figure 5.3.2.1, the values of PMOS γ for each wafer and the correlation between oxide thickness and γ can be seen. The first wafer has a gate oxide of around 700A, the second 800A, the third 900A and the fourth 1000A. The oxide thickness variation is repeated in the same way: 700A, 800A, 900A and 1000A on wafers 5 through to 24. Cosmetic lines have been drawn in figure 5.3.3.3 to highlight the fact that for each set of four wafers, γ increases as oxide thickness increases.

The mobility parameters: μ_o , θ and v_{max} are all dependent upon oxide thickness. The low field mobility correlates positively with t_{ox} in both NMOS and PMOS. The rate of mobility reduction as gate voltage increases is significantly lower for

Parameter	Oxide Thickness	Substrate Implant	P-Tub Implant
V_{to}	0.44	-0.61	0.64
γ	0.82	-0.32	0.36
μ_o	0.55	0.26	-0.18
θ	-0.86	0.11	-0.17
v_{max}	0.49	0.32	-0.53
N_{fs}	-0.50	0.24	0.13
L_d	-0.15	0.28	-0.64
Δ_w	0.33	0.32	-
δ	-0.46	-0.36	-0.15
η	-	0.51	-0.22
κ	0.36	-0.45	0.73

Parameter	Oxide Thickness	Substrate Implant	P-Tub Implant
V_{to}	-0.70	-0.70	-
γ	0.74	0.66	-
μ_o	0.38	0.27	0.28
θ	-0.89	-0.33	-
v_{max}	-0.25	-	0.24
N_{fs}	-0.80	0.27	0.11
L_d	-0.15	-0.66	-
Δ_w	0.27	0.32	0.28
δ	-0.37	-	-0.28
η	-0.27	-0.73	-
κ	0.26	0.67	0.24

Parameters V Process

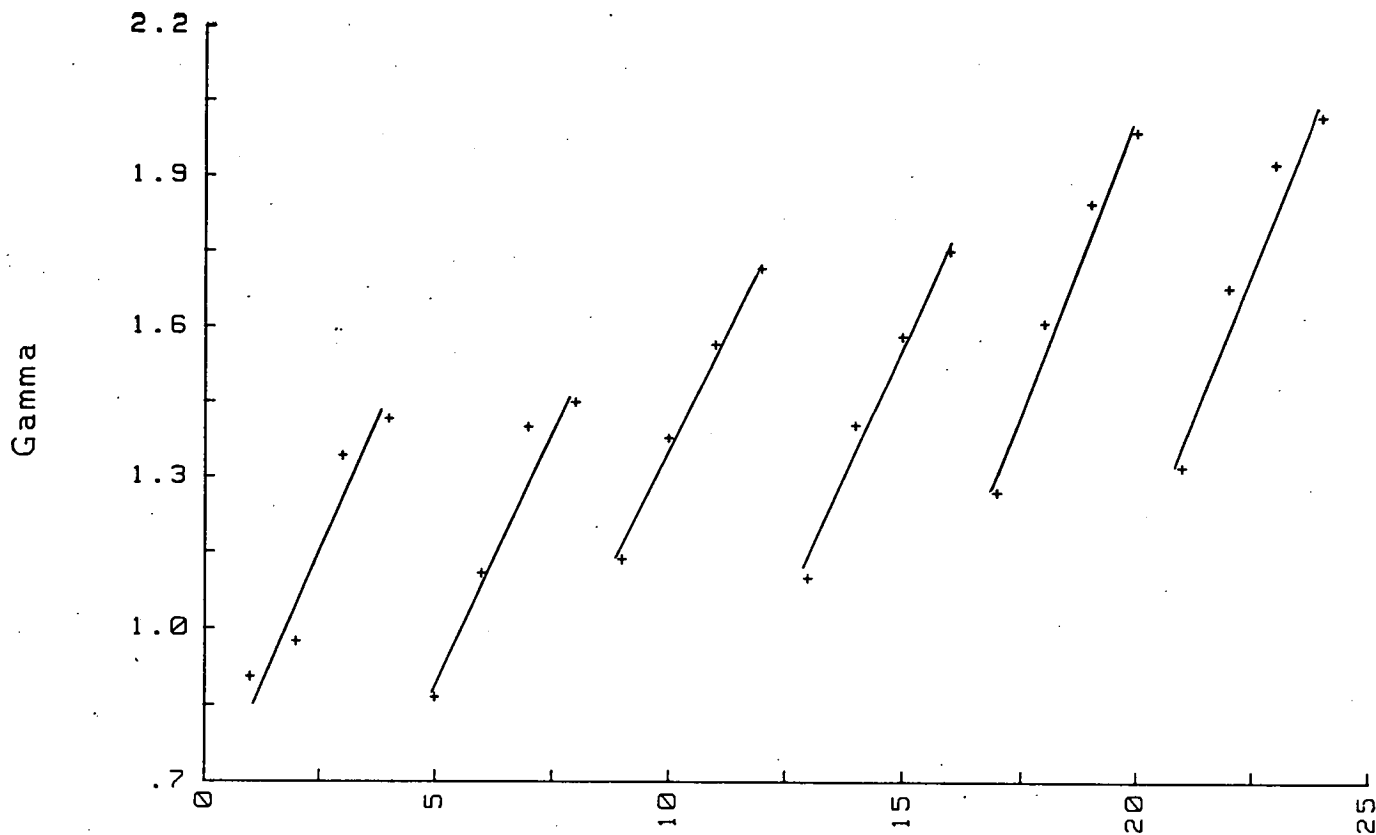


Figure 5.3.3.3

Wafer_No

PMOS

Parameters V Process

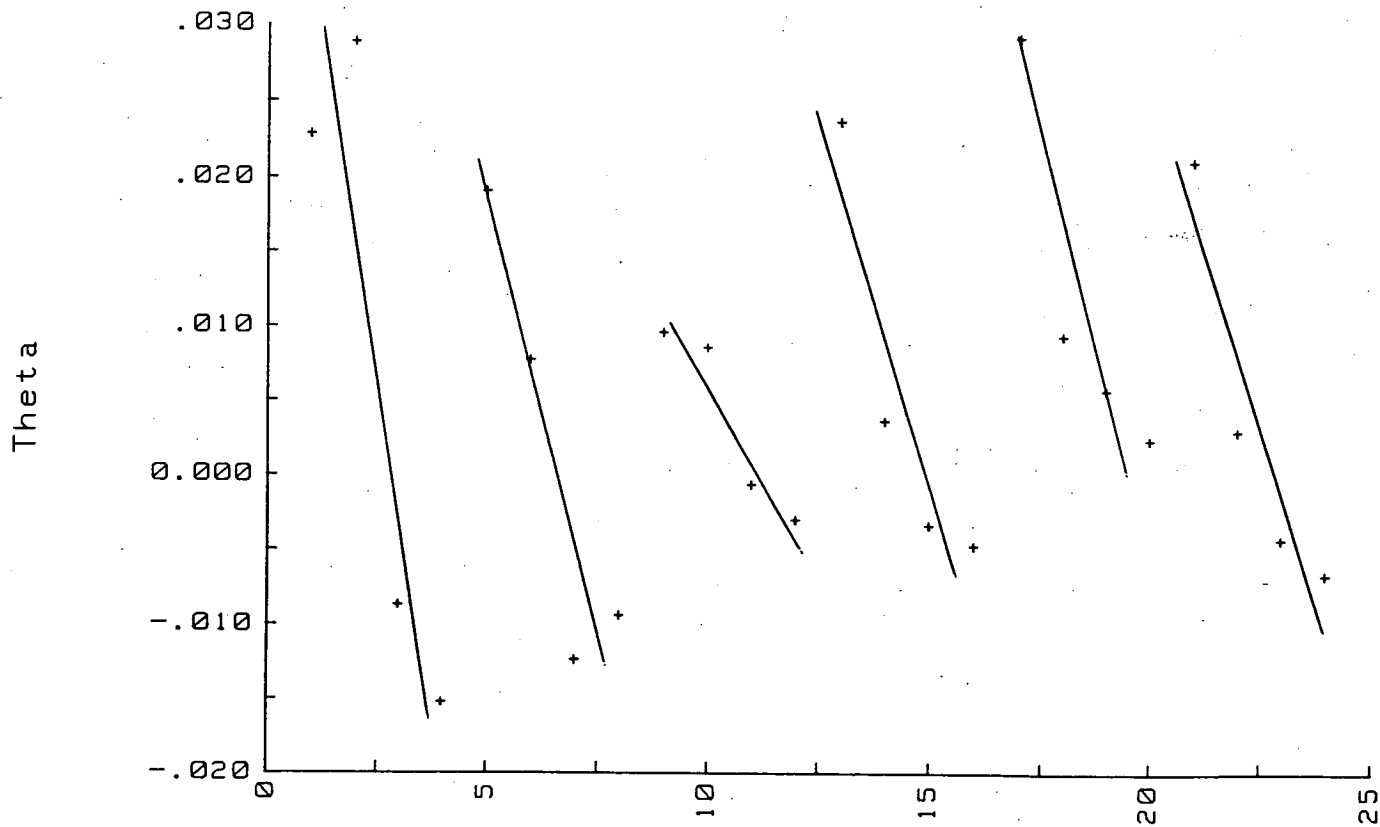


Figure 5.3.3.4

Wafer_No

NMOS

thick gate oxides and correspondingly there is a high negative correlation. A greater increase in gate voltage is required to produce a similar increase in transverse electric field at the silicon-silicon dioxide interface. In NMOS the correlation is -0.86 (see figure 5.3.3.4) and in PMOS, it is -0.89. Cosmetic lines have once again been included to emphasise the correlation. Drain voltage modulation of carrier mobility is also reduced and hence v_{\max} shows a positive correlation with t_{ox} for NMOS. v_{\max} is not a significant parameter in PMOS since p-type carrier mobility is low and little velocity saturation occurs.

The final parameter which negatively correlates with t_{ox} is N_{fs} . In NMOS the correlation is -0.50 and in PMOS it is -0.80. This indicates that the rate of turn-off below threshold is greater for thicker oxides. Theoretically this is not the case²¹ (see Chapter 6). In fact the change in subthreshold slope is only slightly changed by the small changes which have been made in the oxide thickness. The slope is affected much more by the changes made to the impurity profile and consequently the correlation figures between N_{fs} and t_{ox} lose their significance.

The second process variable was the dose of the blanket n-substrate implant used to produce higher surface concentrations for PMOS transistor fabrication without increasing the bulk conductivity. The NMOS channel regions also receive this implant. The higher concentration of n-type impurity raises the p-type thresholds (they become more negative and hence the correlation is in fact negative -0.70) and lowers the n-type thresholds whose correlation with implant dose is -0.61. The body effect coefficients show negative and positive correlation for NMOS and PMOS respectively. The reduction of net p-type impurity in the NMOS channels with higher substrate dose means that the depletion region around the drain extends further into the channel for the same bias. Consequently the static feedback coefficient correlates positively for NMOS 0.51 and negatively for PMOS, -0.73. This is illustrated in figure 5.3.3.5 where PMOS η is plotted against wafer number. The horizontal lines are intended to show the approximate average η values for wafers 1 to 8 where the dose was $1 \times 10^{12} \text{ cm}^{-2}$; for wafers 9 to 16 where the dose was $2 \times 10^{12} \text{ cm}^{-2}$ and thirdly for wafers 17 to 24 where the dose was $3 \times 10^{12} \text{ cm}^{-2}$. This is a second order parameter which only becomes important in short channels but is seen to have a definite dependence upon the channel impurity concentration.

Another small geometry effect, represented by the saturation slope coefficient κ correlates with the substrate dose. In NMOS the factor is -0.45 and in PMOS the factor is 0.67. Again the effect is dependent upon the depletion region around the drain which in turn is dependent upon the channel impurity concentration. In NMOS the net channel impurity concentration decreases and hence the saturation slope goes down (negative correlation) whereas in PMOS the net concentration increases and the saturation slope goes up.

The final process step to be altered was the implant dose for the p-tub in which the NMOS transistors are made. This implant should only affect the NMOS devices and as might be expected no significant correlation was found between the PMOS parameters and the implant dose. In NMOS both V_{to} and γ correlate with coefficients of 0.64 and 0.36 respectively. The higher channel doping causes V_{to} to rise and a higher gate voltage is required to overcome the potential applied to the substrate. The quantity of mobility modulation by drain voltage increases as the doping increases because of the greater number of imperfections in the crystal lattice. Consequently v_{max} is reduced and correlates with the tub implant with a value of -0.53. Finally the saturation slope coefficient κ correlates positively with the tub implant dose with a value of 0.73. The positive correlation with p-tub implant dose results for the opposite reasons than for the negative correlation with the n-substrate implant which was explained above. The graph of κ against wafer number is shown in figure 5.3.3.6. Referring to figure 5.3.2.1, it can be seen that κ increases with oxide thickness, decreases with substrate implant and increases with p-tub implant.

As was explained in Section 5.2, the batch of 1.5 μ m NMOS was fabricated in order to determine the most suitable combination of processing steps to produce 1.5 μ m NMOS. No structured variations took place and so the conclusions which have been drawn are deduced from only a few wafers on which a particular step was varied. The large number of process variables also confuses the correlation statistics. For this reason only the most significant correlations are worthy of discussion. The correlation statistics for this batch are listed in figure 5.3.3.7.

The main objective of the SEPOX process (explained in Chapter 3) was to reduce the bird's beak and this was proved by the correlation of Δ_w with the use of SEPOX. Both gate voltage and drain voltage modulation of mobility increase as the

Parameters V Process

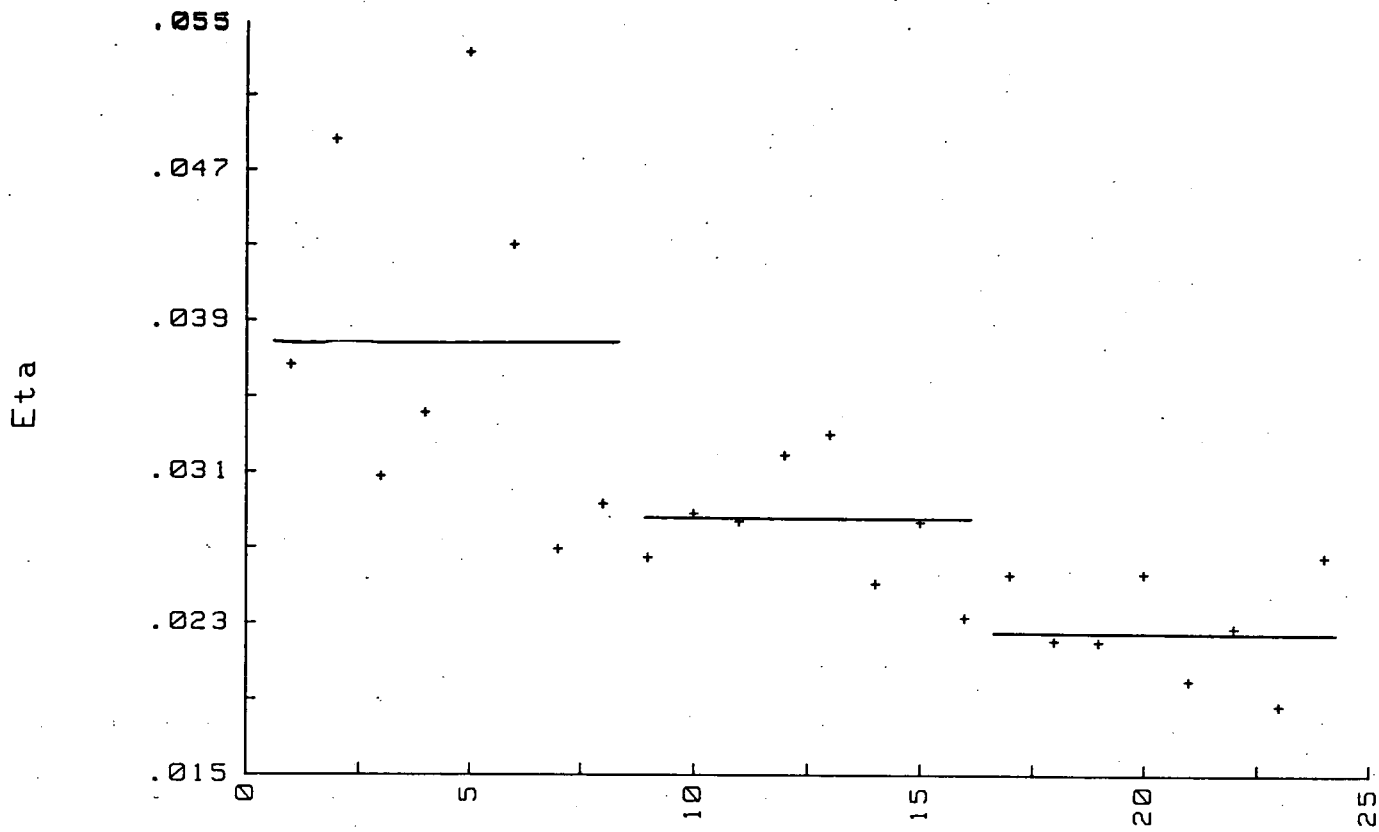


Figure 5.3.3.5

Wafer_No

PMOS

Parameters V Process

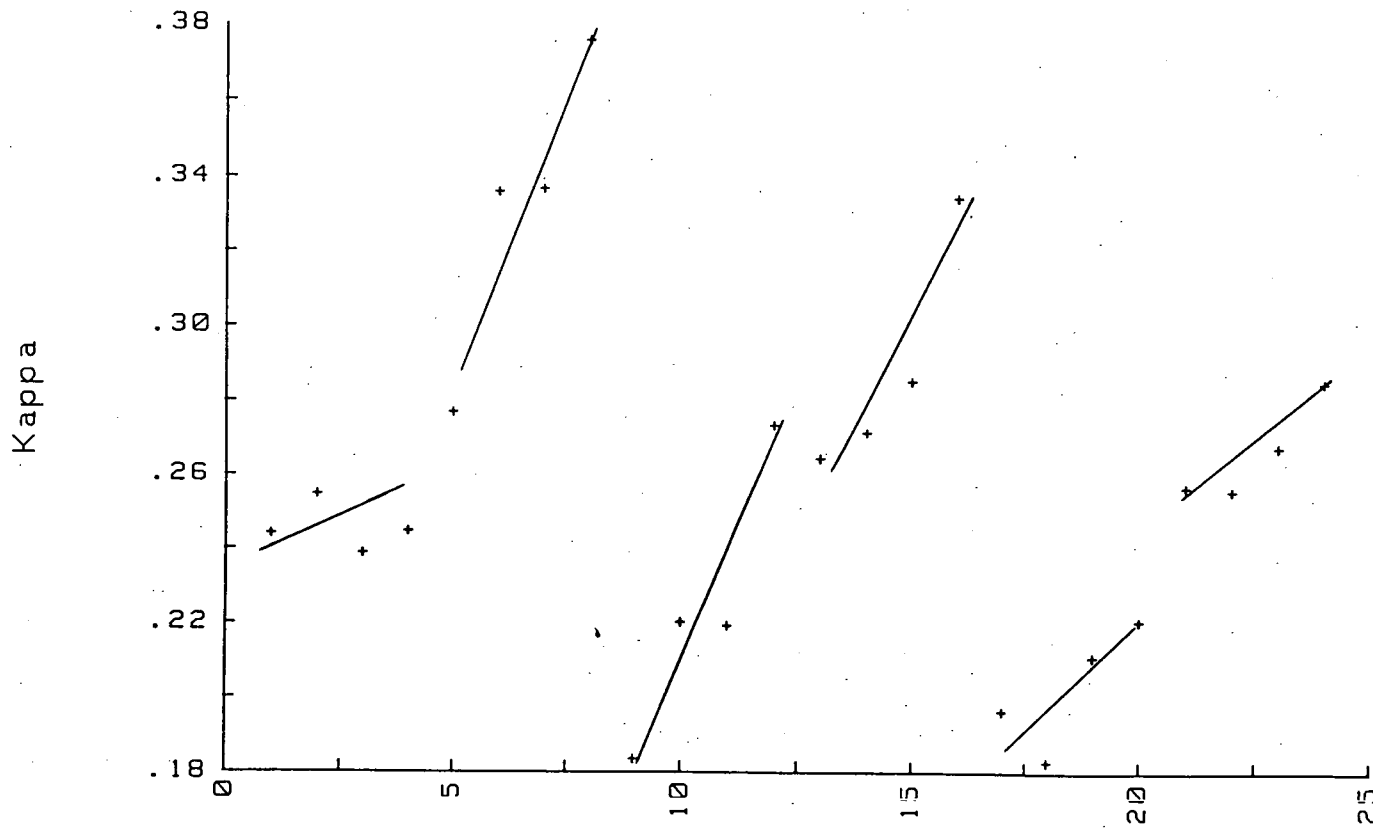


Figure 5.3.3.6

Wafer_No

NMOS

Figure 5.3.3.7 Small Geometry NMOS Parameter-Process Correlations

Parameter	Poly Before LOCOS	Extra Dummy Oxide	As Implant	Dry Oxide	Nitrid- ation	Deep B Implant	Shallow B Implant
V_{to}	-0.30	-0.37	0.22	0.51	-0.25	0.27	-0.11
γ	-0.14	-0.43	0.61	0.37	-	0.91	-0.91
μ_o	-0.39	-0.82	-0.13	-	-0.24	-	-0.28
θ	0.91	0.70	-	-0.13	-	-0.13	0.29
v_{max}	-0.71	-0.77	-0.18	-0.18	-0.21	-	-0.14
N_{fs}	-0.54	-0.37	-0.61	-0.17	-0.17	-0.26	0.16
L_d	-0.48	-0.27	-0.25	-0.64	-0.44	-0.25	0.23
Δ_w	-0.88	-0.74	-	-	-0.24	0.14	-0.31
δ	0.67	0.64	-0.17	-	0.12	-0.44	0.56
η	0.35	-	-0.37	-0.35	-	-0.46	0.33
κ	0.21	-0.56	0.49	0.54	-	0.74	-0.77

width of the channel increases. The gate voltage effect is probably a consequence of the broader channel region over which the transverse electric field can act. The extra sacrificial oxide leads to lower maximum mobilities and more drain modulation of mobility. The high correlation of the device width with extra sacrificial oxide is due to the fact that the wafers which had the extra oxide also underwent SEPOX.

The arsenic implant, the dry/wet oxide growth and nitridation do not correlate significantly with any parameters. The arsenic implant variations only affected the depletion devices and no parameters were extracted for these structures.

The effect of the boron implants is also difficult to determine as the significant correlations seem to be contradictory. The values of γ correlate with the shallow and deep boron implants with values of -0.91 and 0.91 respectively. This arises because wafers which received a light shallow implant received a heavy deep implant and vice versa. Theoretically the major positive correlation should be between the deep implant and gamma and this has been found in practice.

Similarly κ correlates with the shallow and deep boron implants with values of -0.77 and 0.74 respectively. The slope coefficient κ is more dependent on the shallow implant dose i.e. the surface concentration and a light surface concentration results in a high κ value.

5.3.4 Conclusions and Future Experiments

From these experiments, it has been shown that the process variations: oxide thickness and implant doses correlate significantly with device parameters. The oxide thickness relates particularly to θ and N_{fs} , which are dependent the variation in surface field and surface potential with gate potential. The implant doses have more effect on V_{to} , γ and the second order parameters η and κ . The width reduction Δ_w shows the success of the SEPOX process in reducing the bird's beak width.

The high correlations indicate that it might be possible to establish some empirical equations between process variables and SPICE parameters. As an example, suppose V_{to} is to be predicted. The SPICE equation from Chapter 2 is

$$V_{to} = V_{fb} + 2\phi_b + \gamma(2\phi_b)^{\frac{1}{2}} \quad 5.3.4.1$$

If there are two implants of opposite types in the channel then a possible form of the empirical relationship would be

$$V_{to} = a_0 + a_1 \ln[\text{implantA}(\text{dose}, \text{energy})] - a_2 \ln[\text{implantB}(\text{dose}, \text{energy})] + a_3 \left\{ \text{implantA}(\text{dose}, \text{energy}) - \text{implantB}(\text{dose}, \text{energy}) \right\}^{\frac{1}{2}} t_{ox} X \left\{ \ln \left(\text{implantA}(\text{dose}, \text{energy}) - \text{implantB}(\text{dose}, \text{energy}) \right) \right\}^{\frac{1}{2}} \quad 5.3.4.2$$

The (dose,energy) factors for each implant are functions which provide the average doping concentration from the dose and energy variables.

If routine measurement of SPICE parameters takes place, using the minimum parallel test system described in Chapter 4, then through the sort of relationship outlined above, the parameters can provide a very useful extra indicator of process control. With their precise relevance in circuit design, these values can provide more information than the parameters which are currently monitored. The processing industry will probably continue to insist on monitoring contact resistances, sheet resistances and step coverage since they can be related directly to particular stages of processing. However, it may be that these are only measured on a more occasional basis as SPICE parameters are measured currently.

In order to establish an empirical relationship for a particular process, more experimental batches of wafers with carefully planned variations need to be fabricated. It would probably be useful to only vary one quantity at a time. This would not only quantify the relationship between particular process variables and SPICE parameters, but also impose realistic tolerances on various steps in the process. Through device simulation, limits on device parameters which produce functional circuits can be set and then through the empirical equations, tolerances can be set on implant dose or gate length. Referring to equation 5.3.4.2, if it is discovered that the gate oxide thickness is too small then it may be possible to increase the channel implant to produce V_{to} s within specification.

Process simulation tools still need to be further developed in order to yield device parameters. The cumbersome method for this described in Section 5.3.1 using SUPREM and MINIMOS has recently been improved in a suite of software packages

supplied to the EMF by Technology Modelling Associates Inc (TMA). The TMA software includes process and device simulators as well as parameter extraction software and data can be easily transported between them. Parameters measured from fabricated wafers should be compared with those from simulation in order to validate the simulation tools. This is an essential component in parameter based design and wafer fabrication.

Chapter 6 : Subthreshold Operation

6.1 The Subthreshold Region

The subthreshold region of MOS device operation is receiving renewed attention with the advent of smaller geometry and possibly lower voltage digital circuits and the promise of its incorporation in analogue circuits. In the past, MOS transistors have been used mainly in digital circuits with geometries of $2\mu\text{m}$ and greater and a supply voltage of 5V. Unless subthreshold currents are exceptionally large, then the capacitive nodes are sufficiently large that the leakage current is of little significance. Analogue MOS circuits have tended to use devices biased in their linear or saturation regions.

Currently, MOS dimensions are reaching $1\mu\text{m}$ or less and moves have been made to reduce the standard supply voltage, probably to 3.3V.¹⁸ Under these circumstances, capacitive nodes are much smaller and when used with lower supply voltage, they can discharge much more easily. Therefore, the importance of subthreshold currents is greatly increased. In analogue circuits too, noise and gain advantages can be obtained by operating in the subthreshold region.^{69,19} CMOS operational amplifiers designed to operate in subthreshold are very suitable for switched capacitor circuits.

As a consequence of the greater importance of the subthreshold region, the factors which affect the subthreshold device characteristics are of increasing interest. In particular, how sensitive is the subthreshold slope to the impurity profile? As MOS transistor design developed, a channel implant was introduced to adjust the threshold voltage and as channel lengths were reduced, punchthrough was avoided by using a deep channel implant. The primary objectives of these implants were to tailor threshold voltage and prevent punchthrough and their effect on the subthreshold region was largely ignored. As long as the swing was approximately of the order of 100mV/dec then it was considered satisfactory.

The lack of attention received by the subthreshold region is evident in the SPICE MOS models. Level 1 assumes zero current and level 3 uses an exponential

dependence of current upon gate voltage, the slope of which is governed by the parameter N_f . The theory behind this is provided in Section 6.2. According to SPICE, the subthreshold slope is independent of substrate bias.

Other authors, Shannon⁷⁰, Brews²¹ and Buehler^{71,72} have investigated techniques for derivation of doping profiles from subthreshold measurements. Doping profiles can theoretically be obtained from subthreshold swing, S versus V_b curves. As the substrate bias is increased, the depletion region is extended. The distance by which it increases is dependent upon the doping concentration in the region through which the depletion region spreads. Hence the depletion capacitance which is dependent upon the depletion width is in turn dependent on the doping profile. As will be seen in Section 6.2, S is explicitly dependent on the depletion capacitance and so the variation of S with V_b is profile dependent.

Following on from the argument above, some experiments have been carried out to test the sensitivity of the subthreshold slope to the channel profile. The profile, if it is desired, can be experimentally determined using C-V techniques. Here SUPREM has been used to simulate the channel profiles of several devices which received different implants. The devices were then fabricated and the S against V_b characteristics measured.

6.2 Subthreshold Model for Uniformly Doped Devices

Conduction in the subthreshold region is due to diffusion of the minority carriers between source and drain. The model is formed therefore, by treating the MOS transistor in subthreshold as a weak bipolar device.

$$\begin{aligned} I_d &= -q \text{Area} D_n \frac{dn}{dy} \\ &= -q W x_{ch} \frac{\mu_o k T}{q} \frac{n(L) - n(0)}{L} \end{aligned} \quad 6.2.1$$

where Area is the area of the conducting channel in the x and z planes (see figure 2.1.1);

W is the channel width;

x_{ch} is the channel thickness;

D_n is the diffusion coefficient of electrons;
and $\frac{dn}{dy}$ is the gradient of n-type carriers along the length of the channel.

In Chapter 2, the quasi-Fermi levels were found to be

$$\psi_n = \psi_f - V_b \quad \text{at the source, } y=0$$

$$\psi_n = \psi_f + V_d - V_b \quad \text{at the drain, } y=L$$

The n-type carrier concentration is found using equation 2.4.2b:

$$n = n_i \exp \left[\frac{q}{k T} (-\psi_n + \psi - V_b) \right]$$

Therefore the concentration at the source ($y=0$) and drain ($y=L$) are

$$n(0) = n_i \exp \left[\frac{q}{k T} (-\psi_f + \psi_s) \right] \quad 6.2.2a$$

$$\text{and } n(L) = n_i \exp \left[\frac{q}{k T} (-\psi_f + \psi_s - V_d) \right] \quad 6.2.2b$$

$$I_d = q W x_{ch} \frac{\mu_o k T}{q L} n_i \exp \left[\beta (-\psi_f + \psi_s) \right] - q W x_{ch} \frac{\mu_o k T}{q L} n_i \exp \left[\beta (-\psi_f + \psi_s - V_d) \right] \quad 6.2.3$$

$$I_d = q W x_{ch} \frac{\mu_o k T}{q L} n_i \exp \left\{ -\beta \psi_f + \beta \psi_s \right\} \left[1 - \exp(-\beta V_d) \right] \quad 6.2.4$$

where ψ_s is the surface potential due to the gate voltage.

Three quantities need to be found in order to calculate subthreshold current using equation 6.2.4: the Fermi potential, the surface potential and the channel thickness. The Fermi potential is a function of the substrate doping and the surface potential, as a function of the gate voltage, is found by solving Poisson's equation and equating field densities at the silicon-silicon dioxide interface. Finally, the channel thickness is calculated by assuming that the channel extends to where the concentration is e^{-1} X surface concentration.

The Fermi potential is given by

$$\psi_f = \frac{k T}{q} \ln \left(\frac{N_{sub}}{n_i} \right) \quad 6.2.5$$

The surface potential is related to the gate voltage by equating the electric displacement on either side of the silicon-silicon dioxide interface. In order to obtain the electric field in the silicon, Poisson's equation is used. In Chapter 2, Poisson's equation was solved for the general case where acceptors, donors, electrons and holes were considered. Here, a p-type substrate is assumed so that donors are not included and majority carriers are omitted since conduction is by minority carriers. The analysis is carried out using the quasi-Fermi level ψ_n although it is assumed that the drain voltage is very low in order to avoid having a significant depletion region around the drain. Hence the quasi Fermi level at all points in the channel is

$$\psi_n = \psi_f - V_b$$

The Poisson equation for minority carriers and an ionised acceptor impurity is

$$\frac{d^2\psi}{dx^2} = \frac{q n_i}{\epsilon_{si}} \left\{ \exp(\beta\psi_f) + \exp(\beta\psi - \beta\psi_n - \beta V_b) \right\} \quad 6.2.6$$

$$\frac{1}{2} \frac{d}{dx} \left(\frac{d\psi}{dx} \right)^2 = \frac{q n_i}{\epsilon_{si}} \left\{ \exp(\beta\psi_f) + \exp(\beta\psi - \beta\psi_n - \beta V_b) \right\} \frac{d\psi}{dx}$$

Integrate from the interface $x=0$, $\psi = \psi_s$ and $\frac{d\psi}{dx} = \frac{d\psi}{dx}_s$ to the bulk where $x=\infty$, $\psi = V_b$ and $\frac{d\psi}{dx} = 0$.

$$\left[\left(\frac{d\psi}{dx} \right)^2 \right]_0^\infty = \left[\frac{2qn_i}{\epsilon_{si}} \left\{ \psi \exp(\beta\psi_f) + \frac{1}{\beta} \exp(\beta\psi - \beta\psi_n - \beta V_b) \right\} \right]_0^\infty$$

Simplifying

$$\left[\left(\frac{d\psi}{dx} \right)^2 \right]_s = \frac{2qn_i}{\beta\epsilon_{si}} \left\{ (\beta\psi_s - \beta V_b) \exp(\beta\psi_f) \right\} +$$

$$\frac{2qn_i}{\beta\epsilon_{si}} \left\{ \exp(\beta\psi_s - \beta\psi_n - \beta V_b) - \exp(-\beta\psi_n) \right\} \quad 6.2.7$$

The negative square root has to be chosen since the potential is higher at the surface than in the bulk.

$$\left(\frac{d\psi}{dx} \right)_s = - \left(\frac{2qn_i}{\beta\epsilon_{si}} \right)^{\frac{1}{2}} \left\{ (\beta\psi_s - \beta V_b) \exp(\beta\psi_f) + \right.$$

$$\left. \exp(\beta\psi_s - \beta\psi_n - \beta V_b) - \exp(-\beta\psi_n) \right\}^{\frac{1}{2}} \quad 6.2.8$$

This equation is a reduced version of equation 2.4.8 and after some algebraic manipulation, it becomes

$$\left(\frac{d\psi}{dx}\right)_s = -\frac{2^{\frac{1}{2}}}{\beta L_B} \left\{ (\beta\psi_s - \beta V_b) + \left(\frac{n_i}{N_{sub}}\right)^2 \exp(\beta\psi_f - \beta\psi_n) \left[\exp(\beta\psi_s - \beta V_b) - 1 \right] \right\}^{\frac{1}{2}} \quad 6.2.9$$

Assuming that the $\left(\frac{n_i}{N_{sub}}\right)^2$ term is very low in weak inversion and that $\exp(\beta\psi_s)$ is large in strong inversion then

$$\left(\frac{d\psi}{dx}\right)_s = -\frac{2^{\frac{1}{2}}}{\beta L_B} \left\{ (\beta\psi_s - \beta V_b) + \left(\frac{n_i}{N_{sub}}\right)^2 \exp(\beta\psi_f - \beta\psi_n) \exp(\beta\psi_s - \beta V_b) \right\}^{\frac{1}{2}}$$

If $\psi_n = \psi_f - V_b$ then

$$\left(\frac{d\psi}{dx}\right)_s = -\frac{2^{\frac{1}{2}}}{\beta L_B} \left\{ (\beta\psi_s - \beta V_b) + \left(\frac{n_i}{N_{sub}}\right)^2 \exp(\beta\psi_s) \right\}^{\frac{1}{2}} \quad 6.2.10$$

Justification for the above approximation, which was also used by Brews³⁷, can be found in figure 6.2.1 where ψ_s is plotted against $\frac{d\psi}{dx}$ for substrate biases of 0, -1 and -2V, using both equation 6.2.9 and 6.2.10. The difference is barely visible.

Equating the field densities at the silicon-silicon dioxide interface yields:

$$\begin{aligned} \epsilon_{ox} \frac{(V_g - \psi_s)}{t_{ox}} &= \epsilon_{si} \frac{2^{\frac{1}{2}}}{\beta L_B} \left\{ (\beta\psi_s - \beta V_b) + \left(\frac{n_i}{N_{sub}}\right)^2 \exp(\beta\psi_s) \right\}^{\frac{1}{2}} \\ \Rightarrow \psi_s &= V_g - \frac{t_{ox}}{\epsilon_{ox}} \frac{\epsilon_{si} 2^{\frac{1}{2}}}{\beta L_B} \left\{ (\beta\psi_s - \beta V_b) + \left(\frac{n_i}{N_{sub}}\right)^2 \exp(\beta\psi_s) \right\}^{\frac{1}{2}} \quad 6.2.11 \end{aligned}$$

Unfortunately it is difficult to solve for ψ_s from this equation but the reverse procedure (V_g from ψ_s) is straight forward. A lookup table or iterative method

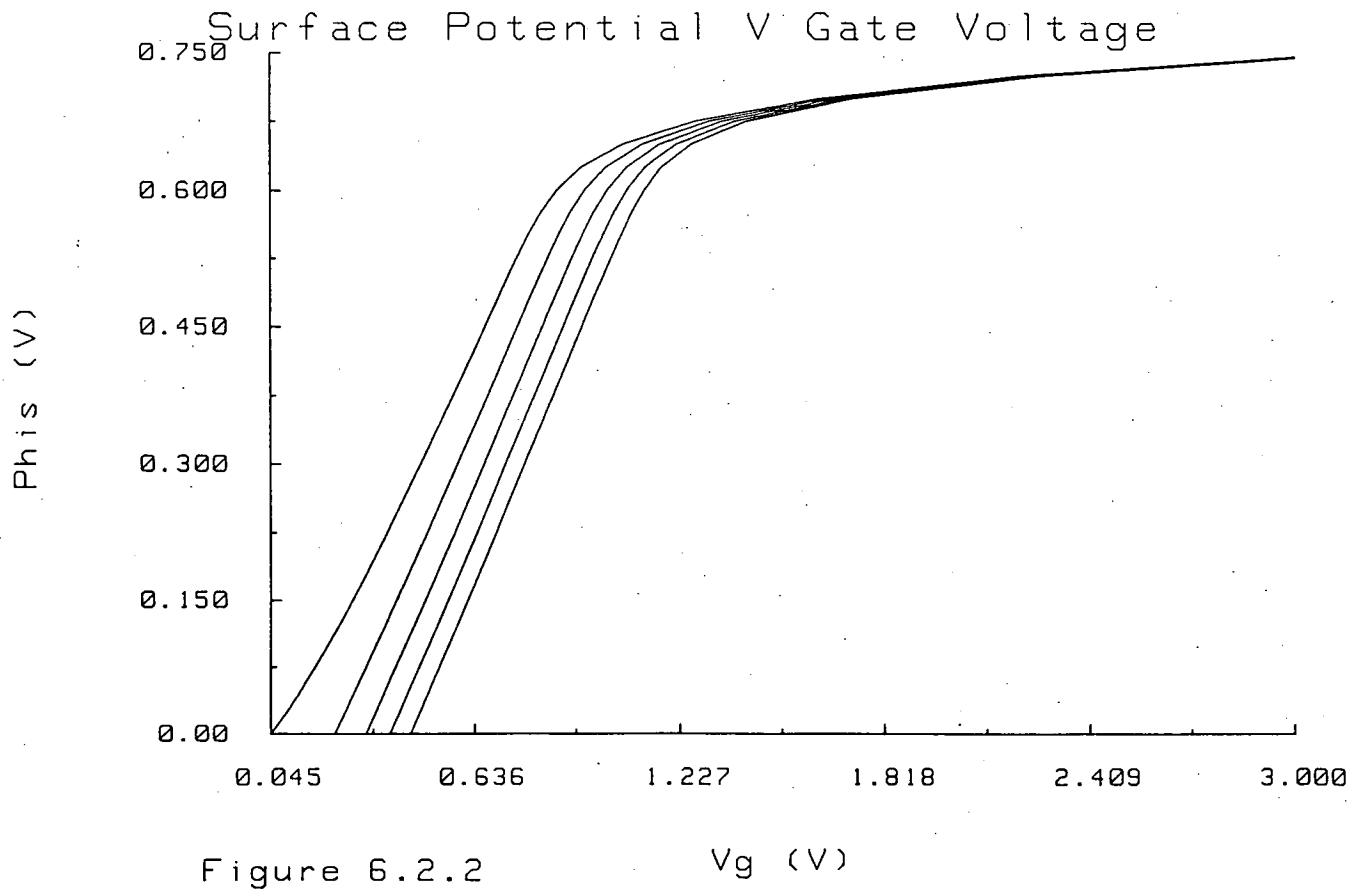
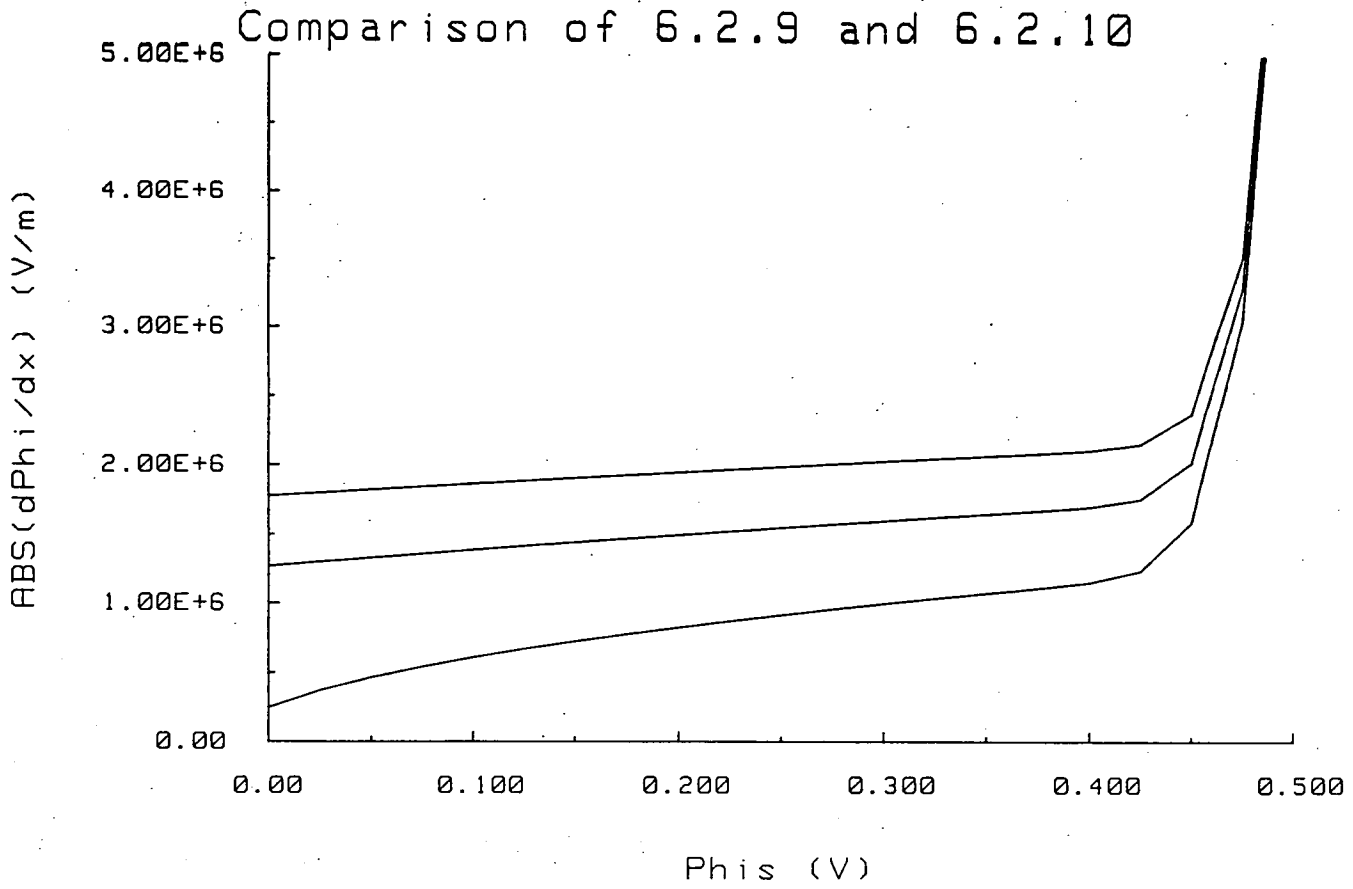


Figure 6.2.2

V_g (V)

is needed. The relationship between ψ_s and V_g according to equation 6.2.11 at substrate biases of 0,-0.5,-1,-1.5 and -2V is shown in figure 6.2.2. The third quantity which needs to be calculated in order to use equation 6.2.4 is the channel thickness x_{ch} . A commonly adopted procedure^{73,20} is to estimate the channel depth as being the depth at which the carrier density is $e^{-1}X$ surface carrier density. Using Boltzmann statistics

$$\begin{aligned} n_s &= n_i \exp \left[\beta(-\psi_n + \psi_s - V_b) \right] \\ n(x) &= n_i \exp \left[\beta(-\psi_n + \psi(x) - V_b) \right] \\ \Rightarrow n(x) &= n_s \exp \left[\beta(\psi(x) - \psi_s) \right] \end{aligned} \quad 6.2.12$$

where ψ_s , n_s are the surface potential and concentration respectively.

$$\psi(x_{ch}) = \psi_s - \frac{kT}{q} \quad 6.2.13$$

If the electric field in the x-direction is constant then

$$x_{ch} = - \frac{kT}{qE_x} \quad 6.2.14$$

Using equation 6.2.10.

$$x_{ch} = - \frac{kT}{q} \left[- \frac{2^{\frac{1}{2}}}{\beta L_B} \left\{ (\beta \psi_s - \beta V_b) + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta \psi_s) \right\}^{\frac{1}{2}} \right]^{-1} \quad 6.2.15$$

A model consisting of equations 6.2.4, 6.2.5, 6.2.11 and 6.2.15 has now been developed. A typical set of device characteristics calculated using these equations are shown in figure 6.2.3. The device parameters used are:

$$N_{sub} = 10^{21} m^{-3}$$

$$\mu_o = 600 cm^2 V_s^{-1}$$

$$t_{ox} = 600 \text{ \AA}$$

$$\text{and } W \text{ and } L = 5 \mu m$$

A most useful quantity for assessing the subthreshold operation is the subthreshold swing, S. This is the change in gate voltage required to reduce the subthreshold current by one decade. It is defined by

$$S = \ln(10) \frac{dV_g}{d(\ln I_d)} \quad 6.2.16$$

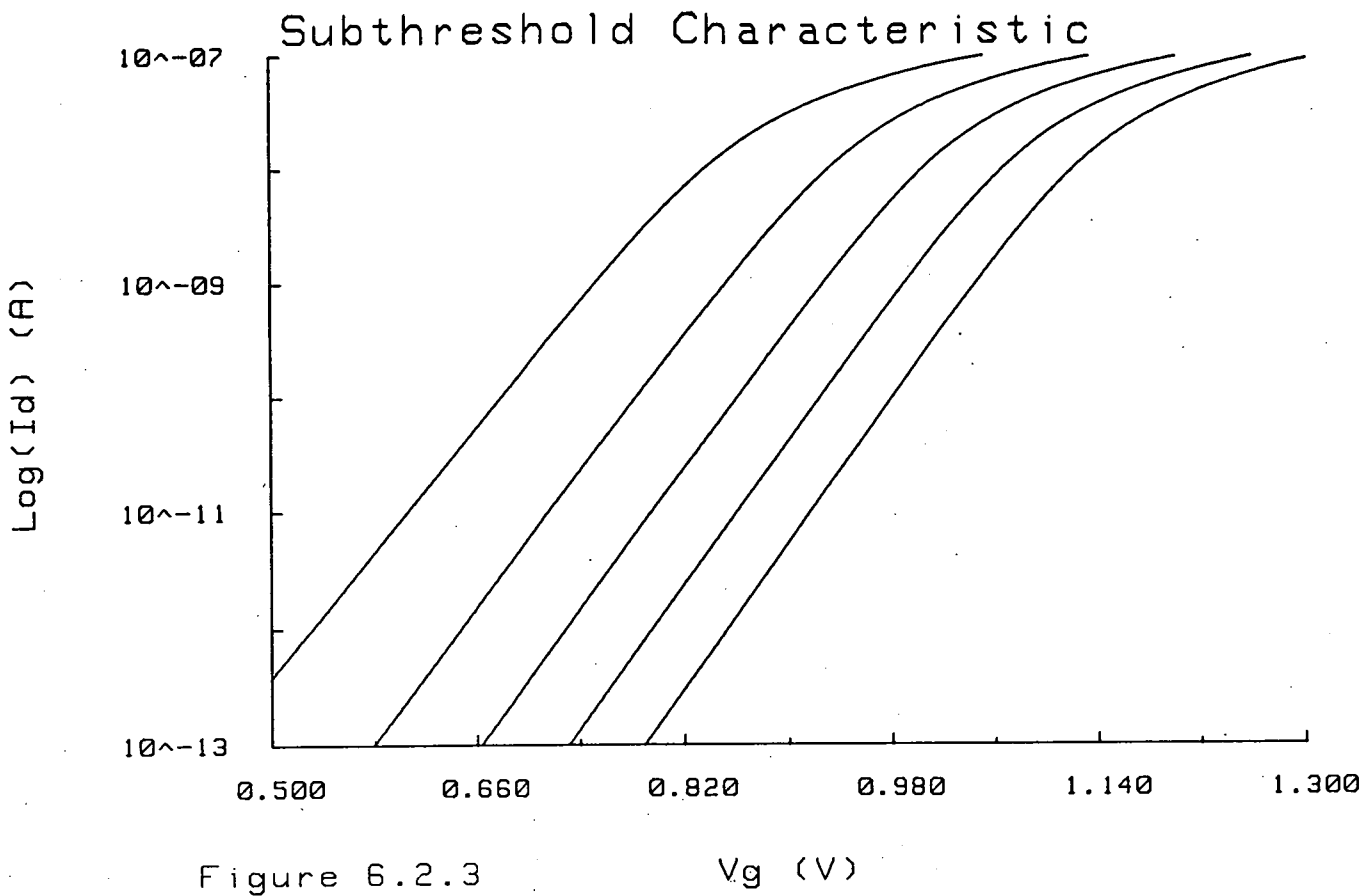


Figure 6.2.3

V_g (V)

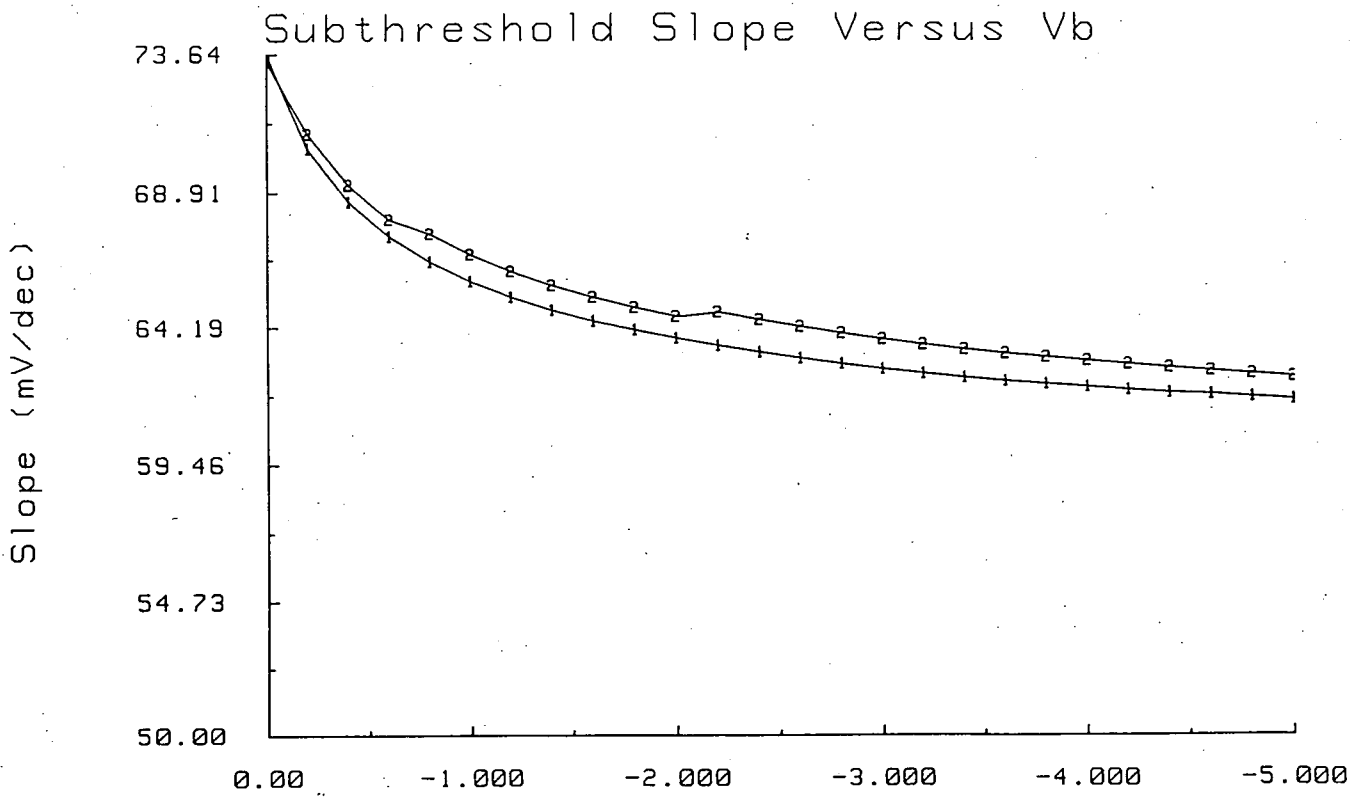


Figure 6.2.4

V_b (V)

$$S = \frac{kT}{q} \ln(10) \frac{d\beta V_g}{d\beta\psi_s} \left\{ \frac{d(\ln I_d)}{d\beta\psi_s} \right\}^{-1} \quad 6.2.17$$

Differentiating 6.2.11

$$\frac{d\beta V_g}{d\beta\psi_s} = 1 + \frac{t_{ox}}{\epsilon_{ox}} \frac{\epsilon_{si}}{L_B} \frac{1}{2^{\frac{1}{2}}} \left\{ (\beta\psi_s - \beta V_b) + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\}^{-\frac{1}{2}} X$$

$$\left\{ 1 + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\} \quad 6.2.18$$

From equations 6.2.4 and 6.2.11

$$\frac{d(\ln I_d)}{d\beta\psi_s} = 1 - \frac{1}{2} \left\{ \beta\psi_s - \beta V_b + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\}^{-1} X$$

$$\left\{ 1 + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\} \quad 6.2.19$$

The depletion layer capacitance is useful for minimising the expression for subthreshold swing.

$$C_d(\psi_s) = \frac{\delta Q_s}{\delta\psi_s}$$

$$\frac{C_d(\psi_s)}{C_{ox}} = \frac{a}{2} \left\{ \beta\psi_s - \beta V_b + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\}^{-\frac{1}{2}} \left\{ 1 + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\} \quad 6.2.20$$

where

$$a = 2^{\frac{1}{2}} \frac{\epsilon_{si}}{L_B} \frac{t_{ox}}{\epsilon_{ox}}$$

Now

$$S = \frac{kT}{q} \ln(10) \left\{ 1 + \frac{a}{2} \left\{ (\beta\psi_s - \beta V_b) + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\}^{-\frac{1}{2}} \left\{ 1 + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\} \right\} X$$

$$S = \frac{kT}{q} \ln(10) \left\{ 1 + \frac{C_d(\psi_s)}{C_{ox}} \right\} \left\{ 1 - \frac{2}{a^2} \left(\frac{C_d(\psi_s)}{C_{ox}} \right)^2 \left\{ 1 + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\}^{-1} \right\}^{-1} \left\{ 1 - \frac{1}{2} \left\{ \beta\psi_s - \beta V_b + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\}^{-1} \left\{ 1 + \left(\frac{n_i}{N_{sub}} \right)^2 \exp(\beta\psi_s) \right\} \right\}^{-1} \quad 6.2.21$$

If interface traps are present then another capacitor is located in parallel with the depletion layer capacitance. Equation 6.2.21 is modified by replacing $C_d(\psi_s)$ by $C_d(\psi_s) + C_{it}$. The SPICE model includes the parameter N_{fs} , intended to represent the fast state density, and this is used to set the subthreshold slope.

Figures 6.2.4 and 6.2.5 show the relationships between S and V_b and between S and N_{sub} as they were simulated using equation 6.2.21. In 6.2.4, the slope was found at current levels of $10^{-10.5}$ (1) and 10^{-9} (2) and in 6.2.5, a current level of $10^{-10.5}$ was used.

The theoretical treatment required to predict the subthreshold operation of non-uniformly doped devices is extremely complex. Brews²¹ makes a thorough analysis of the effect of implants on threshold by replacing the depleted portion of the implant by a lumped dose placed at the centroid. As a first step, in order to explain the results obtained here when investigating the sensitivity of the subthreshold slope to implant dose, the theory developed for uniform doping will be used.

6.3 Subthreshold Slope Versus Doping Profile Experiment

Six samples were prepared with different channel profiles. The implants used on each sample are listed in figure 6.3.1. The profiles resulting from the different combinations of implants were simulated using SUPREM II. The simulated profiles just after implant and at the end of process for sample 1 are shown in figures 6.3.2 and 6.3.3 respectively. The impurity redistribution during the high temperature steps means that there are no large impurity gradients in the final channel profiles. Profiles after implant and at the end of process for samples 2 to 6 are shown in figures 6.3.4 to

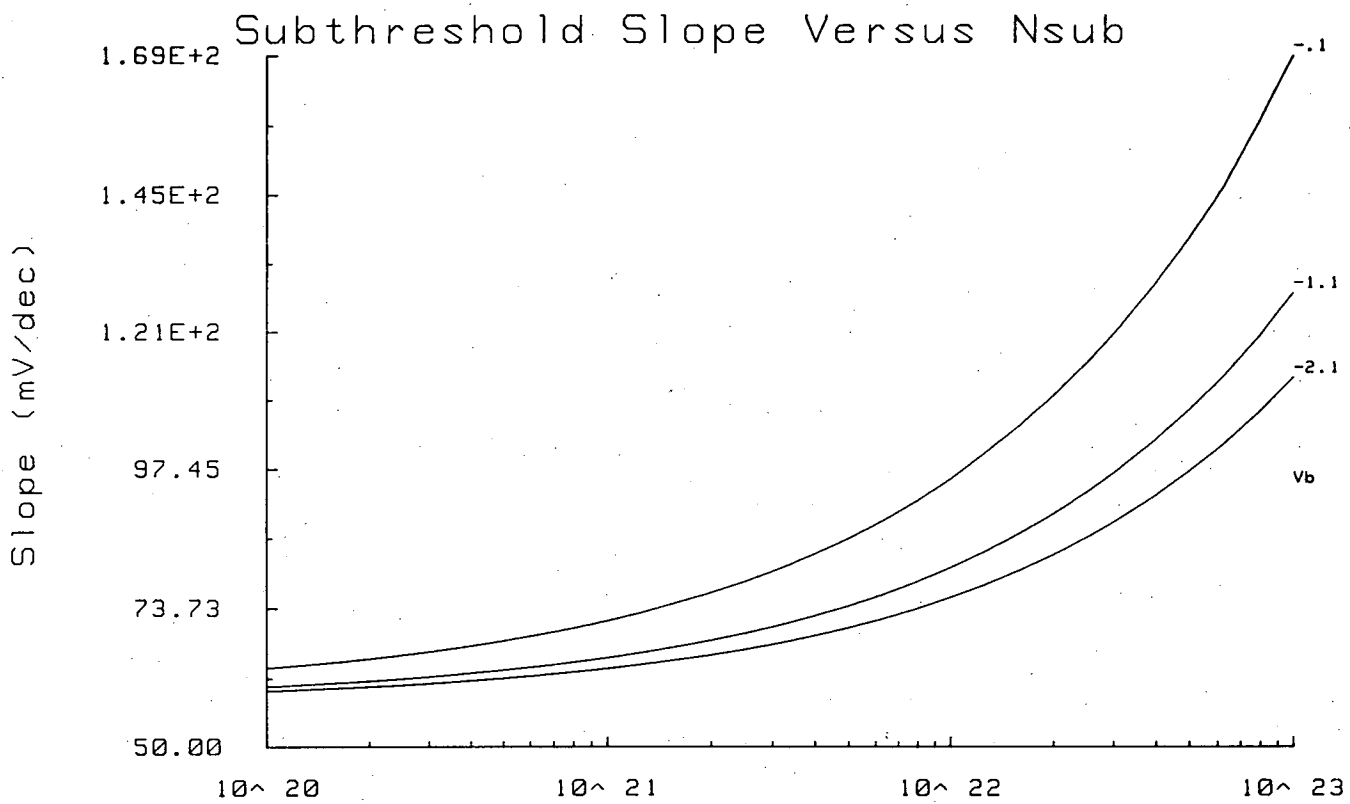


Figure 6.2.5

N_{sub} (m^{-3})

Figure 6.3.1 Experimental Implants

Sample Number	Boron Implants	
	Dose (cm^{-2})	Energy (keV)
1	3E11	40
	8E11	180
2	4E11	40
	4E11	320
3	5E12	180
4	3E11	40
	1E13	320
5	4E11	40
	1E12	320
6	4E11	40

Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

13: 11: 41 29Jan88

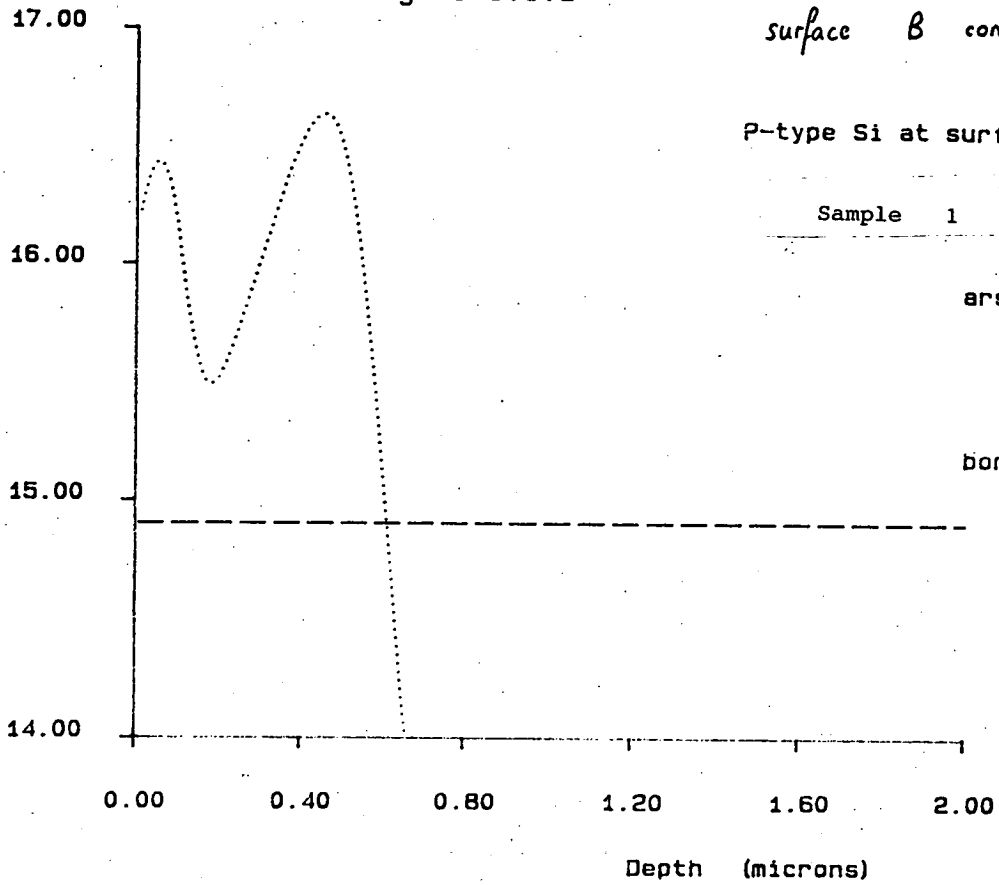
Channel Implant : B 3E11 @ 40 keV
B 8E11 @ 180 keV

Figure 6.3.2

surface B conc : 1.369E16

P-type Si at surface

Sample 1



Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

13: 16: 44 29Jan88

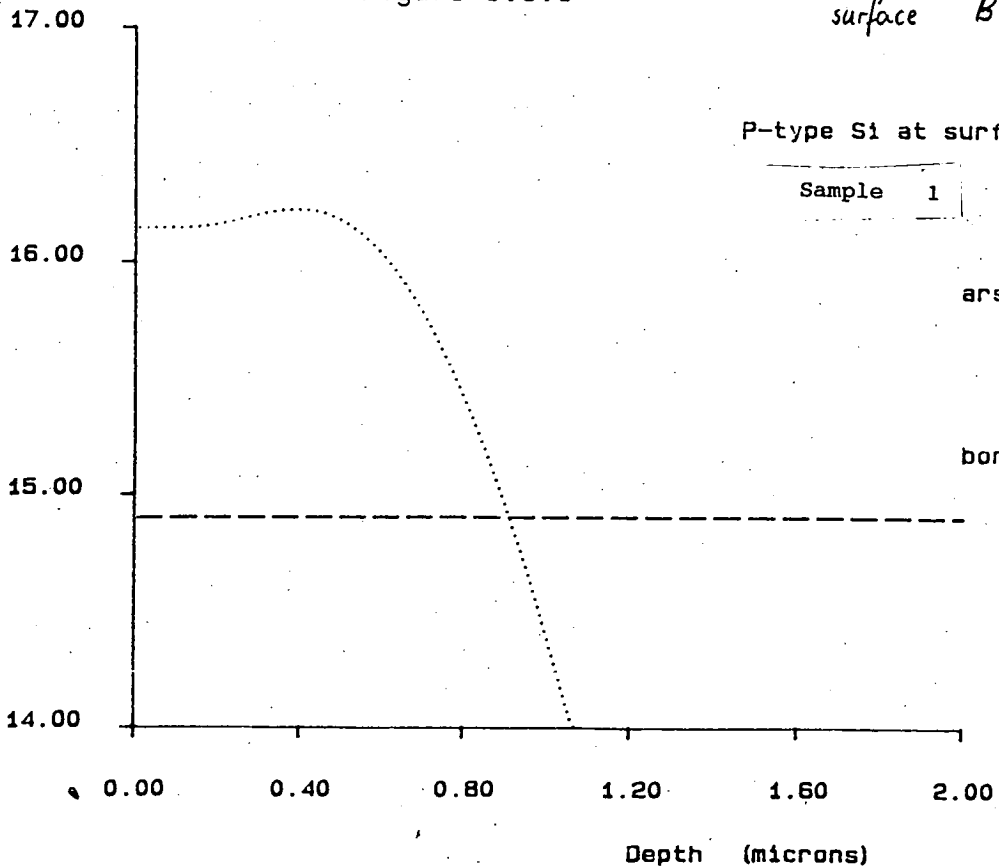
Final Channel Profile : B 3E11 @ 40 keV
B 8E11 @ 180 keV

Figure 6.3.3

surface B conc : 1.404 E6

P-type Si at surface

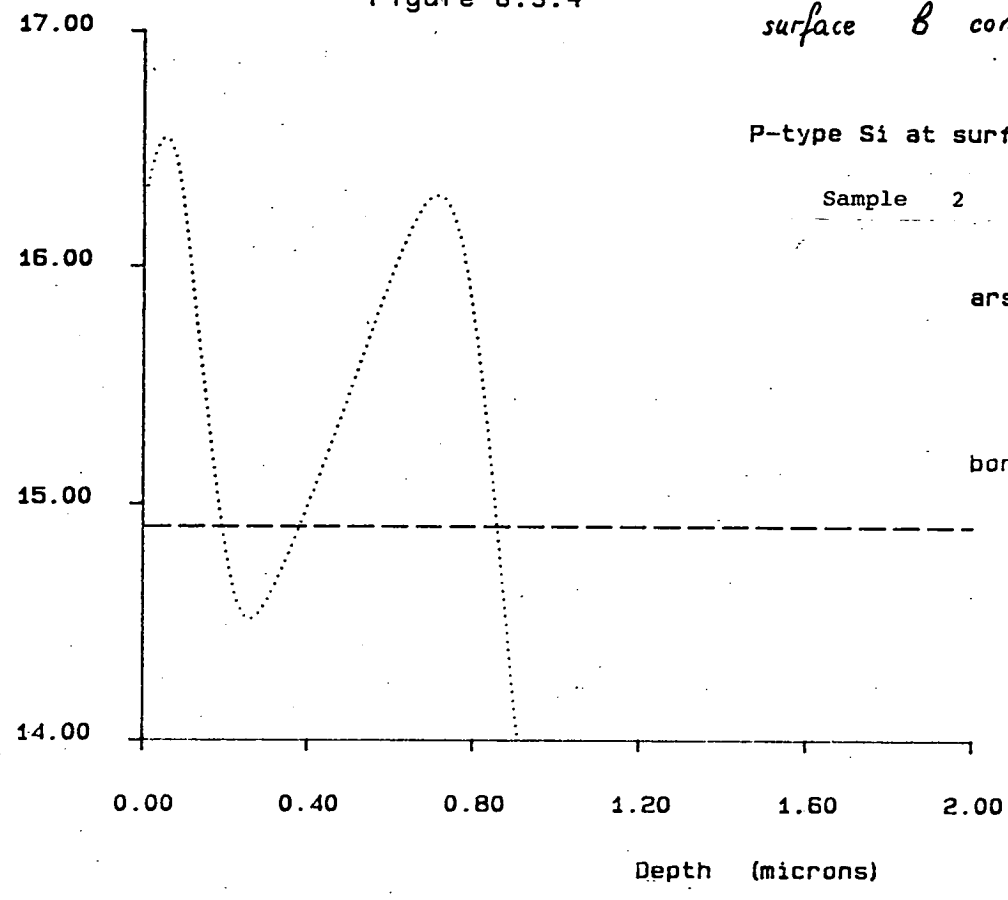
Sample 1



Log Conc. (cm⁻³) SUPREM DOPING PROFILE 13: 19: 54 29Jan86
 Channel Implant : β 4E11 @ 40 keV
 β 4E11 @ 320 keV

Figure 6.3.4

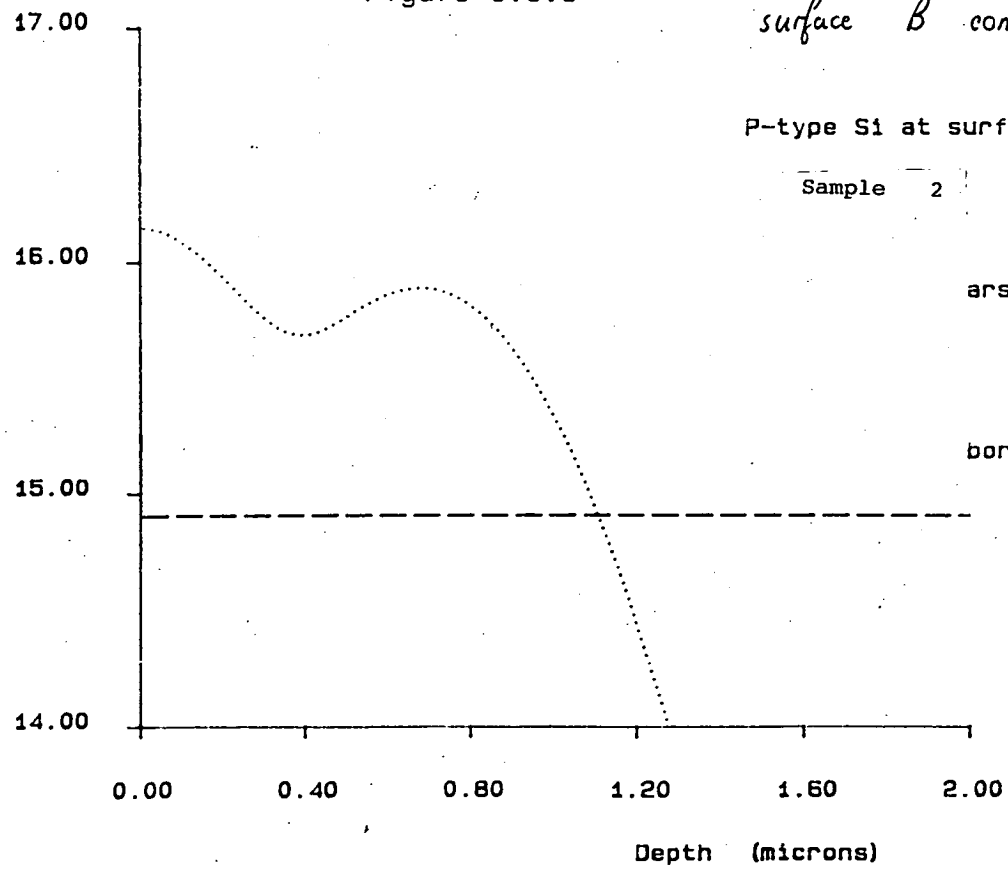
surface β conc : 1.785E16



Log Conc. (cm⁻³) SUPREM DOPING PROFILE 13: 23: 54 29Jan86
 Final Channel Profile : β 4E11 @ 40 keV
 β 4E11 @ 320 keV

Figure 6.3.5

surface β concentration : 1.397E16



Log Conc. (cm⁻³)

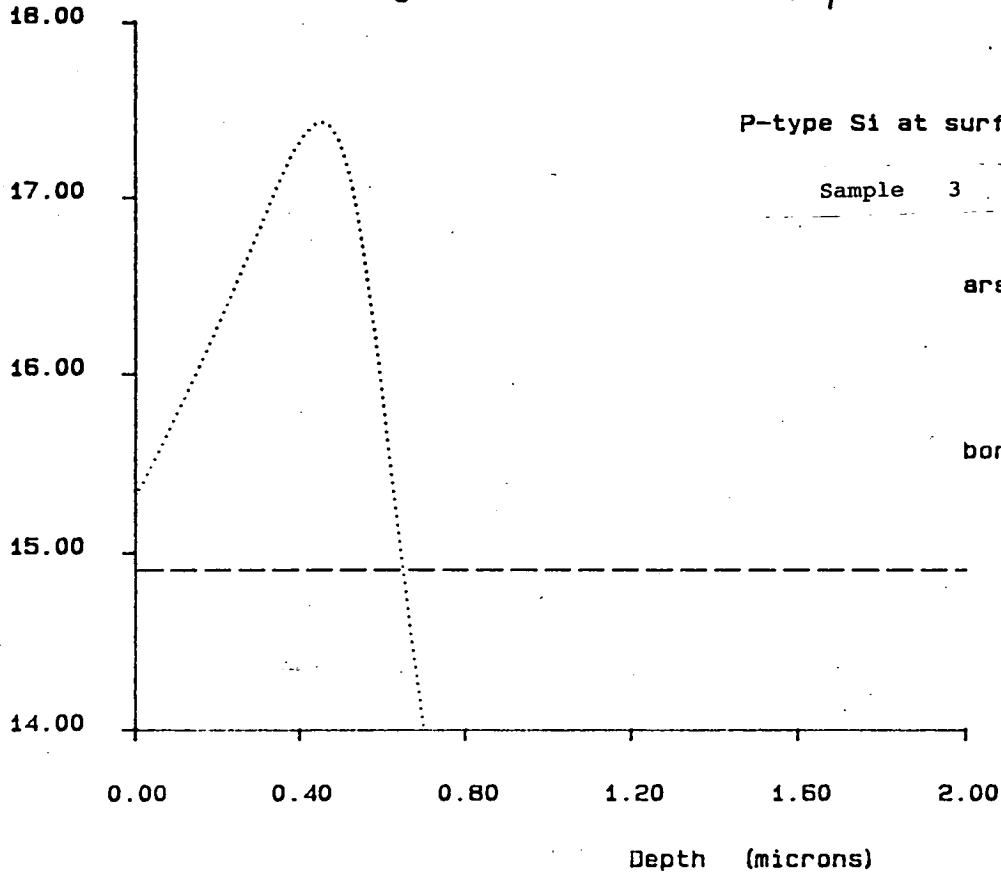
SUPREM DOPING PROFILE

13:33:42 29Jan86

Channel Implant : β SE12 @ 180keV

Figure 6.3.6

surface β conc : $2.083 \text{ E}15$



Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

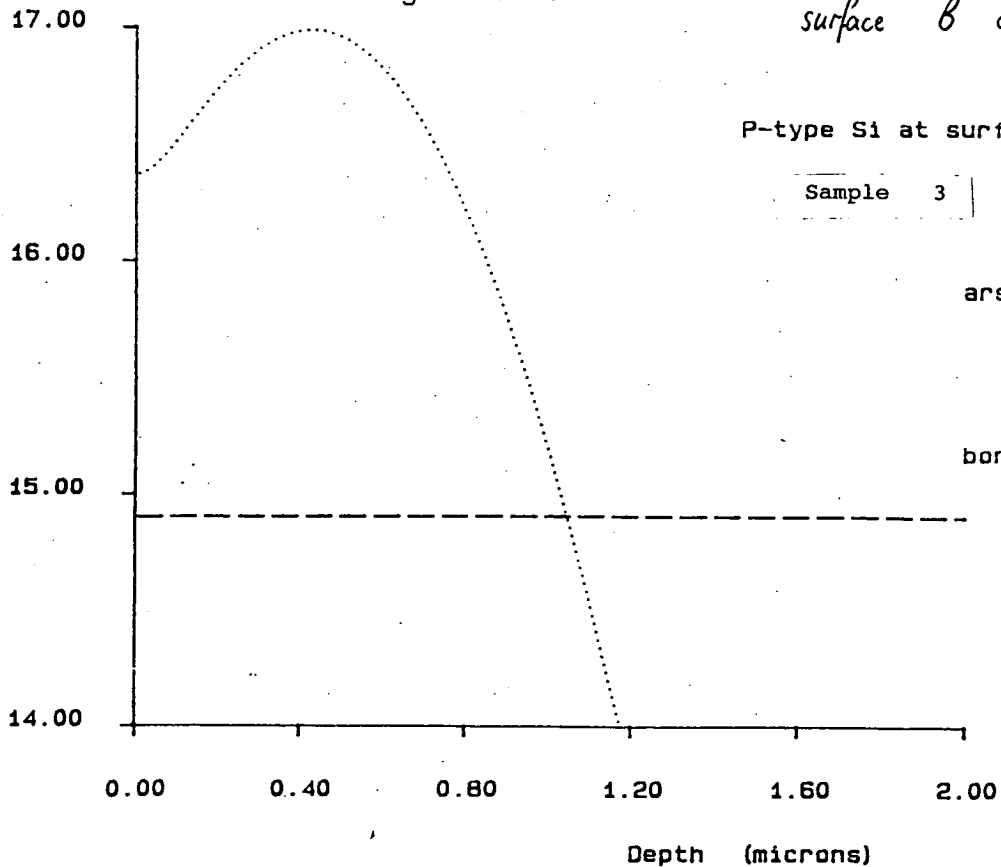
13:30:27 29Jan86

Final Channel Profile :

β SE12 @ 180keV

Figure 6.3.7

surface β conc : $2.323 \text{ E}16$



Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

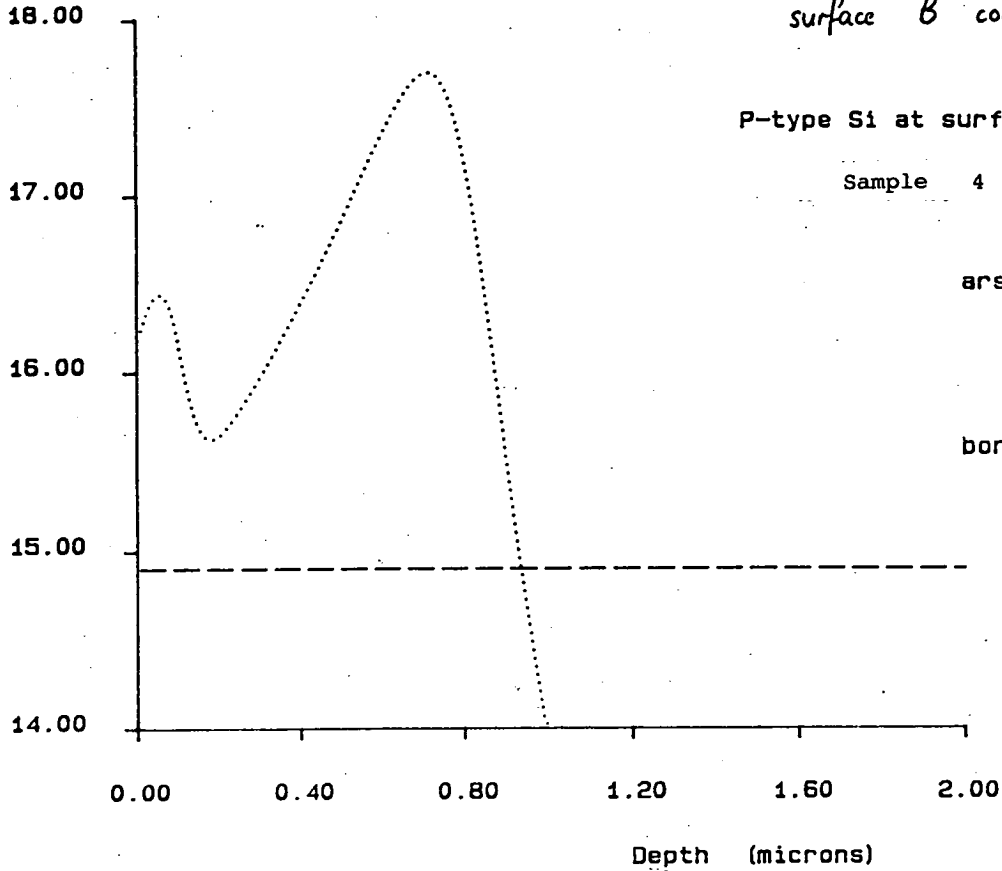
13: 44: 27 29Jan86

Channel Implant :

8 3E11 @ 40 keV
8 1E13 @ 320 keV

Figure 6.3.8

surface B conc : 1.432 E16



Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

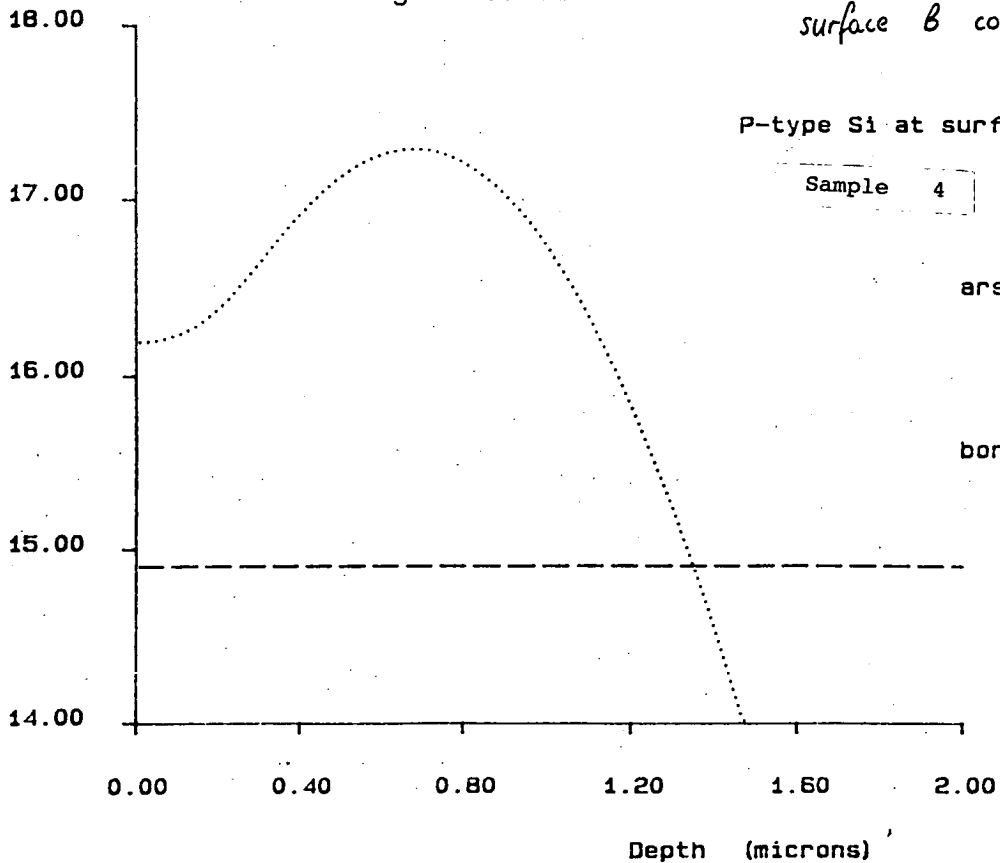
13: 47: 37 29Jan86

Final Channel Profile :

8 3E11 @ 40 keV
8 1E13 @ 320 keV

Figure 6.3.9

surface B conc : 1.546 E16



Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

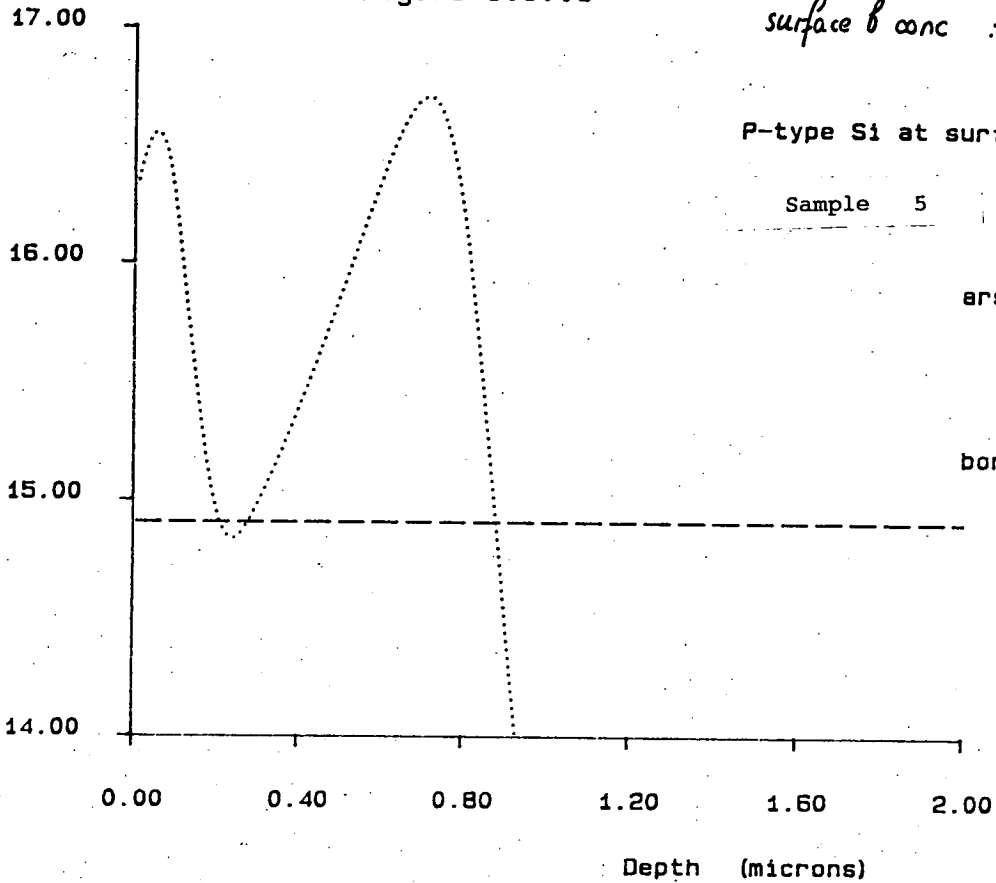
14: 18: 03 29Jan86

Channel Implant :

B 4E11 @ 40 keV
B 1E12 @ 320 keV

Figure 6.3.10

surface B conc : 1.790E16



Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

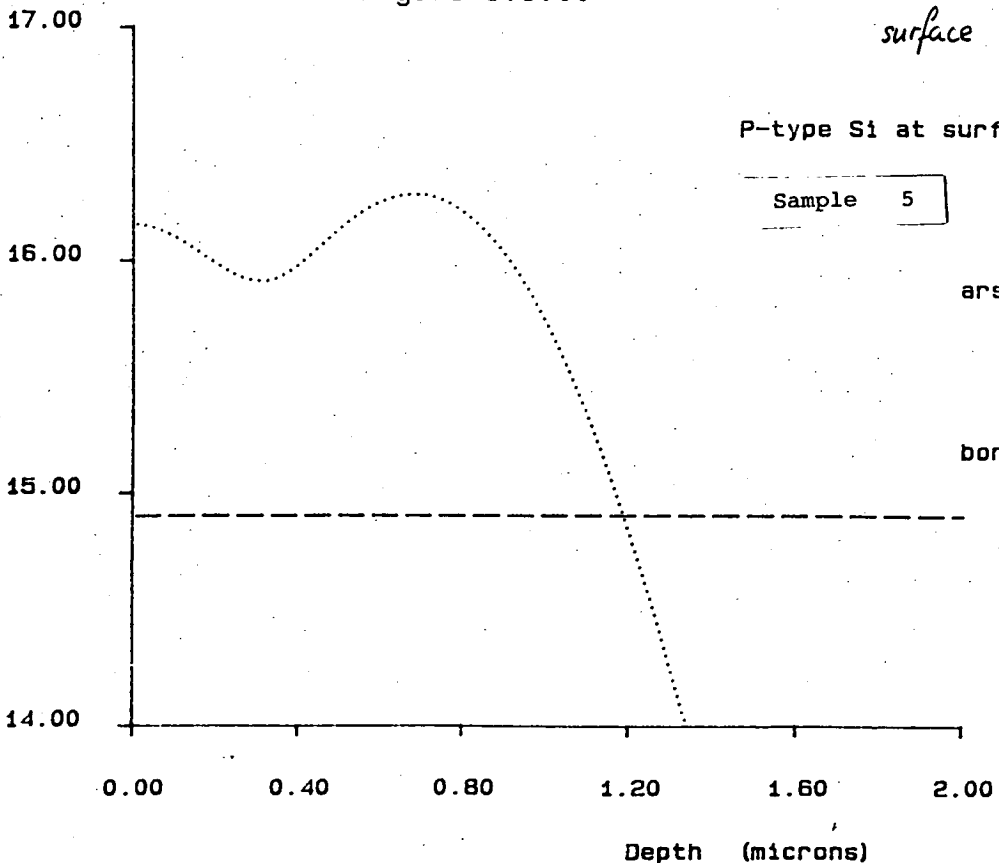
14: 19: 48 29Jan86

Final Channel Profile :

B 4E11 @ 40 keV
B 1E12 @ 320 keV

Figure 6.3.11

surface B conc : 1.428E16



Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

14: 41: 11 29Jan86

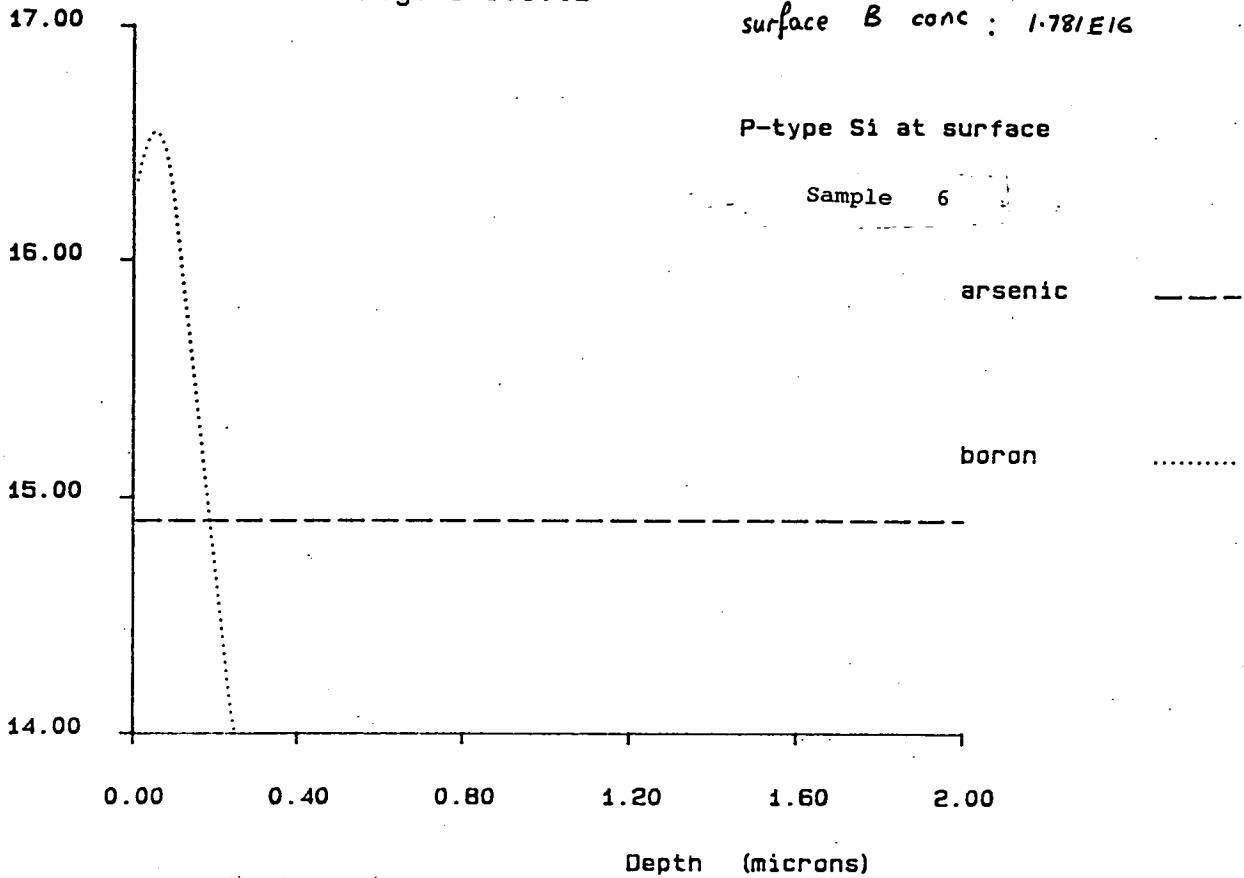
Channel Implant : B 4E11 @40 keV

Figure 6.3.12

surface B conc : 1.781E16

P-type Si at surface

Sample 6



Log Conc. (cm⁻³)

SUPREM DOPING PROFILE

14: 50: 23 29Jan86

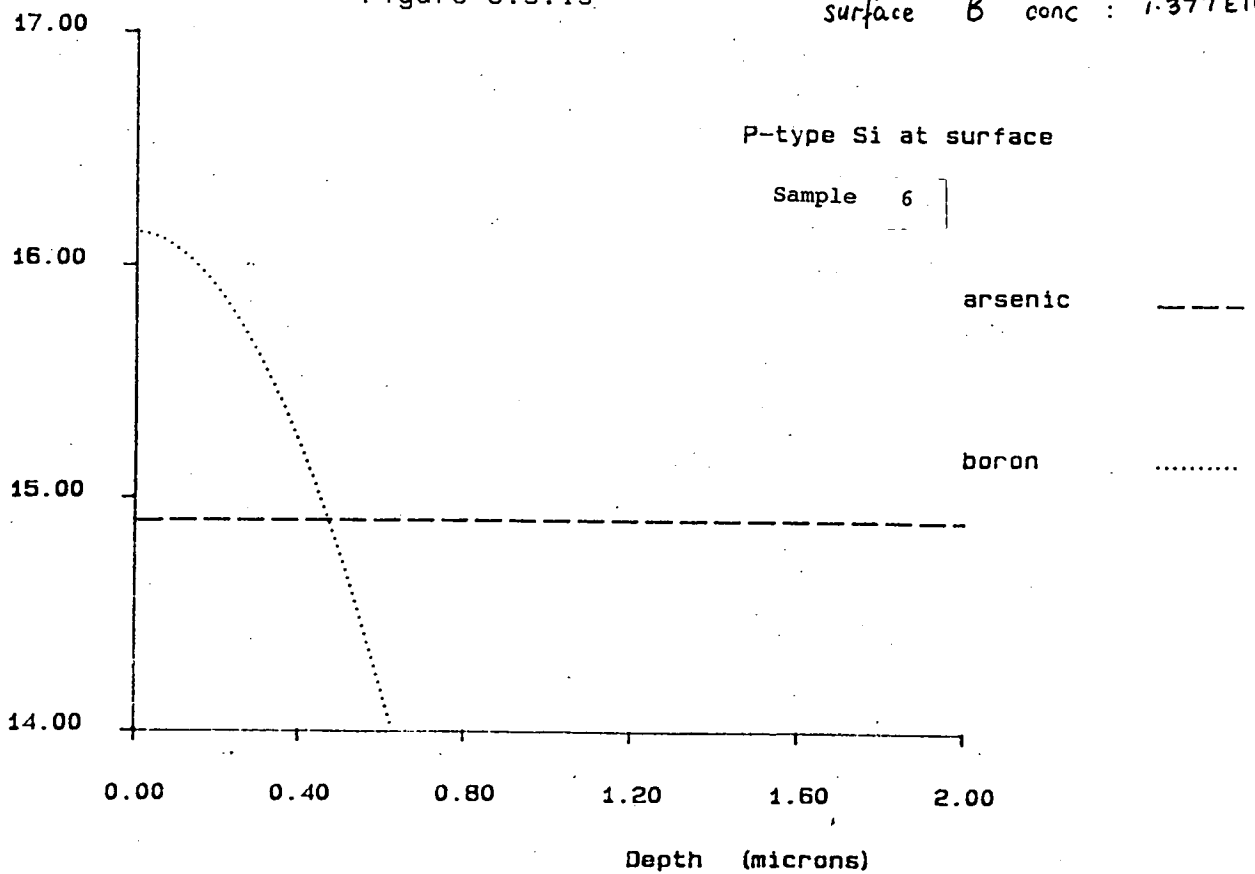
Final Channel Profile : B 4E11 @40 keV

Figure 6.3.13

surface B conc : 1.377E16

P-type Si at surface

Sample 6



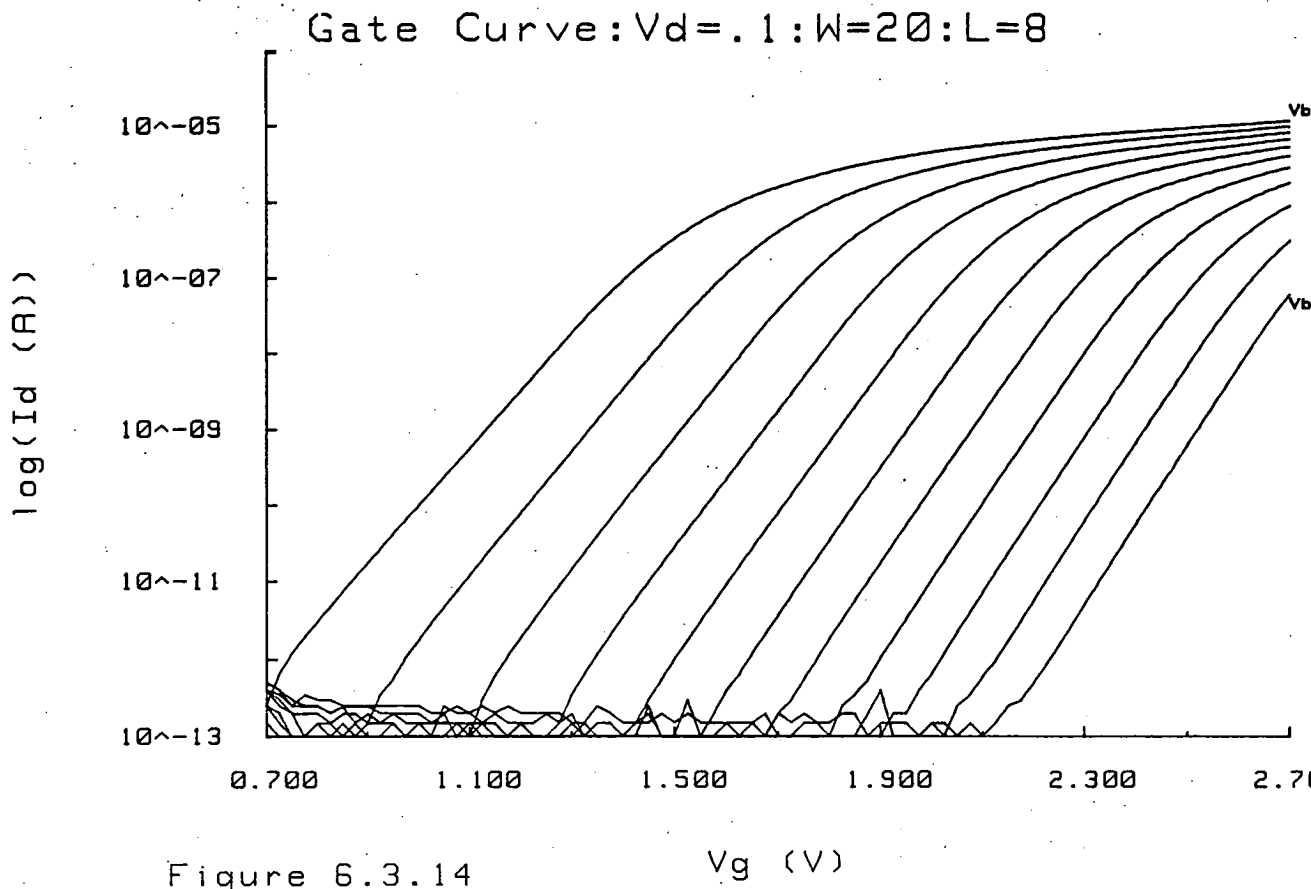
6.3.13. Sample 6 was fabricated using the standard implant used in the EMF $6\mu\text{m}$ NMOS process.

In order to test these specimens, the wafers were diced and some chips of each type were packaged. The packaged devices were placed in turn in an enclosed test jig on the HP4062B test system (described in Chapter 4) and the subthreshold characteristics were measured. For each sample, currents were measured over the range 1pA to $1\mu\text{A}$ at 11 values of substrate bias ranging from 0 to -2V inclusive. The long integration time was selected on the measuring instrument in order to obtain maximum accuracy and time was allowed between forcing voltages and measuring current in order to allow the device to settle in a specified bias condition. The subthreshold characteristics for specimens 1 to 6 are shown in figures 6.3.14 to 6.3.19.

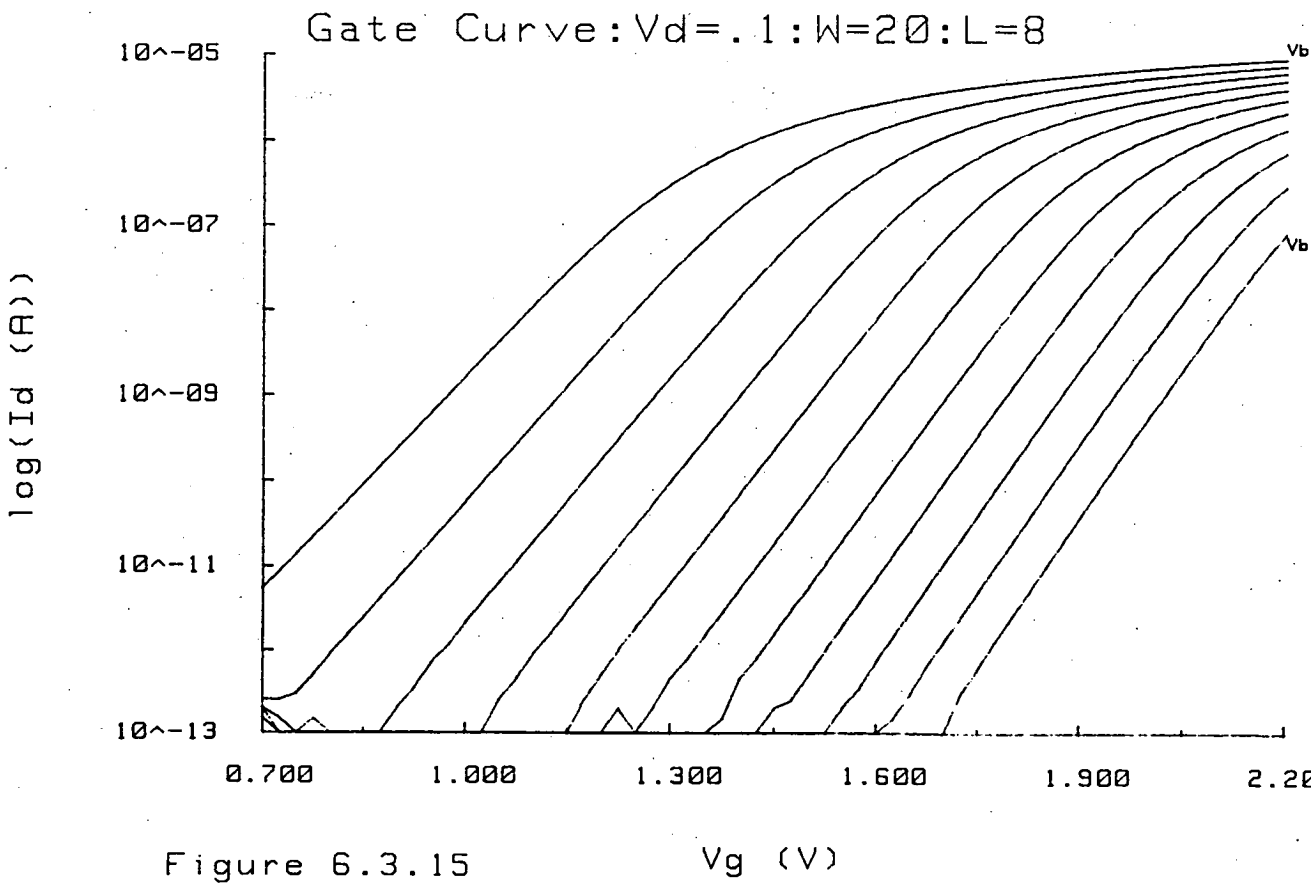
The fourth sample does not turn off completely and there is a leakage current of around 10^{-10}A which is substrate bias dependent. This sample received the highest dose $1\text{X}10^{13}$ at very high energy (320 keV) and also a surface implant of $3\text{X}10^{11}$ at 40 keV. Possibly, despite the high temperature processing, some positive boron ions have not diffused out of the gate oxide because of the high substrate concentration, and hence the channel is weakly turned on even at $V_g = 0$.

The measurement program includes a section to thoroughly analyse the characteristics. Numerical methods (summarised in APPENDIX E) are used to differentiate the curves in order to find the subthreshold slope. Forward and backward difference formulae are used at the beginning and end of the measured curve respectively and central differences on the rest of the curve. Three measurements are used to calculate the derivative at each point. The results are used to plot subthreshold slope against substrate bias at specific current levels. Figure 6.3.20 shows the results for sample 1 at current levels of $10^{-10.5}$ (1) and $10^{-8.5}$ (2).

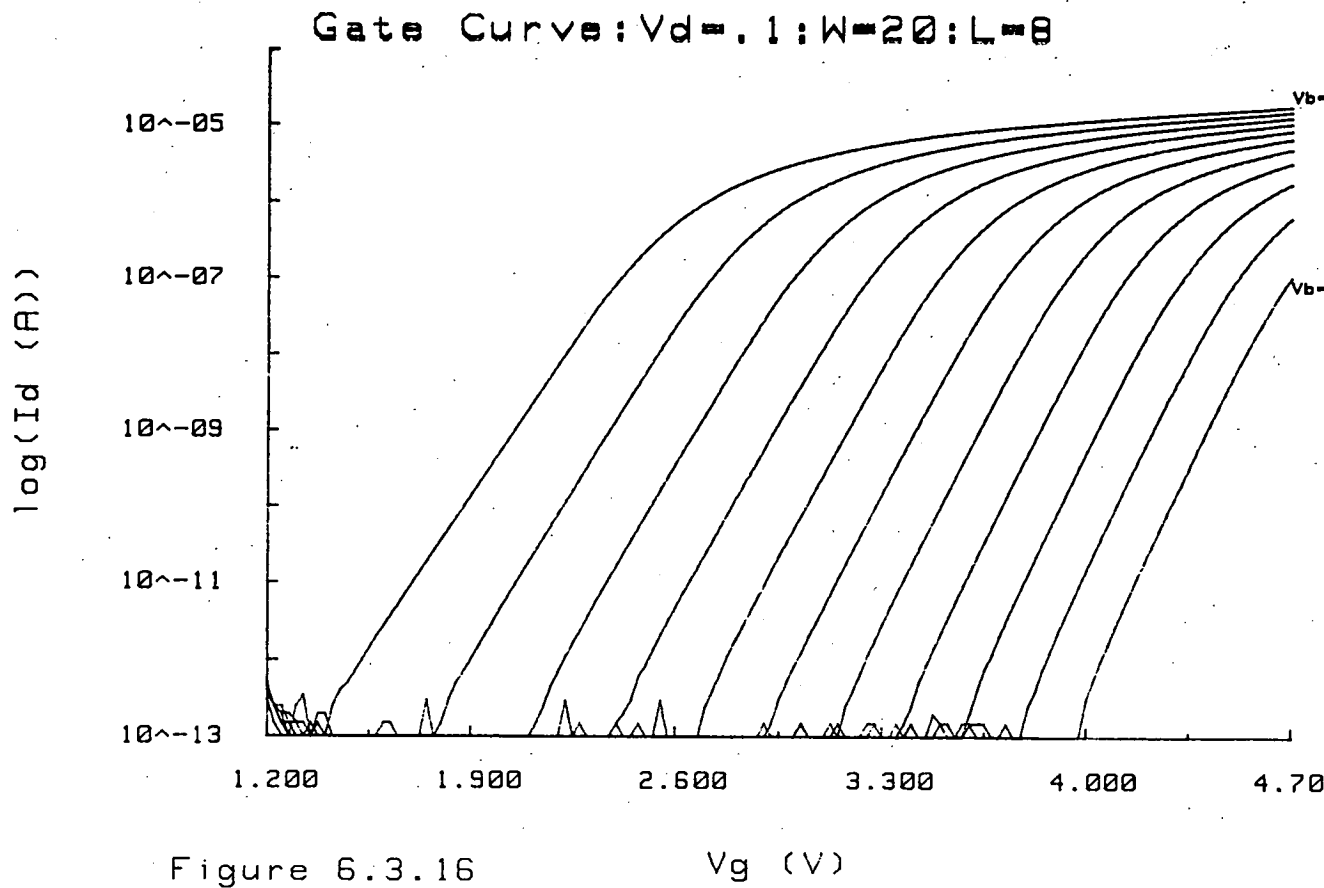
Inverse interpolation is used to find the gate voltages corresponding to particular currents on the subthreshold characteristics. The derivative, and hence subthreshold slope, at that point was then computed using the Stirling formula (see APPENDIX E). Slope against current level for sample 1 is plotted in figure 6.3.21. The analysis results for the other samples, apart from sample 4, are shown in figure 6.3.22 to 6.3.29.



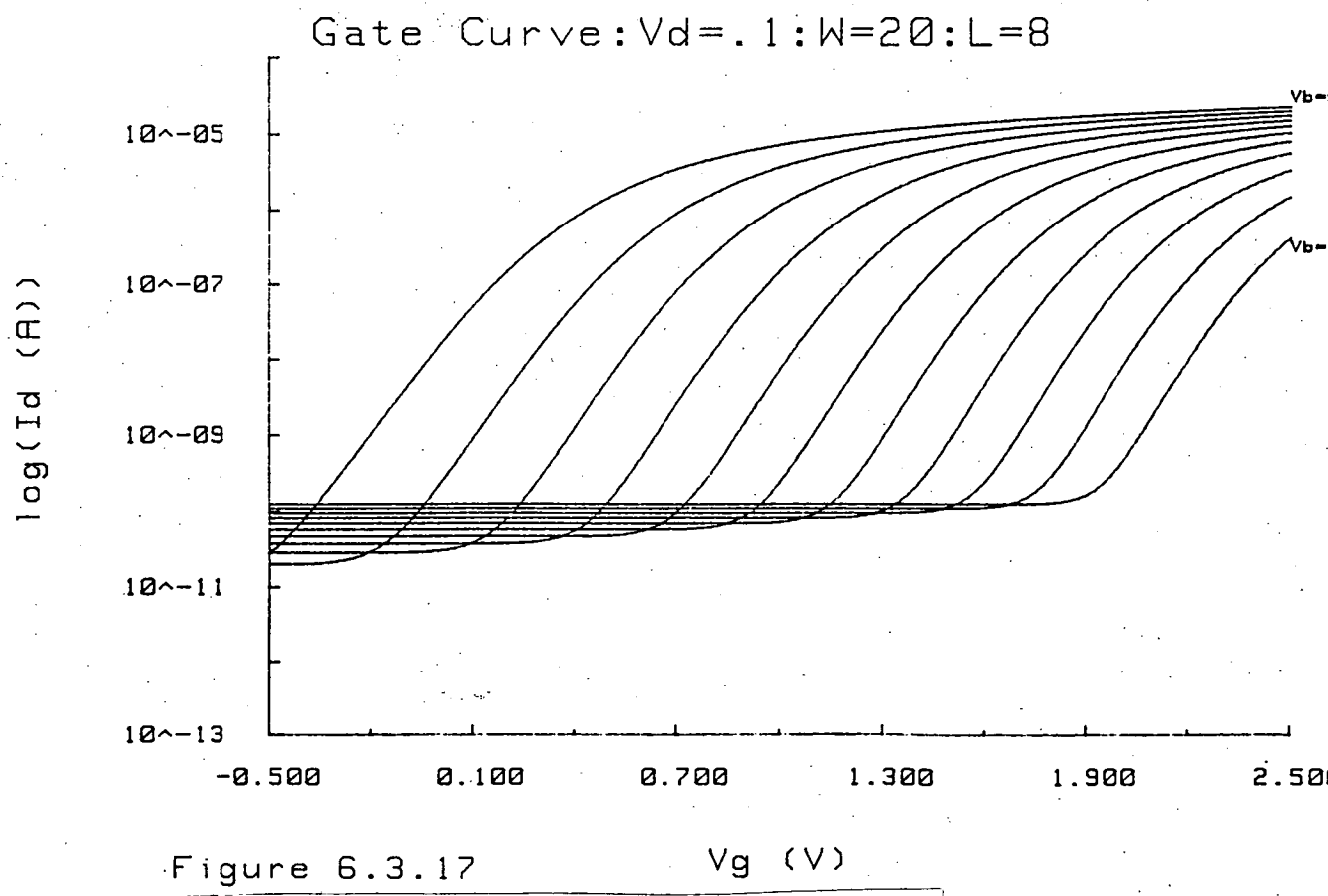
Sample 1 Boron Implants 3E11 at 40keV and 8E11 at 180keV



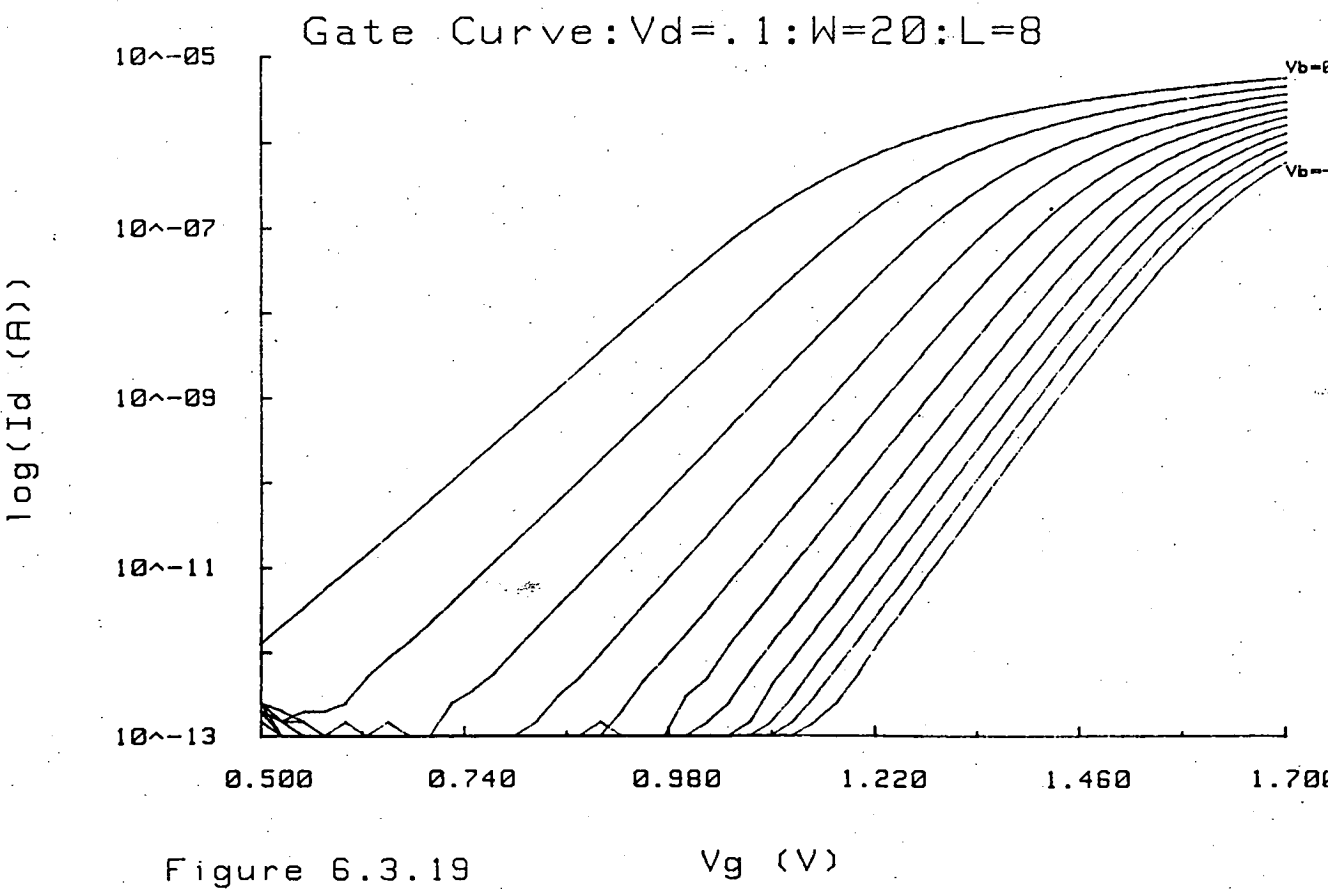
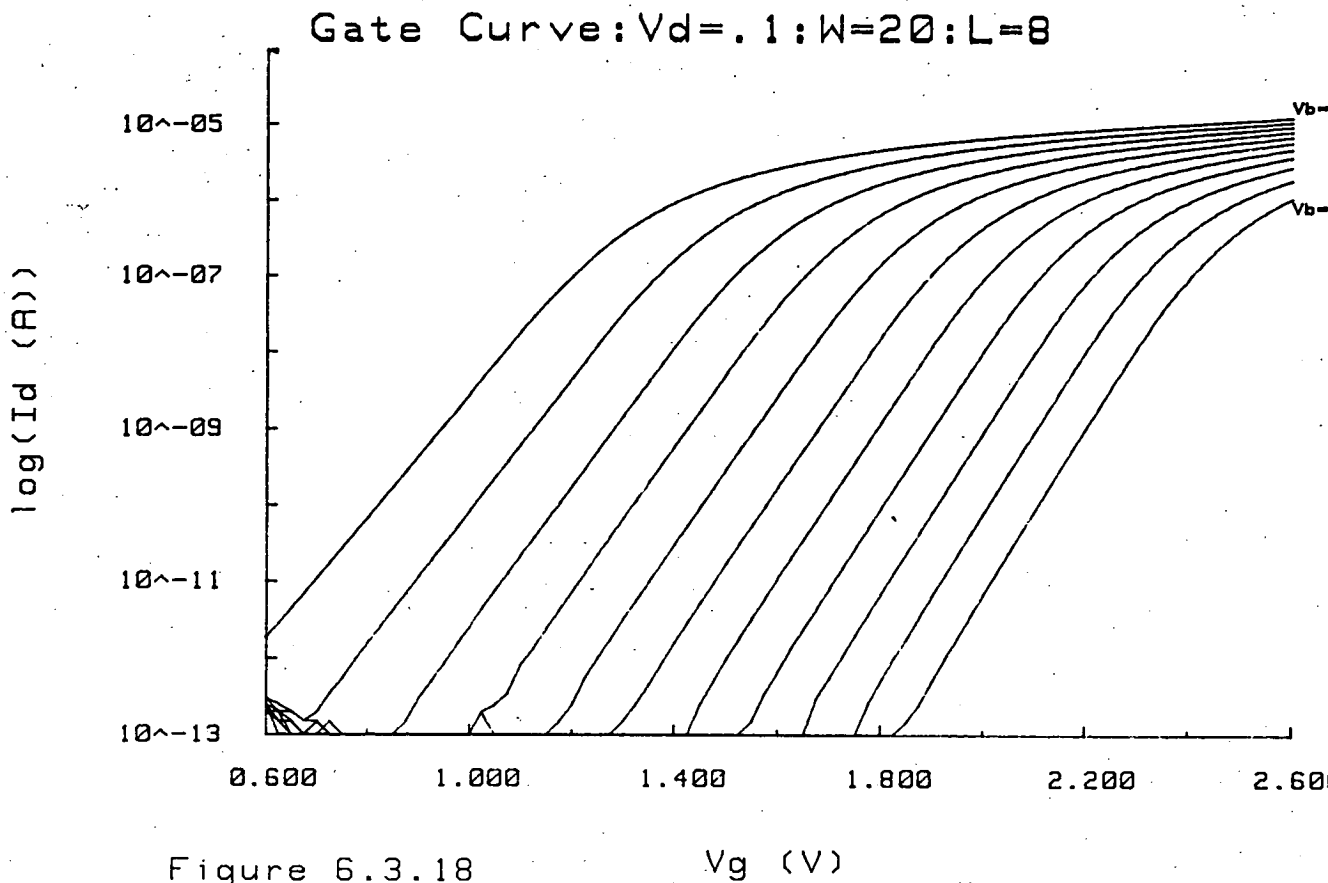
Sample 2 Boron Implants 4E11 at 40keV and 4E11 at 320keV



Sample 3 Boron Implant: $5E12$ at $180keV$



Sample 4 Boron Implants $3E11$ at $40keV$ and $1E13$ at $320keV$



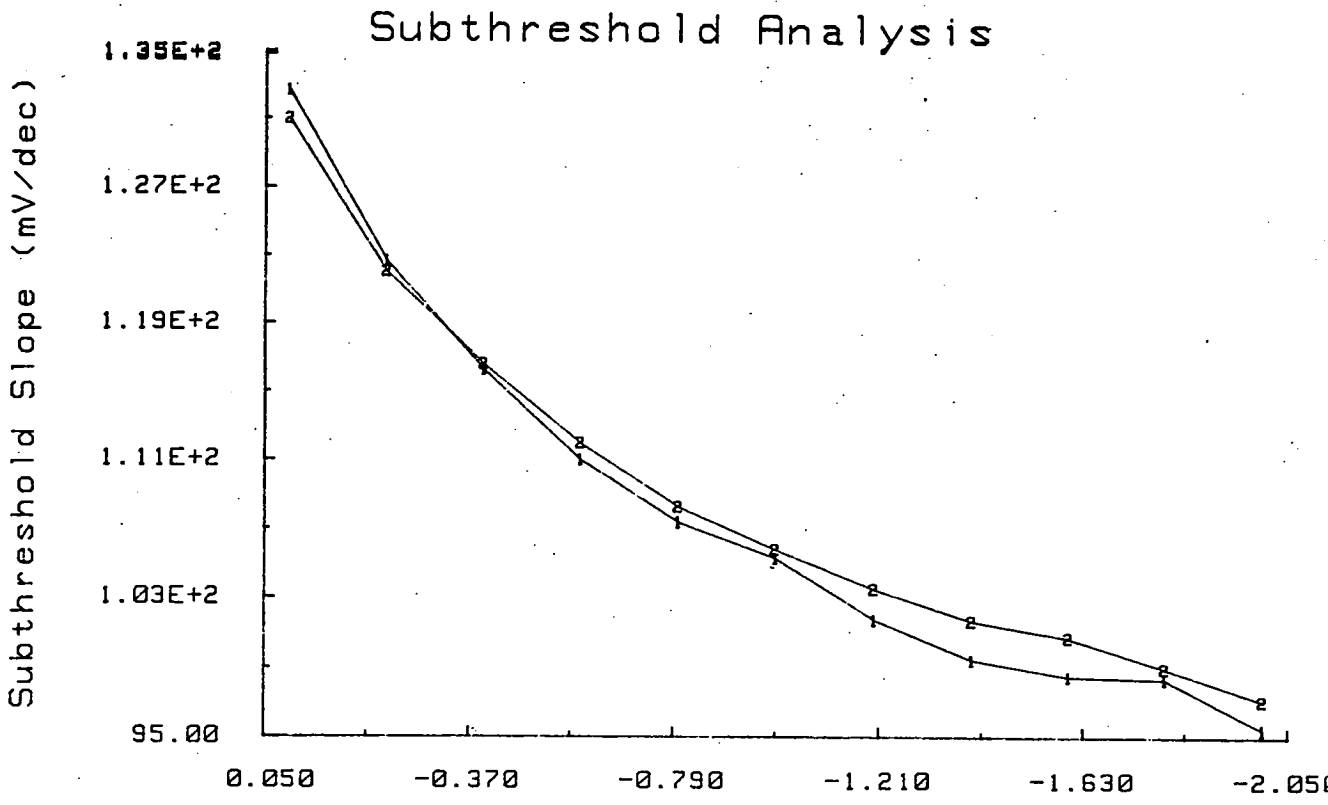


Figure 6.3.20 Substrate Bias (V)

Sample 1 Boron Implants 3E11 at 40keV and 8E11 at 180keV

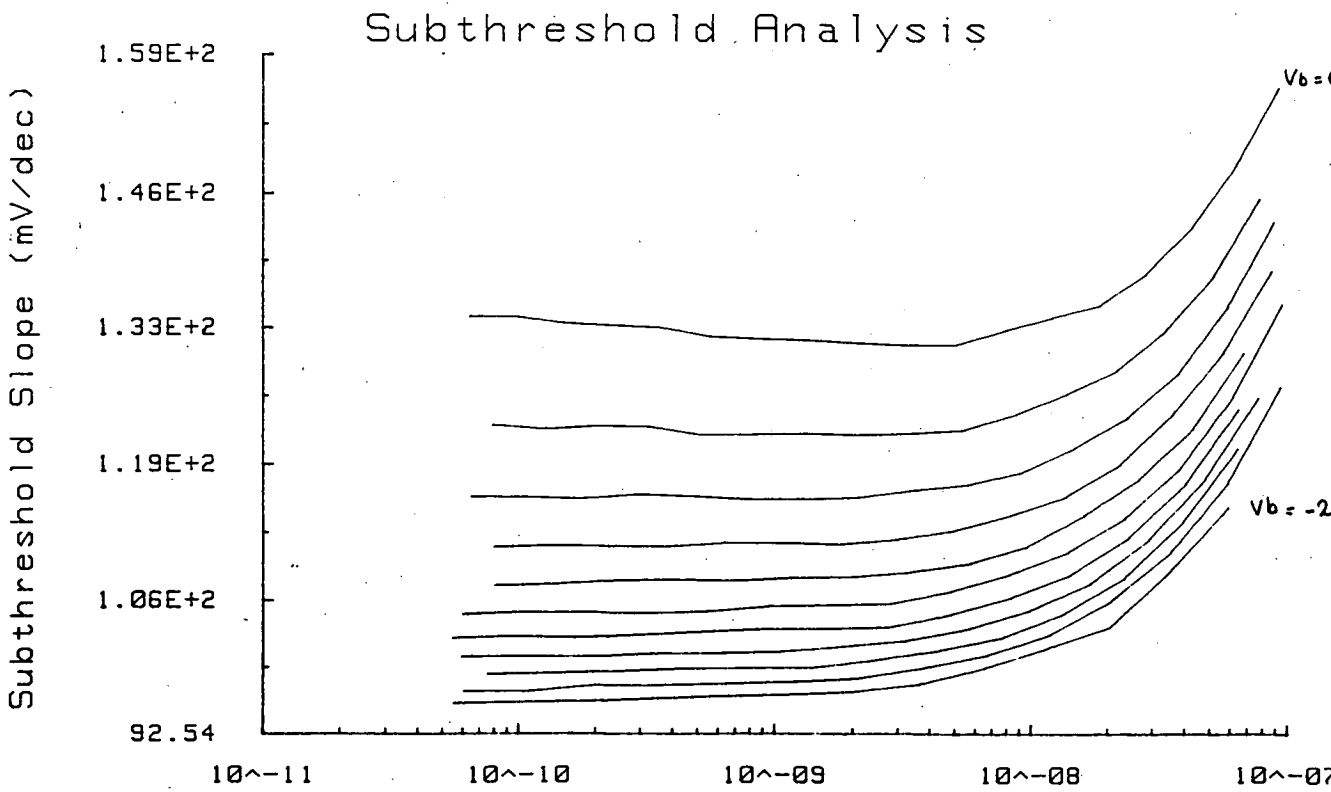


Figure 6.3.21 Log (I_d (A))

Sample 1 Boron Implants 3E11 at 40keV and 8E11 at 180keV

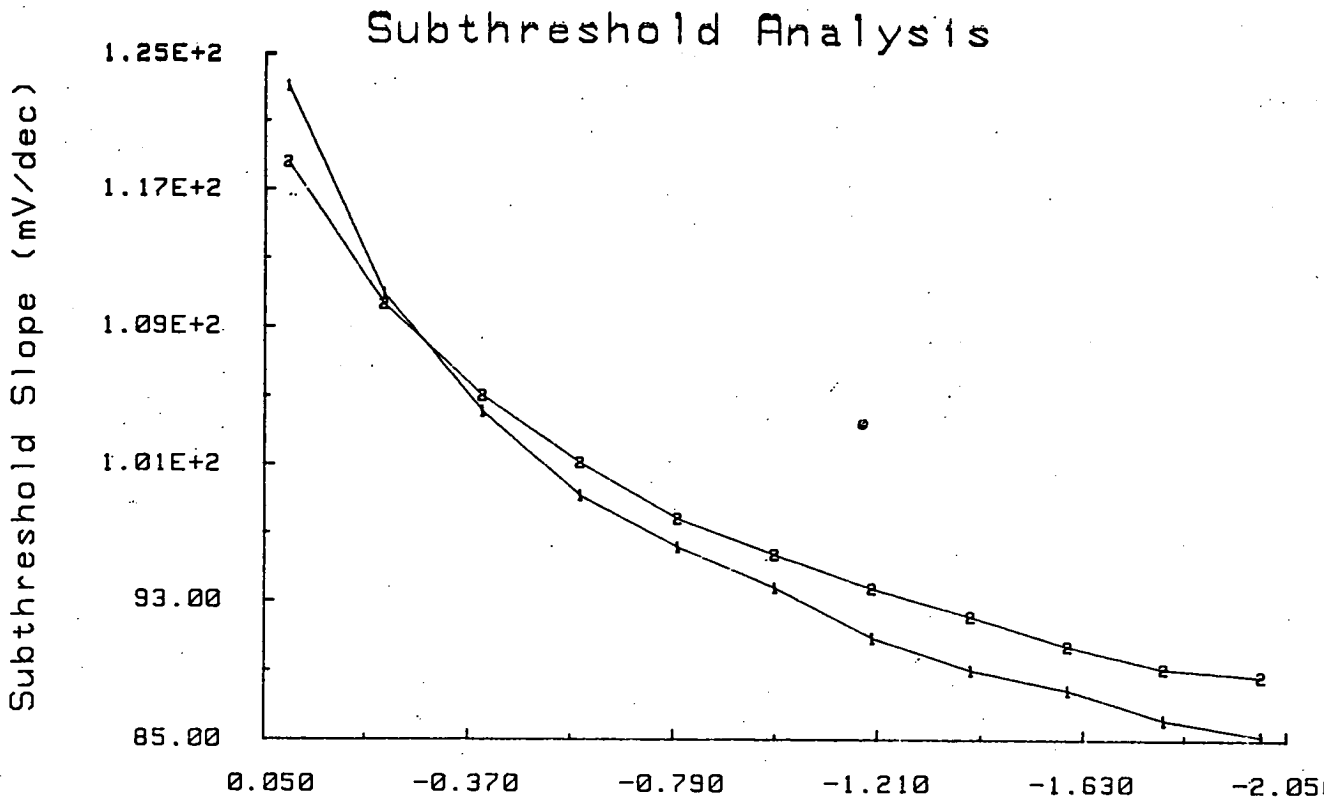


Figure 6.3.22 Substrate Bias (V)

Sample 2 Boron Implants 4E11 at 40keV and 4E11 at 320keV

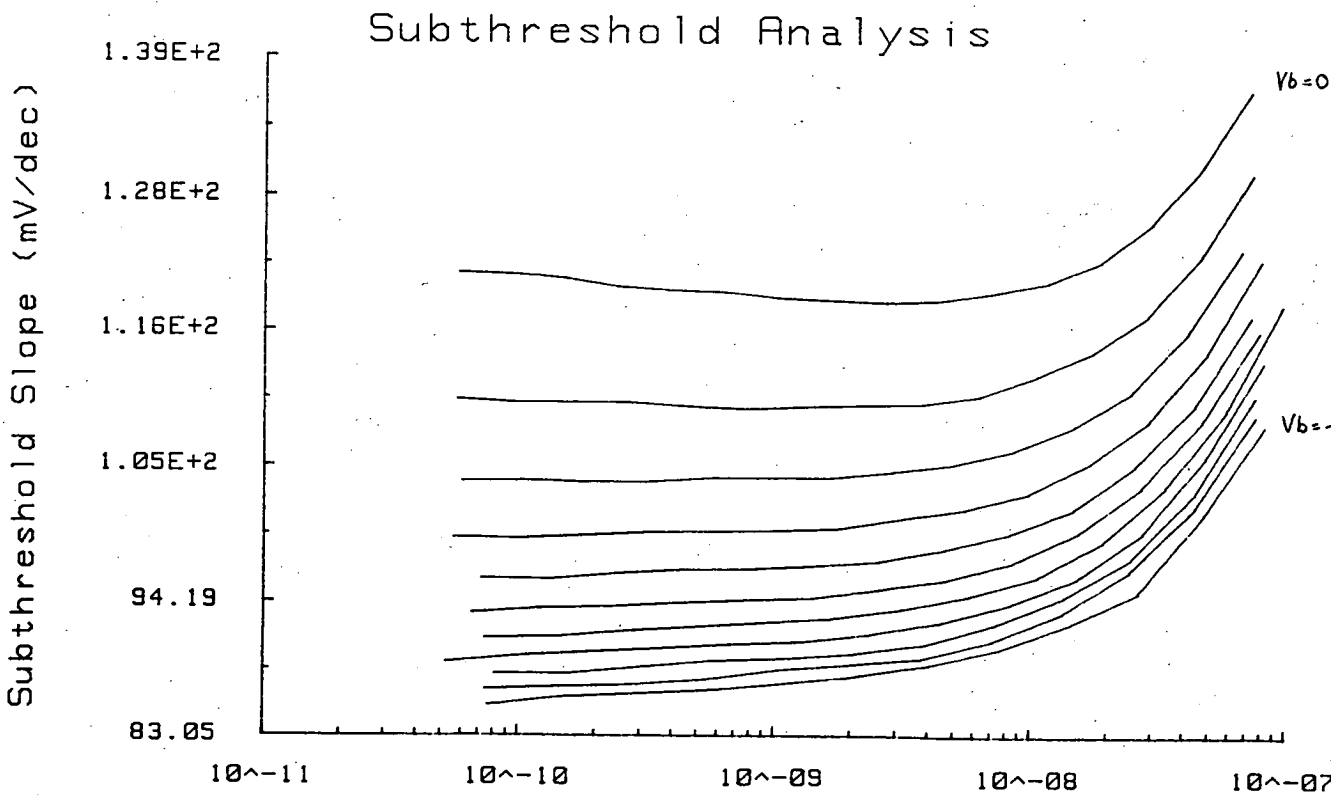


Figure 6.3.23 Log (I_d (A))

Sample 2 Boron Implants 4E11 at 40keV and 4E11 at 320keV

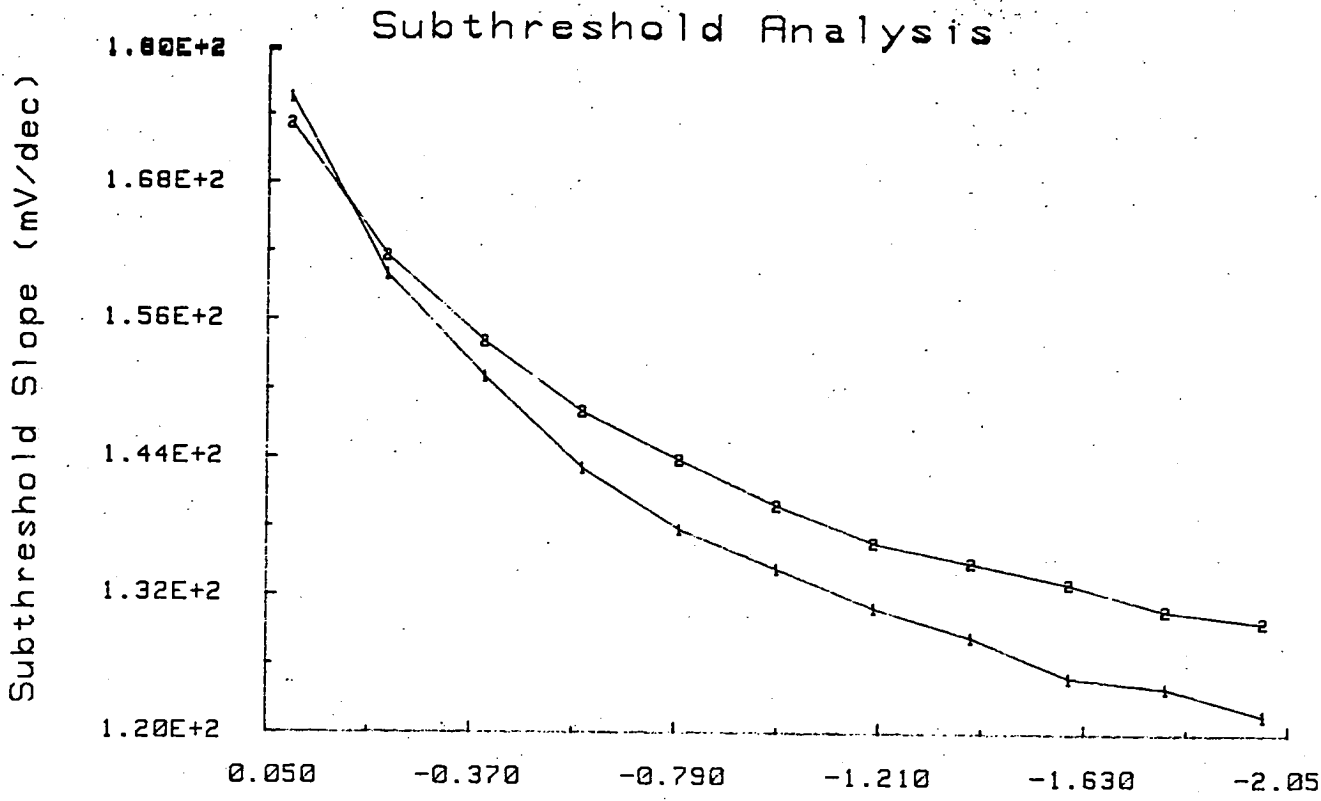


Figure 6.3.24 Substrate Bias (V)

Sample 3 Boron Implant, 5E12 at 180keV

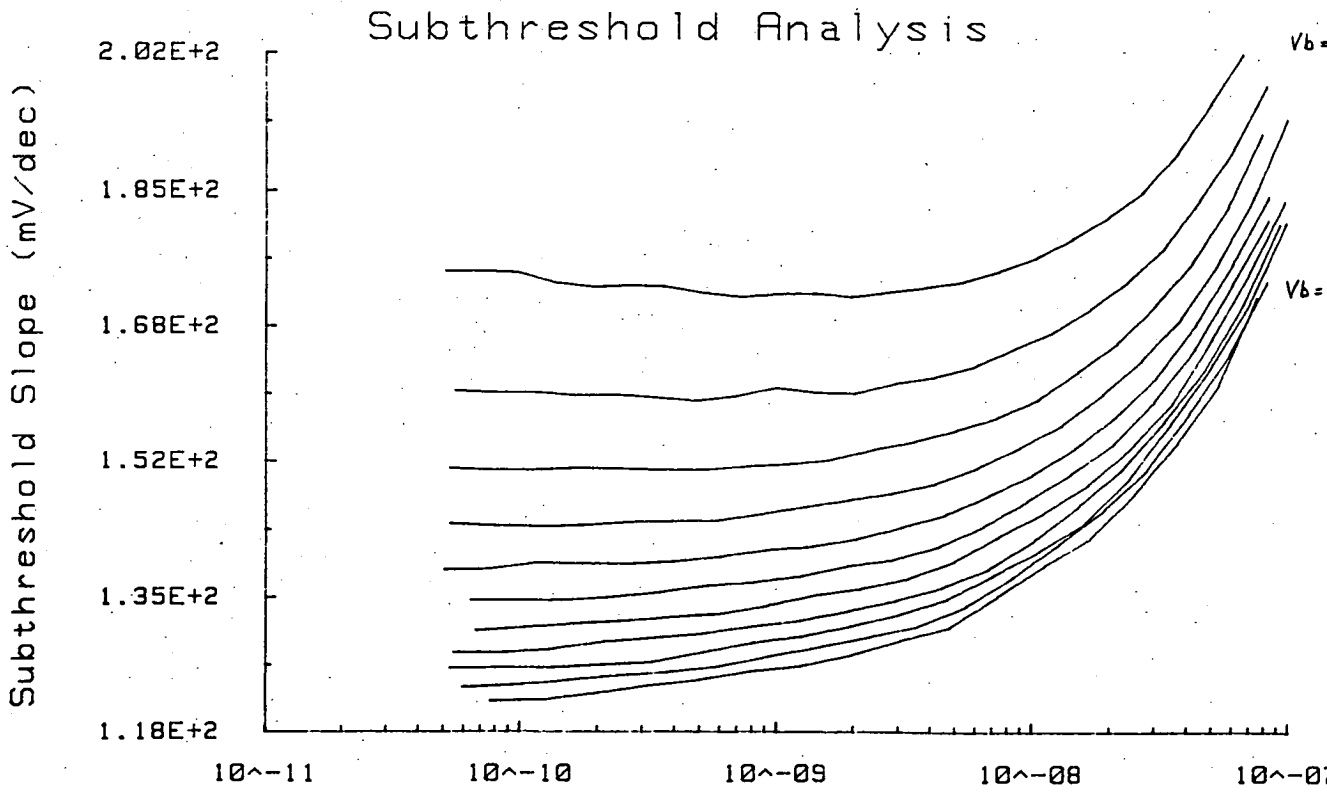


Figure 6.3.25 Log (Id (A))

Sample 3 Boron Implant, 5E12 at 180keV

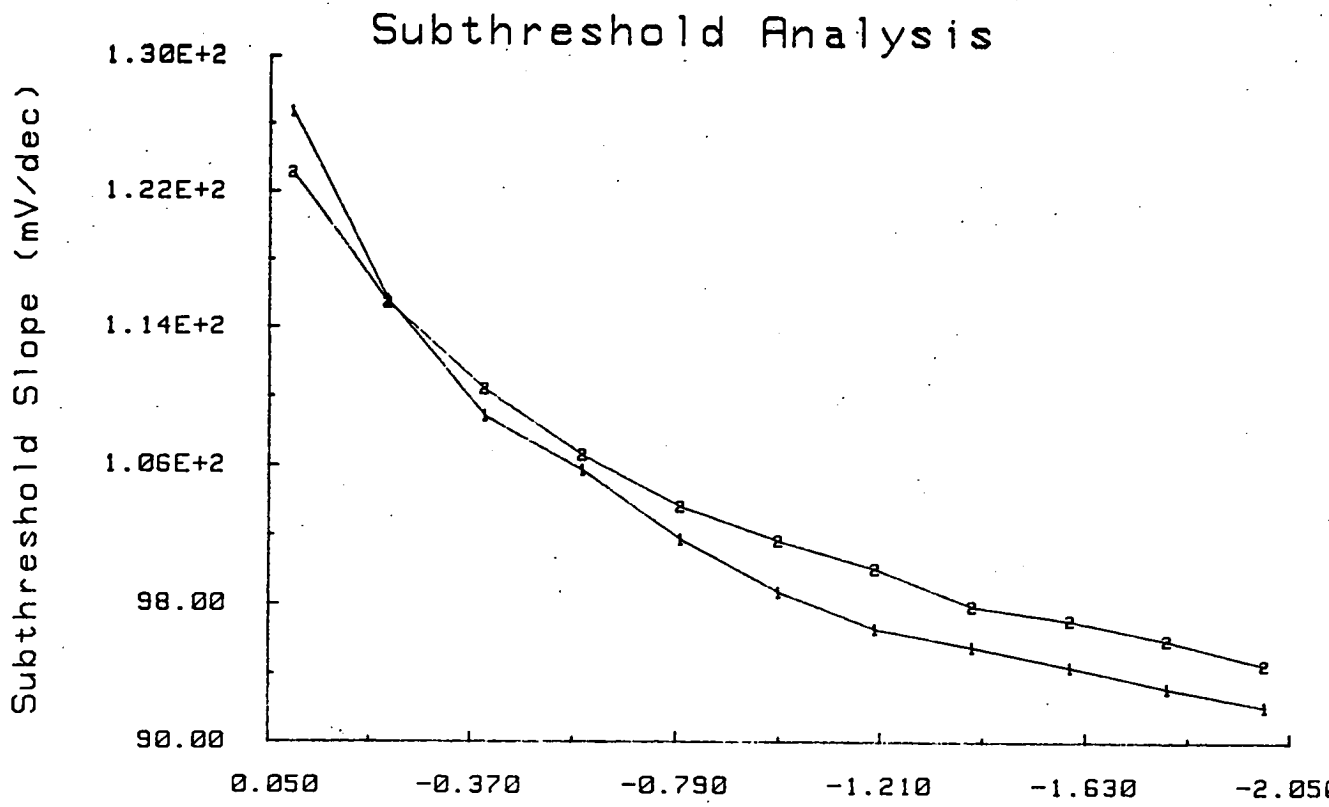


Figure 6.3.26 Substrate Bias (V)

Sample 5 Boron Implants 4E11 at 40keV and 1E12 at 320keV

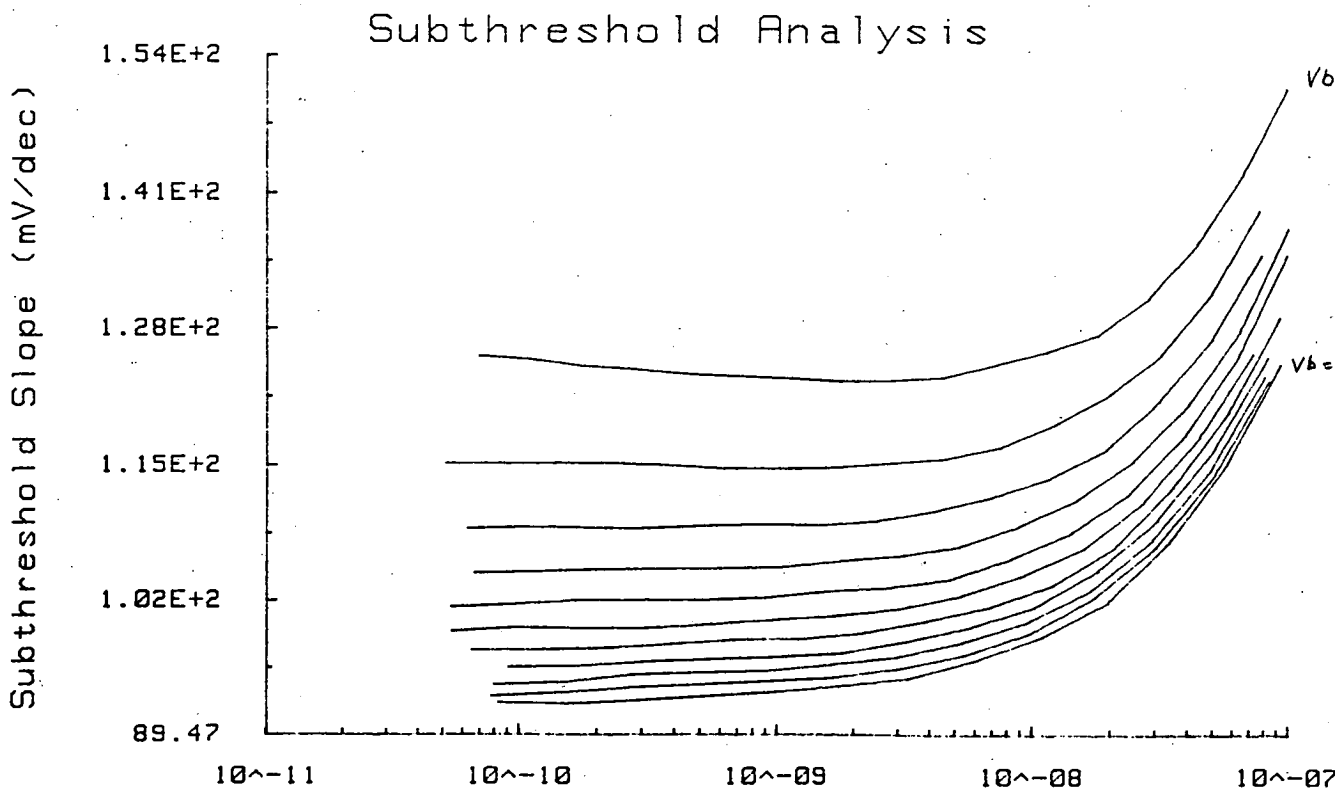


Figure 6.3.27 Log (Id (A))

Sample 5 Boron Implants 4E11 at 40keV and 1E12 at 320keV

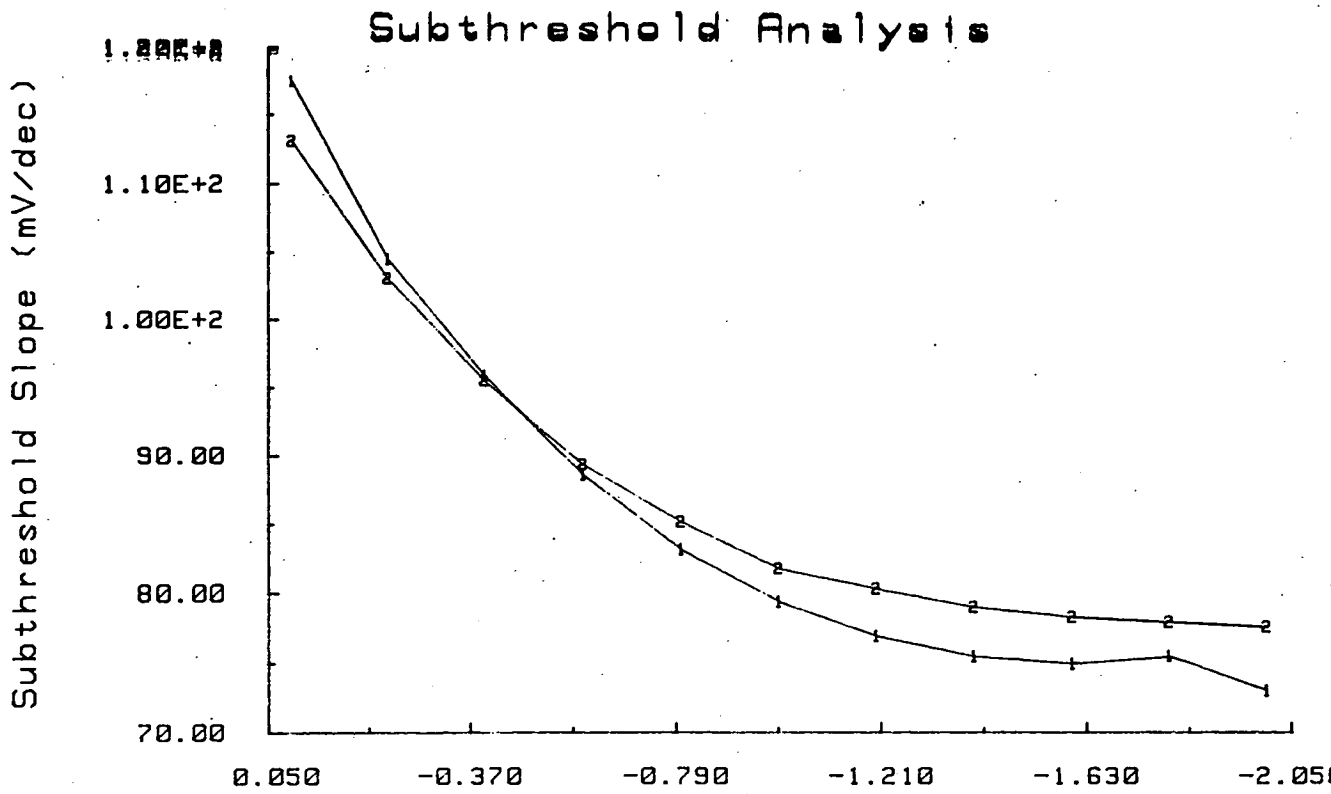


Figure 6.3.28 Substrate Bias (V)

Sample 6 Boron Implant 4E11 at 40keV

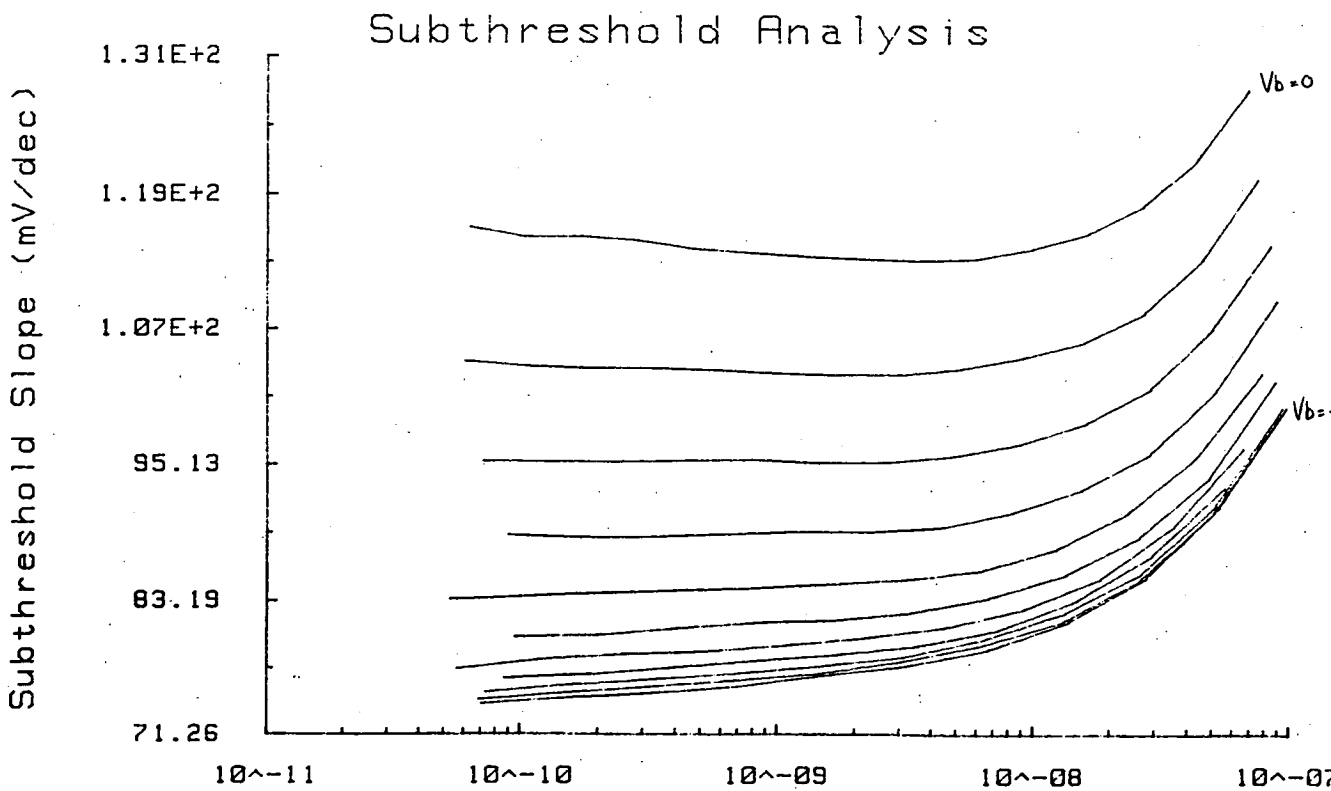


Figure 6.3.29 Log (I_d (A))

Sample 6 Boron Implant 4E11 at 40keV

The leakage current in the fourth sample means that subthreshold slopes cannot be determined at low current levels and that at higher current levels the slope may also be affected. Hence further analysis has not been carried out on the fourth sample.

6.4 Discussion and Conclusions

Table 6.4.1 lists some of the information provided in the plots of subthreshold slope against substrate bias in Section 6.3. For each sample, a brief description of the profile is provided and the values of S at $V_b = 0$ and $V_b = -2$ for current levels of $10^{-10.5}A$ and $10^{-8.5}A$.

The first conclusion to be drawn from these figures is that S increases as the impurity concentration increases. This is in agreement with the theory which is illustrated in figure 6.2.5. Sample 3 has an average concentration of around $5 \times 10^{16} \text{ cm}^{-3}$ to a depth of $0.8 \mu\text{m}$ and it has the highest value of S ; around $175.5 \text{ mV dec}^{-1}$ at $V_b = 0$ and $I_d = 10^{-10.5}A$. Although the standard NMOS device (specimen 6) has a surface concentration of $1.4 \times 10^{16} \text{ cm}^{-3}$, this concentration falls off very rapidly beyond $0.2 \mu\text{m}$ and consequently, it exhibits the lowest S of $117.4 \text{ mV dec}^{-1}$ under the same conditions.

This result is important in small geometry processing where a heavy implant is used to prevent punchthrough. Since capacitive nodes are smaller in small geometry circuits and the device turns off more slowly so that leakage currents are higher, the danger of a node discharging when it isn't intended to is greater. Therefore, consideration must be given to S when designing the structure of small geometry devices.

There is a gradual reduction of S as V_b varies from 0 to $-2V$. At $V_b = 0$, S at $10^{-10.5}A$ is greater than at $10^{-8.5}A$ whereas the opposite is true at $V_b = -2V$. The greatest variation occurs near $V_b = 0$ and it is in this region where the curves measured at different current levels cross. This is in accordance with the theoretical variation in figure 6.2.4.

The standard device (number 6) changes by the largest percentage as the

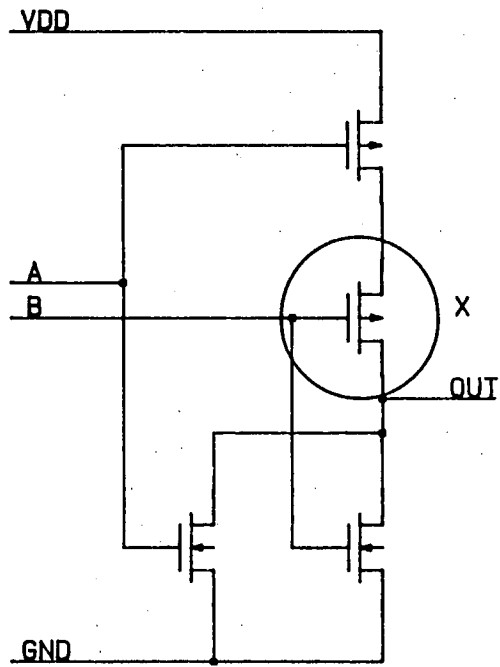
Table 6.4.1 Subthreshold Slope ($mV.dec^{-1}$) Variation with V_b					
Current Level : $I_d = 10^{-10.5}A$					
Sample Number	Approximate Profile	Substrate Bias (V)		Change ($mV.dec^{-1}$)	Percentage Change
		0	-2		
1	1.5E16 to 0.5 μm	132.6	95.4	37.2	28.0
2	1E16 to 0.8 μm	123.0	85.2	37.8	30.7
3	5E16 to 0.8 μm	175.5	121.6	53.9	30.7
5	1.4E16 to 0.8 μm	126.7	92.1	34.6	27.3
6	1.4E16 to 0.2 μm	117.4	73.0	44.4	37.8

Table 6.4.1 Subthreshold Slope ($mV.dec^{-1}$) Variation with V_b (cont)					
Current Level : $I_d = 10^{-8.5}A$					
Sample Number	Approximate Profile	Substrate Bias (V)		Change ($mV.dec^{-1}$)	Percentage Change
		0	-2		
1	1.5E16 to 0.5 μm	130.9	97.1	33.8	25.8
2	1E16 to 0.8 μm	118.5	88.6	29.9	25.2
3	5E16 to 0.8 μm	173.2	129.7	43.5	25.1
5	1.4E16 to 0.8 μm	123.1	94.5	28.6	23.2
6	1.4E16 to 0.2 μm	113.0	77.5	35.5	31.4

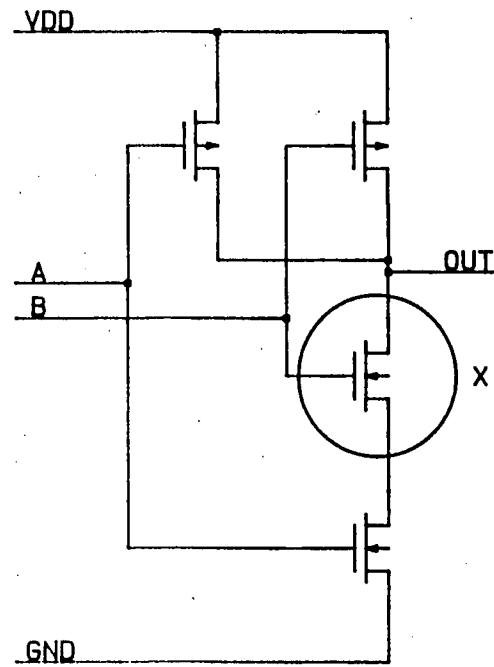
substrate bias is varied from 0 to -2V and most rapidly as the substrate bias leaves 0V. For biases of between -1V and -2V, the swing is dependent on the ratio of $C_{ox}:C_d$. Since the channel is lightly doped, the depletion region quickly becomes large, further increases in substrate bias do not have such a large effect on $C_{ox}:C_d$. For all other samples, which are more heavily doped, the variation in S continues as the substrate bias approaches -2V.

The figures showing the variation of S with current level at different substrate voltages indicate that S only varies slightly until threshold is neared. The phenomenon pointed out above, where the swing decreases as current increases at $V_b = 0$ and S increases as the current decreases at $V_b = -2V$ can be seen.

The importance of the variation of subthreshold swing with substrate bias can be seen in simple CMOS NOR and NAND gates (figure 6.4.1). In these gates the devices which are marked with X are operating with non-zero source to substrate voltages and correspondingly will turn off more quickly. This implies that the simulation tools, which assume that subthreshold slope is irrespective of substrate bias will predict higher leakage currents and slower switching operation than is actually the case. Although these are second order effects, in small geometry circuits where the value of S has been pushed up through the higher substrate doping, it is important when pursuing optimum speed and minimum power dissipation.



(i) CMOS NOR



(ii) CMOS NAND

Figure 6.4.1 Devices at non-zero V_b in CMOS gates

Chapter 7 : Conclusions

A set of physical parameter extraction algorithms for the SPICE 2 levels 1 and 3 MOSFET models have been derived. Each parameter is extracted separately and from a particular device characteristic. A program called PARAMEX, consisting of approximately 5000 lines of advanced BASIC, measures device characteristics, extracts parameters using the algorithms mentioned above and also simulates transistor operation. No commercially available software uses physical parameter extraction and copies of PARAMEX have been supplied to INMOS, Motorola, Jet Propulsion Laboratories and Rutherford Laboratories. Accurate simulation has been achieved for NMOS and PMOS enhancement devices, NMOS depletion devices and devices with drawn dimensions of $1.5\mu\text{m}$. Further research which has used PARAMEX includes minimising the number of measurement points, making parameter measurements in parallel and measuring parameters through on-chip switches.

Parameters were measured on devices whose dimensions ranged from $1\mu\text{m} \times 1\mu\text{m}$ to $30\mu\text{m} \times 30\mu\text{m}$. The values of V_{to} , γ and μ_o showed significant variation with device size and changed more rapidly as channel lengths became shorter. From the extraction graphs, it can be seen that the short channel effects modelled by v_{max} , η and κ have most influence on the shortest channels and little influence on devices of length greater than $3\mu\text{m}$. For a $2\mu\text{m}$ or $3\mu\text{m}$ process, parameters for only three or four different lengths would be required to provide accurate simulation of all devices and hence a lookup table was suggested. Since parameters change more quickly with length in shorter channels, the CASMOS approach of empirically modelling parameter variations with geometry may become worthwhile for submicron processes. Proposals were made for deriving a single set of parameters capable of providing reasonable accuracy for all geometries of device.

On each of four wafers, 2 NMOS and 2 PMOS, parameters were measured at up to 170 sites and from these best and worst case parameter sets were obtained by taking mean $\pm 3\sigma$ values for each parameter. However for all four wafers, this produced a much wider variation in transistor characteristics than was found by looking for maximum and minimum currents. This was due to correlation between parameters and the correlation was investigated and explained. The rise time and switching voltage of a CMOS inverter were simulated with different best and worst case parameter sets in

order to illustrate the effect of the different parameter sets on circuit operation. A method for deriving realistic best and worst case parameter sets by only making a limited number of measurements was proposed.

Parameters were evaluated for two batches of wafers in which various steps in the fabrication process had been varied. In the first batch of 24 CMOS wafers, parameters for each wafer were found for both NMOS and PMOS by measuring parameters at 10 sites and averaging them. Only 5 sites from each wafer were averaged in the second batch of small geometry NMOS wafers. The correlation between process variations and parameters was calculated and the most significant correlations were θ and N_f , with oxide thickness and V_{to} and γ with implant dose. In the short channel NMOS, the second order parameters, η and κ exhibited high correlation with channel doping. The very high correlation between Δ_w and the use of the SEPOX process indicated that it was successful in reducing the bird's beak at the edge of the field oxide. Further experiments were proposed in which wafers were fabricated with only one process step being varied in order to produce more specific conclusions for use in process control and process simulation.

An expression for the subthreshold swing S of a uniformly doped device was derived by standard techniques. Using SUPREM 2, implant recipes were determined for producing six devices with different channel profiles. The subthreshold characteristics of samples fabricated according to these recipes were measured and by numerical methods graphs of S :substrate bias and S :transistor current were plotted for 5 out of the 6 samples. Subthreshold current was found to increase with channel doping and the consequences with regards leakage current or the occurrence of a spurious state were pointed out. The subthreshold swing was found to vary with substrate bias in accordance with the theory. The effect of swing changing with substrate bias, which is not accounted for in the SPICE 2 level 3 model, was discussed with reference to CMOS NAND and NOR gates.

Bibliography

References

1. D. Kahng, "A Historical Perspective on the Development of MOS Transistors and Related Devices," *IEEE Transactions on Electron Devices*, vol. ED-23, pp. 655-657, July 1976.
2. C.T. Sah, "Characteristics of the Metal-Oxide-Semiconductor Transistors," *IEEE Transactions on Electron Devices*, vol. ED-11, pp. 324-345, July 1964.
3. J.T. Wallmark, "The Field-Effect Transistor - an old device with new promise," *IEEE Spectrum*, vol. 1, pp. 182-191, March 1964.
4. W. Shockley and G.L. Pearson, "Modulation of Conductance of Thin Films of Semiconductors by Surface Charges," *Physical Review*, vol. 74, pp. 232-233, June 1948.
5. W.L. Brown, "n-Type Surface Conductivity on p-Type Germanium," *Physical Review*, vol. 91, pp. 518-527, August 1953.
6. M.M. Atalla and Scheibner, "Stabilisation of Silicon Surfaces by Thermally Grown Oxides," *Bell System Technical Journal*, pp. 749-783, May 1959.
7. E.H. Snow, A.S. Grove, B.E. Deal, and C.T. Sah, "Ion Transport Phenomena in Insulating Films," *Journal of Applied Physics*, vol. 36, pp. 1664-, May 1965.
8. J.C. Sarace, R.E. Kerwin, D.L. Klein, and R. Edwards, "Metal-Oxide-Silicon Field Effect Transistors with Self-Aligned Gates," *Solid State Electronics*, vol. 11, pp. 653-660, 1968.
9. "Introduction," in *VLSI Technology*, ed. S.M. Sze, pp. 1-7, McGraw-Hill, Tokyo, 1983.
10. M.H. Woods, "The Scaling of VLSI; The reliability Impact on Technology," *Dynamite course at Leuven*, pp. 41-56, June 1986.
11. J.M. Robertson, "MOS Scaling Options," *MSc Course Notes, Edinburgh University*, 1984.
12. G. Declerck, K. De Meyer, and L. Dupas, "MOS Technology for VLSI," *Microelectronics Reliability*, vol. 24, pp. 205-225, 1984.
13. Robert H. Dennard, Fritz H. Gaensslen, Hwa-Nien Yu, V. Leo Rideout, Ernest Bassous, and Andre LeBlanc, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *IEEE J. of Solid-State Circuits*, vol. SC-9, no. 5, pp.

256-267, October 1974.

14. J.R. Brews, W. Fichtner, E.H. Nicollian, and S.M. Sze, "Generalised Guide for MOSFET Miniaturization," *IEEE Electron Device Letters*, vol. EDL-1, pp. 2-4, Jan 1980.
15. A. Silburt, "Linking MOSFET Design and Fabrication," *Microelectronics Manufacturing and Testing*, vol. 9, pp. 1,13-15, April 1986.
16. K. Nichols, "Simulation - Inside out," *Silicon Design*, vol. 2, pp. 22-24, November 1985.
17. Ebrahim Khalily, Peter H. Decher, and Darrell A. Teegarden, "TECAP2: An Interactive Device Characterisation and Model Development System," *Hewlett-Packard Internal Paper*.
18. Sunlin Chou and Carl Simonsen, "Chip Voltage: Why Less is Better," *IEEE Spectrum*, vol. 24, pp. 39-43, April 1987.
19. P.R. Gray and R.G. Meyer, "MOS Operational Amplifier Design - A Tutorial Overview," *IEEE J. of Solid-State Circuits*, vol. SC-17, no. 6, pp. 969-982, December 1982.
20. S.M. Sze, "Subthreshold Region in MOSFETs," in *Physics of Semiconductor Devices*, pp. 446-448, John Wiley and Sons Inc., New York, 1981.
21. John R. Brews, "Subthreshold Behaviour of Uniformly and Nonuniformly Doped Long-Channel MOSFET," *IEEE Trans. Electron Devices*, vol. ED-26, no. 9, pp. 1282-1291, September 1979.
22. Andrei Vladimirescu and Sally Liu, "The Simulation of MOS Integrated Circuits using SPICE2," UCB/ERL M80/7, February 1980.
23. John R. Brews, "A Charge-Sheet Model of the MOSFET," *Solid-State Electronics*, vol. 21, pp. 345-355, June 1977.
24. R. E. Oakley and R. J. Hocking, "CASMOS - an accurate MOS model with geometry-dependent parameters:1," *IEE Proceedings*, vol. 128 pt 1, no. 6, pp. 239-247, December 1981.
25. G. T. Wright, "Simple and Continuous MOSFET Models for the CAD of VLSI," *IEE Proceedings*, vol. 132 pt 1, no. 4, pp. 187-194, August 1985.
26. A.S. Grove, "Elements of Semiconductor Physics," in *Physics and Technology of Semiconductor Devices*, pp. 91-105, John Wiley and Sons Inc., New York, 1967.

27. S.M. Sze, "Physics and Properties of Semiconductors - A Resume," in *Physics of Semiconductor Devices*, pp. 12-27, John Wiley and Sons Inc., New York, 1981.
28. S.M. Sze, "Ideal MIS Diode," in *Physics of Semiconductor Devices*, pp. 362-369, John Wiley and Sons Inc., New York, 1981.
29. Andrew S. Grove, "The Ideal MIS Structure," in *Physics and Technology of Semiconductor Devices*, pp. 271-285, John Wiley and Sons Inc., New York, 1967.
30. J.R. Brews, "Physics of the MOS Transistor - The MOS Capacitor," in *Applied Solid State Science Supplement 2A*, ed. D. Kahng, pp. 3-11, Academic Press Inc., New York, 1981.
31. A.S. Grove, "Space Charge Region for Step Junctions," in *Physics and Technology of Semiconductor Devices*, pp. 153-161, John Wiley and Sons Inc., New York, 1967.
32. H. K. J. Ihantola and J. L. Moll, "Design Theory of a Surface Field-Effect Transistor," *Solid-State Electronics*, vol. 7, pp. 423-430, 1964.
33. S.M. Sze, "MOSFET - Basic Device Characteristics," in *Physics of Semiconductor Devices*, pp. 431-443, John Wiley and Sons Inc., New York, 1981.
34. J. Mavor, M.A. Jack, and P.B. Denyer, "MOS Transistor Theory and Inverter Circuit," in *Introduction to MOS LSI Design*, pp. 18-36, Addison-Wesley, London, 1983.
35. A.S. Grove, "Surface Field-Effect Transistor," in *Physics and Technology of Semiconductor Devices*, pp. 317-327, John Wiley and Sons Inc., New York, 1967.
36. H.C. Pao and C.T. Sah, "Effects of Diffusion Current on Characteristics of Metal-Oxide (Insulator)-Semiconductor Transistors," *Solid-State Electronics*, vol. 9, pp. 927-937, 1966.
37. J.R. Brews, "Physics of the MOS Transistor - The MOSFET," in *Applied Solid State Science Supplement 2A*, ed. D. Kahng, pp. 15-29, Academic Press Inc., New York, 1981.
38. Prof M. Martelli and Prof E. Cumberbatch et al, "Modelling Short Channel MOSFETs for use in VLSI," *Jet Propulsion Laboratory Report*, May 1986.
39. G. T. Wright and H. M. Gaffur, "Pre-Processor Modelling of Parameter and Geometry Dependence of Short and Narrow MOSFET's for VLSI Circuit Simulation, Optimization, and Statistics with SPICE," *IEEE Trans. Electron Devices*, vol. ED-32, no. 7, pp. 1240-1245, July 1985.

40. G. Baum and H. Beneking, "Drift Velocity Saturation in MOS Transistors," *IEEE Trans. Electron Devices*, pp. 481-482, June 1970.
41. D. Frohman-Bentchkowsky and A. S. Grove, "Conductance of MOS Transistors in Saturation," *IEEE Trans. Electron Devices*, vol. ED-16, no. 1, pp. 108-113, January 1969.
42. Siegfried Selberherr, Alfred Schutz, and Hans Wolfgang Potzl, "MINIMOS - A Two-Dimensional MOS Transistor Analyzer," *IEEE J. of Solid-State Circuits*, vol. SC-15, no. 4, pp. 605-615, August 1980.
43. A.M. Gundlach, "EMF NMOS Process," *EMF Internal Report, Edinburgh University*, May 1983.
44. J.M. Robertson, "Microelectronic Circuit Fabrication," in *The Impact of Microelectronics Technology*, ed. M. Jack, pp. 12-25, Edinburgh University Press, 1982.
45. J.A. Appels, E. Kooi, M.M. Paffen, J.J.H. Schatorje, and W.H.C.G. Verkuylen, "Local Oxidation of Silicon and its Application in Semiconductor-Device Technology," *Philips Research Reports*, vol. 25, pp. 118-132, March 1970.
46. J.T.M. Stevenson, "Lithography," in *SERC School on Microfabrication*, Edinburgh, June 1985.
47. D.A. McGillis, "Lithography," in *VLSI Technology*, ed. S.M. Sze, pp. 267-300, McGraw-Hill, Tokyo, 1983.
48. R. Holwill, "Implantation," in *SERC School on Microfabrication*, Edinburgh, June 1985.
49. T.E. Seidel, "Ion Implantation," in *VLSI Technology*, ed. S.M. Sze, pp. 219-264, McGraw-Hill, Tokyo, 1983.
50. N. Matsukawa, H. Nozawa, and J. Matsunaga, "Selective Polysilicon Oxidation Technology for VLSI Isolation," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 561-567, April 1982.
51. D. Burkman, "Optimizing the Cleaning Procedure for Silicon Wafers Prior to High Temperature Operations," *Semiconductor International*, pp. 103-116, July 1981.
52. W. Kern and D.A. Puotinen, "Cleaning Solutions Based on Hydrogen Peroxide for use in Silicon Semiconductor Technology," *RCA Review*, pp. 187-206, June 1970.

53. C.P. Ho, J.D. Plummer, S.E. Hansen, and R.W. Dutton, "VLSI Process Modelling - SUPREM III," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 1438-1454, November 1983.
54. D.A. Antoniadis and R.W. Dutton, "Models for Computer Simulation of Complete IC Fabrication Process," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 490-500, April 1979.
55. J. Anguita, "OSIRIS: Two Dimensional Process Simulation - Its Implementation at the EMF," *EMF Internal Report, Edinburgh University*, April 1986.
56. S.E. Hansen, D.A. Antoniadis, and R.W. Dutton, *SUPREM II User's Manual*, Stanford University, July 1978.
57. Dezsoe Takacs, Wolfgang Muller, and Ulrich Schwabe, "Electrical Measurement of Feature Sizes in MOS Si² -Gate VLSI Technology," *IEEE Trans. Electron Devices*, vol. ED-27, no. 8, pp. 1368-1373.
58. K. L. Peng, S. Y. Oh, M. A. Afromowitz, and J. L. Moll, "Basic Parameter Measurement and Channel Broadening Effect in the Submicrometer MOSFET," *IEEE Trans. Electron Devices*, vol. EDL-5, no. 11, pp. 473-475, November 1984.
59. Keithley Instruments Inc., *MOSFIT Parameter Extraction Software*, Cleveland, Ohio.
60. W. Maes, K.M. De Meyer, and L.H. Dupas, "SIMPAN: A Versatile Technology Independent Parameter Extraction Program Using a new Optimised Fit Strategy," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, no. 2, pp. 320-325, April 1986.
61. Hewlett-Packard, *HP 9445A TECAP Software*, Palo Alto, California.
62. A. Gribben, A.J. Walton, and J.M. Robertson, "Accurate Physical Parameter Extraction for Small Geometry Devices," *Semiconductor International Conference Proceedings*, pp. 186-202, Birmingham, 1986.
63. A. Gribben, A.J. Walton, and J.M. Robertson, "Parametric Testing to Link Design and Fabrication," *IEE Colloquium on Testing and Inspection of Electronic Components and Circuits*, pp. 3/1-3/3, London, 1987.
64. A.V. Ferris-Prabhu, L.D. Smith, H.A. Bonges, and J.K. Paulsen, "Radial Yield Variations in Semiconductor Wafers," *IEEE Circuits and Devices Magazine*, vol. 3, no. 2, pp. 42-47, March 1987.
65. Jacob Millman, "MOSFET Inverters," in *Microelectronics: Digital and Analog Circuits and Systems*, pp. 253-266, McGraw-Hill Inc., Tokyo, 1979.

66. H. Taub and D. Schilling, "MOS Gates," in *Digital Integrated Electronics*, pp. 262-273, McGraw-Hill, Tokyo, 1977.
67. P. Tuohy, A. Gribben, A.J. Walton, and J.M. Robertson, "Realistic Worst-case Parameters for Circuit Simulation," *IEE Proceedings*, vol. 134 Pt. 1, no. 5, pp. 137-140, October 1987.
68. A. Gribben and A.J. Walton, "A Review of Parametric Testing," *Semiconductor International Conference Proceedings*, pp. 39-63, Birmingham, 1987.
69. W.M. Sansen, "Design of Analog CMOS Building Blocks," *Crest Vacation School*, Edinburgh, April 1984.
70. J. M. Shannon, "D.C. Measurement of the Space Charge Capacitance and Impurity Profile beneath the Gate of an MOST," *Solid-State Electronics*, vol. 14, pp. 1099-1106, 1971.
71. M.G. Buehler, "Dopant Profiles Determined from Enhancement-mode MOSFET dc measurements," *Applied Physics Letters*, vol. 31, no. 12, pp. 848-850, December 1977.
72. M.G. Buehler, "The D-C MOSFET Dopant Profile Method," *J. Electrochem. Soc. : Solid-State Science and Technology*, vol. 127, no. 3, pp. 701-704, March 1980.
73. A. L. Silburt, A. R. Boothroyd, and M. DiGiovanni, *Automated Parameter Extraction and Modelling of the MOSFET Below Threshold*, October 1985.
74. Stanley C. Lennox and Mary Chadwick, "Chapter 14 : Numerical Methods II," in *Mathematics for Engineers and Applied Scientists*, pp. 354-391, Heinemann, London, 1979.

APPENDIX A - SPICE MOSFET Models

A1 Level 1 Model

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad A1.1$$

$$W = W_m - 2\Delta_w \quad A1.2$$

$$L = L_m - 2L_d \quad A1.3$$

$$V_{th} = V_{to} + \gamma |V_b|^{1/2} \quad A1.4$$

$$\mu_s = \frac{\mu_o}{1 + \theta(V_g - V_{th})} \quad A1.5$$

Subthreshold Region $V_g < V_{th}$

$$I_d = 0 \quad A1.6$$

Linear Region $V_g > V_{th}$ and $V_g > V_{th} + V_d$

$$I_d = \frac{W}{L} \mu_s C_{ox} \left(V_g - V_{th} - \frac{V_d}{2} \right) V_d \quad A1.7$$

Saturation Region $V_g > V_{th}$ and $V_g < V_{th} + V_d$

$$I_d = \frac{W}{L} \mu_s C_{ox} \frac{(V_g - V_{th})^2}{2} \quad A1.8$$

A2 Level 3 Model

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad \text{A2.1}$$

$$L = L_m - 2 L_d \quad \text{A2.2a}$$

$$W = W_m - 2 \Delta_w \quad \text{A2.2b}$$

$$N_{sub} = \frac{\gamma^2}{2 q \epsilon_{si}} C_{ox}^2 \quad \text{A2.3}$$

$$X_d = \left[\frac{2 \epsilon_{si}}{q N_{sub}} \right]^{\frac{1}{2}} \quad \text{A2.4}$$

$$2\phi_b = 2 \frac{k T}{q} \ln \left[\frac{N_{sub}}{n_i} \right] \quad \text{A2.5}$$

$$V_{fb} = V_{io} - 2\phi_b - \gamma |2\phi_b|^{\frac{1}{2}} \quad \text{A2.6}$$

if $V_b \leq 0$ then

$$Phibs = 2\phi_b - V_b \quad \text{A2.7a}$$

else

$$Phibs = \frac{2\phi_b}{\left[1 + \frac{V_b}{4 \phi_b} \right]^2} \quad \text{A2.7b}$$

end if

$$Sqphbs = Phibs^{\frac{1}{2}}$$

if $x_j \neq 0$ and $X_d \neq 0$ then

$$W_{ps} = X_d Sqphbs \quad A2.8$$

$$\frac{W_c}{x_j} = D0 + D1 \frac{W_{ps}}{x_j} + D2 \left[\frac{W_{ps}}{x_j} \right]^2 \quad A2.9$$

where

$$D0 = 0.0631353$$

$$D1 = 0.8013292$$

$$D2 = -0.01110777$$

$$F_s = 1 - \frac{x_j}{L} \left[\left[\frac{W_c}{x_j} + \frac{L_d}{x_j} \right] \left[1 - \left[\frac{W_{ps}}{x_j + W_{ps}} \right]^2 \right]^{\frac{1}{2}} - \frac{L_d}{x_j} \right] \quad A2.10$$

else

$$W_c = 5E-8$$

$$F_s = 1$$

end if

$$\gamma_s = \gamma F_s$$

$$F_{bodys} = \frac{\gamma_s}{4 Sqphbs} \quad A2.11$$

$$F_n = \frac{2 \pi \delta \epsilon_{si}}{4 W C_{ox}} \quad A2.12$$

$$F_b = F_{bodys} + F_n \quad A2.13$$

$$\frac{Q_b}{C_{ox}} = \gamma_s S q \phi_{bs} + F_n Phibs \quad A2.14$$

$$\sigma = \frac{\eta C_0}{C_{ox} L^3} \quad A2.15$$

where

$$C_0 = 8.15 \times 10^{-22} F m^{-3}$$

$$V_{bix} = V_{fb} + 2\phi_b - \sigma V_d \quad A2.16$$

$$V_{th} = V_{bix} + \frac{Q_b}{C_{ox}} \quad A2.17$$

$$V_{on} = V_{th}$$

if $N_{fs} \neq 0$ then

$$\frac{C_s}{C_{ox}} = \frac{q N_{fs}}{C_{ox}} \quad A2.18$$

$$\frac{C_d}{C_{ox}} = \frac{Q_b}{2 C_{ox} Phibs} \quad A2.19$$

$$N = 1 + \frac{C_s}{C_{ox}} + \frac{C_d}{C_{ox}} \quad A2.20$$

$$V_{on} = V_{th} + \frac{N k T}{q} \quad A2.21$$

end if

if $N_{fs}=0$ and $V_g < V_{on}$ then

$$I_d = 0$$

else

$$V_{gsx} = \max(V_g, V_{on}) \quad \text{A2.22}$$

$$\mu_s = \frac{\mu_o}{1 + \theta(V_{gsx} - V_{th})} \quad \text{A2.23}$$

if $v_{max} > 0$ then

$$V_{dsat} = \frac{V_{gsx} - V_{th}}{1 + F_b} + \frac{v_{max} L}{\mu_s} - \left[\left[\frac{V_{gsx} - V_{th}}{1 + F_b} \right]^2 + \left[\frac{v_{max} L}{\mu_s} \right]^2 \right]^{\frac{1}{2}} \quad \text{A2.24a}$$

else

$$V_{dsat} = \frac{V_{gsx} - V_{th}}{1 + F_b} \quad \text{A2.24b}$$

end if

$$V_{dsx} = \min(V_d, V_{dsat}) \quad \text{A2.25}$$

$$\text{Beta} = \mu_s C_{ox} \frac{W}{L} \quad \text{A2.26}$$

$$I_d = \text{Beta} V_{dsx} \left[V_{gsx} - V_{th} - \frac{1+F_b}{2} V_{dsx} \right] \quad \text{A2.27}$$

if $v_{max} \neq 0$ then

$$F_{drain} = \frac{1}{1 + \frac{V_{dsx} \mu_s}{L v_{max}}} \quad \text{A2.28}$$

$$\mu_{eff} = \mu_s F_{drain}$$

A2.29

$$I_d = I_{dsat} F_{drain}$$

end if

if $V_d \geq V_{dsat}$ then

if $v_{max} \neq 0$ then

$$I_{dsat} = I_d$$

$$G_{dsat} = I_{dsat} (1 - F_{drain})$$

A2.30

$$E_{max} = \frac{I_{dsat}}{G_{dsat} L}$$

A2.31

$$L_{del} = \left[\left[\frac{X_d^2 E_{max}}{2} \right]^2 + \kappa X_d^2 (V_d - V_{dsat}) \right]^{\frac{1}{2}} - \frac{X_d^2 E_{max}}{2}$$

A2.32a

else

$$L_{del} = [\kappa X_d^2 (V_d - V_{dsat})]^{\frac{1}{2}}$$

A2.32b

end if

if $L_{del} > \frac{1}{2}L$ then

$$L_{del} = L - \frac{L^2}{4 L_{del}}$$

A2.33

end if

$$L_{fact} = \frac{1}{1 - \frac{L_{del}}{L}}$$

A2.34

$$I_d = I_d L_{fact}$$

end if

if $V_g < V_{on}$ then

$$W_{fact} = \exp \left[\frac{q (V_g - V_{on})}{N k T} \right] \quad \text{A2.35}$$

$$I_d = I_d W_{fact}$$

end if

end if

APPENDIX B - OSIRIS Process Simulation Summary

VLSI 2-D Channel Profile

DESCRIPTION OF THE GRID

Depth= 1.0 (microns) Length= 1.3 (microns)
Number of grid points= 20*25

Impurity 1 : BORON

DESCRIPTION OF THE SILICON

Orientation= 100 Concentration= 1.00e+14
Element= b

** 1 **

OXIDE DEPOSITION

Thickness= .025 (microns)

** 2 **

ION IMPLANTATION

Impurity 1 : BORON
Energy= 25. (KeVs) Dose= 5.00e+11 (cm-2)

IMPLANTATION WITH TRANSLUCENT MASK

** 3 **

ION IMPLANTATION

Impurity 1 : BORON

Energy=140. (KeVs) Dose= 2.00e+12 (cm-2)

IMPLANTATION WITH TRANSLUCENT MASK

** 4 **

DIFFUSION OF 1 IMPURITY

Ambient :Inert

Temperature= 925. (degrees C) Time= 20. (min)

Number of iterations= 25

** 5 **

DIFFUSION OF 1 IMPURITY

Ambient :Inert

Temperature= 1000. (degrees C) Time= 15. (min)

Number of iterations= 28

** 6 **

DIFFUSION OF 1 IMPURITY

Ambient :Inert

Temperature= 925. (degrees C) Time= 20. (min)

Number of iterations= 13

** 7 **

DIFFUSION OF 1 IMPURITY

Ambient :Inert

Temperature= 950. (degrees C) Time= 15. (min)

Number of iterations= 11

** 8 **

PHOTORESIST DEPOSITION

Position= r Edge= .50 (microns)

** 9 **

PROFILES HAVE BEEN SAVED IN previous.d

**10 **

ION IMPLANTATION

Impurity 2 : ARSENIC

Energy= 90. (KeVs) Dose= 7.00e+15 (cm-2)

IMPLANTATION WITH OXIDE

**11 **

OXIDE ETCH

Thickness= .025 (microns)

**12 **

DIFFUSION OF 2 IMPURITIES

Ambient :Inert

Temperature= 950. (degrees C) Time= 15. (min)

Number of iterations = 106

**13 **

DIFFUSION OF 2 IMPURITIES

Ambient :Inert

Temperature= 1000. (degrees C) Time= 20. (min)

Number of iterations = 153

**14 **

DIFFUSION OF 2 IMPURITIES

Ambient :Inert

Temperature= 950. (degrees C) Time= 25. (min)

Number of iterations = 56

**15 **

DIFFUSION OF 2 IMPURITIES

Ambient :Inert

Temperature= 925. (degrees C) Time= 20. (min)

Number of iterations = 38

**16 **

DIFFUSION OF 2 IMPURITIES

Ambient :Inert

Temperature= 1000. (degrees C) Time= 15. (min)

Number of iterations = 68

**17 **

DIFFUSION OF 2 IMPURITIES

Ambient :Inert

Temperature= 925. (degrees C) Time= 20. (min)

Number of iterations = 34

**18 **

OSIRIS DONE

APPENDIX C - Paramex User Guide

Contents

Title Page

Contents

Introduction and Hardware Requirements

General Program Outline

Common Blocks

Setpars

Measure

Extract 1 and Extract 3

Store and Read Parameters

Simulate

Quick Start Guide

Introduction and Hardware Requirements

Device models used in circuit simulation packages require certain input parameters in order to link the mathematical equations to particular devices. Commercially available packages use numerical optimisation to evaluate these parameters. In order to preserve the physical meaning of the parameters, this program extracts the parameters one by one and from the same characteristics each time they are measured. No numerical optimisation is used.

The program runs on an HP4062B semiconductor parametric test system. This system contains an HP4141 SMU unit similar to the HP4145 curve tracer, an HP4280A Capacitance Meter and a switching matrix. The system is controlled by an HP9836 computer which uses the BASIC 3.0 operating system. Alternatively, the software has now been translated to run on an HP9817 controller connected to an HP4145 curve tracer. The main omission from the system compared with the HP4062 is the switching matrix so connections must be made manually. Unless a capacitance meter is included, the oxide thickness cannot be measured.

General Program Outline

The program is written in a tree-like fashion and controlled by the user using the softkeys (k0-k9). An overview of all the different menus is provided in figure B1 and each option will be described in subsequent sections. At each stage selecting the 'END' key returns the user to one level above.

The logical order to follow when running the program is:

Setpars → Measure → Extract 1/3 → Simulate

First, 'Setpars' is used to set up the hardware and set up the input parameters. 'Measure' is used to obtain several sets of characteristics from which 'Extract 1' and 'Extract 3' evaluate the level 1 and level 3 d.c. SPICE parameters respectively. Finally 'Simulate' can be used to compare measured and simulated characteristics.

Common Blocks

Numerous common blocks are used in the program. An outline of their contents follows:

Datas - The number of variables, the number of parameter sets and the parameters waiting to be stored on disc.

Dimpins - The pins and dimensions of the device to be measured or simulated.

Extrac - The variable 'fast' which controls whether graphs are to be plotted during extraction.

Flags - Information regarding the plotting of measured and simulated data in the 'Simulate' section.

Graphdet and Simvars - The specifications of the graph, the voltage ranges and the most recently measured and simulated characteristics for the 'Simulate' section of the program.

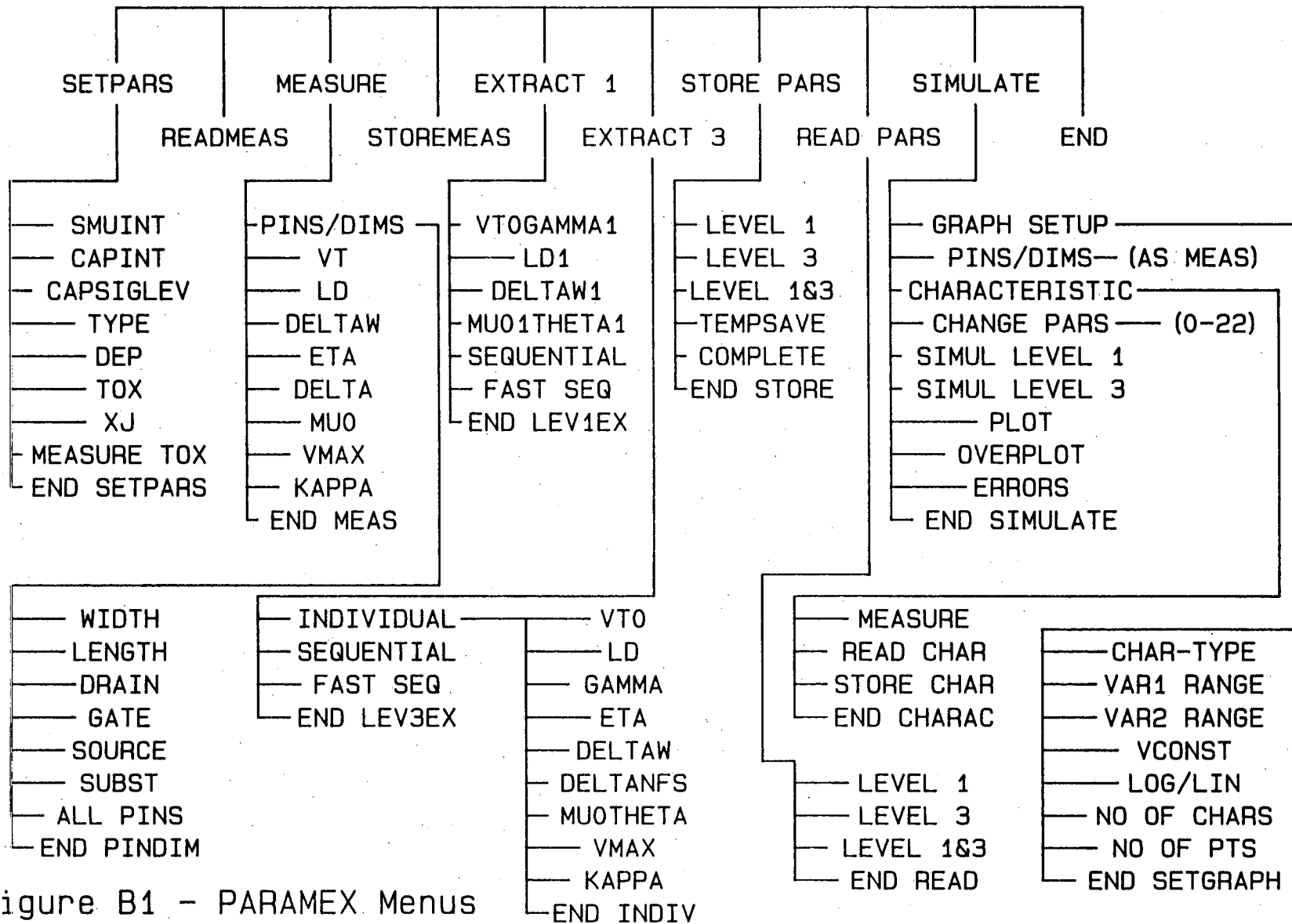


Figure B1 - PARAMEX Menus

Levl - Electrically extracted parameters for the level 1 model.

Meas 1 and Meas 2 - The most recently measured characteristics for parameter extraction.

Pars1 - Input parameters for both the level 1 and level 3 models. x_j only applies to level 3.

Pars2, Pars3 and Pars4 - Electrically extracted parameters for the level 3 model.

Phys - Physical Constants.

Setpars - Instrument addresses, integration times and signal levels.

Waferdet - Information describing the sample and time of extraction for storing with the parameters.

Setpars

The functions of this menu are two fold:

(i) to set up the hardware

and (ii) to set up the input parameters for the extraction program.

When this section of the program is entered, the current values of these quantities will be displayed on the screen. The following softkeys can then be used:

SMUINT - This option enables the integration time of the SMU's to be set. There are three integration times: short, medium and long corresponding to four, sixteen and two hundred and fifty-six conversions of the input A to D respectively.

CAPINT - This option enables the integration time of the capacitance meter to be set. Three integration times can be selected: 1ms, 10ms and 10X10ms.

CAPSIGLEV - The amplitude of the a.c. signal superimposed on the d.c. bias level supplied by the capacitance meter is set using this option. Either 10 mVr.m.s. or 30

mVr.m.s. may be used.

TYPE - This option allows the user to choose the type of device for which the parameters are to be evaluated. Type is 1 for NMOS and -1 for PMOS.

DEP - This option allows the user to choose the nature of the device for which the parameters are to be evaluated. Dep is 0 for enhancement transistors and 1 for depletion transistors.

TOX - The oxide thickness in Angstroms is entered after this key has been selected.

XJ - The approximate junction depth in μm is entered after this key has been selected.

MEASURE TOX - This option allows the user to measure the oxide thickness by measuring the capacitance of an MOS capacitor biased in accumulation.

END SETPARS - This key returns the program to the main menu.

Measure, Readmeas and Storemeas

These commands are all associated with the device characteristics used to evaluate the parameters. The voltage ranges over which the currents are measured for an NMOS device are given in Table B1. The measurements are stored in a set of arrays in the common blocks Meas1 and Meas2. The information in these arrays can be stored on disc at any time by selecting STOREMEAS. This data can then be read back into the arrays by selecting READMEAS. In order to measure the device characteristics, enter the measure menu by selecting 'MEASURE'. The softkeys in the measure menu have the following functions:

PINS/DIMS - This enters another menu(see below) which is used to set up the pin numbers and the dimensions of the device which is to be measured.

VT - This command instructs the hardware to measure the characteristics required for V_{to} evaluation on the device defined in PINS/DIMS. When the characteristic has been

Table B1 Voltage Ranges for Parameter Extraction Characteristics

Parameter	V_g	V_d	V_b	No. of Points
V_{to}	-5-0/0-5	.1	0,-1,-2,-3,-4,-5	6X51
η	-6-3.5/-1-1.5	.1,1,2,3,4,5	0	6X51
δ	-5-0/0-5	.1	-2.5	51
μ_o	-5-0/0-5	.1,2	0	2X151
v_{max}	0/5	0-4	0	51
κ	-3/2	2-5	0	31
L_d	-5-0/0-5	.1	0	51Xm
Δ_w	-5-0/0-5	.1	0	51Xn

where m and n are the number of devices used for the L_d and Δ_w measurements respectively. The different gate voltages which are specified are for depletion/enhancement. All the voltages are multiplied by -1 for PMOS.

measured it is plotted on the screen and then the program returns to the measure menu.

LD - This command starts the measurement of characteristics of devices of different lengths but of the same width which are used to determine L_d . The user is prompted for the number of devices which are to be used (between 2 and 5 inclusive) and for the dimensions and pin numbers of each device in turn just before they are measured. Finally a graph showing all the characteristics is plotted on the screen.

DELTAW - This command starts the measurement of characteristics of devices of different widths but of the same length which are used to determine Δ_w . The user is prompted for the number of devices which are to be used (between 2 and 5 inclusive) and for the dimensions and pin numbers of each device in turn just before they are measured. Finally a graph showing all the characteristics is plotted on the screen.

ETA - Measure the characteristic for evaluating η on the device defined in PINS/DIMS and plot it on the screen.

DELTA - Measure the characteristic for evaluating δ on the device defined in PINS/DIMS and plot it on the screen.

MU0 - Measure the characteristic for evaluating μ_o on the device defined in PINS/DIMS and plot it on the screen.

VMAX - Measure the characteristic for evaluating v_{max} on the device defined in PINS/DIMS and plot it on the screen.

KAPPA - Measure the characteristics for evaluating κ on the device defined in PINS/DIMS and plot it on the screen.

END MEAS - Return to the main menu.

When the PINS/DIMS option is entered, the current values for the pins and dimensions of the device to be measured are displayed on the screen. The pins and dimensions defined here currently will be used when making any parameter

measurement except L_d and Δ_w and also for the measured curve when assessing the simulated output. It is however possible to make different measurements on different devices by selecting PINS/DIMS between characteristic measurements. The PINS/DIMS softkeys have the following functions:

WIDTH - This option prompts the user for a new device width in μm and alters the display accordingly.

LENGTH - This option prompts the user for a new device length in μm and alters the display accordingly.

DRAIN - This option allows the user to alter the drain pin number. Any integer from 1 to 48 may be used.

GATE - This option allows the user to alter the gate pin number. Any integer from 1 to 48 may be used.

SOURCE - This option allows the user to alter the source pin number. Any integer from 1 to 48 may be used.

SUBST - This option allows the user to alter the substrate pin number. Any integer from 1 to 48 may be used.

ALL PINS - This option allows the user to alter all the device pin numbers.

END PINDIM - This option returns the user to either the MEASURE menu or the simulate menu depending on where it was called from.

Extract1 and Extract3

These two options from the main menu are where the characteristics contained in the two measurement arrays Meas1 and Meas2 are numerically manipulated to obtain level 1 and level 3 SPICE Parameters. In both cases there is a specific order in which the parameters should be extracted which is the order of the

softkeys k0 up to k9. The order is followed automatically if sequential extraction is selected and if fast sequential is used, no graphs illustrating the extraction process are plotted on the screen. When one of the extract menus is entered the current parameters are displayed on the screen and this list is updated after each parameter has been evaluated.

The EXTRACT1 menu will be dealt with first.

VT0GAMMA1 - $V_{t0,1}$ and γ_1 are extracted from the measurements in the array Vtmeas and the resulting parameters update those currently appearing on the screen. Two graphs are plotted; one showing the linear extrapolation of the $V_g : I_d$ curves to obtain the threshold voltages and one showing the variation of threshold voltage with substrate bias.

LD1 - $L_d,1$ is extracted from the measurements in Ldelmeas and the new value for $L_d,1$ is displayed in the parameter list. A graph is plotted of 1 over transistor gain against mask length.

DELTAW1 - $\Delta_w,1$ is extracted from the measurements in Wdelmeas and the new value for $\Delta_w,1$ is displayed in the parameter list. A graph is plotted of the variation of transistor gain with mask width.

MU01THETA1 - $\mu_o,1$ and θ_1 are extracted from the measurements in Mobmeas and their values are updated on the parameter list. Mobility is plotted against gate voltage and then $\mu_o,1$ over Mobility is plotted against gate voltage to obtain θ_1 .

SEQUENTIAL - This option extracts all the level 1 parameters in turn, in the correct order, plotting the graphs as described for each individual parameter and updates their values in the parameter list.

FAST SEQ - This option extracts all the level 1 parameters in turn in the correct order and updates their values in the parameter list. No graphs are plotted.

END LEV1EX - This key returns the program to the main menu.

The level 3 extraction menu softkeys have the following functions

INDIVIDUAL - This option takes the user into another menu where the parameters are extracted one at a time with each softkey. The keys should be pressed in the order k0 to k8. A full description of this menu is provided below.

SEQUENTIAL - This option extracts all the level 3 parameters in turn in the correct order, plots the graphs described in the individual menu description below and updates the parameter values on the list on the screen.

FAST SEQ - This option extracts all the level 3 parameters in turn in the correct order and updates the parameter values on the screen. No graphs are plotted.

END LEV3EX - This key returns the user to the main menu.

The INDIVIDUAL menu softkeys have the following functions:

VT0 - V_{to} is extracted by linear extrapolation of the $V_g:I_d$ characteristic contained in V_{tmeas} and its value is updated in the parameter list. Graphs are plotted illustrating the second derivative of I_d against V_g and the linear extrapolation of $I_d:V_g$.

LD - L_d is extracted from the data in $L_{delmeas}$ using V_{to} and the value appearing in the parameter list is updated. A graph is plotted of 1 over gain against mask length.

GAMMA - Gamma is also extracted from the information in V_{tmeas} and the values of L_d and V_{to} which have already been found are used in its derivation. A graph showing the extrapolation to find threshold voltages at different substrate biases is plotted and also a graph of threshold voltage against the square root of substrate bias including the short channel factor F_s .

ETA - η is extracted by finding the threshold voltages at different drain biases from the

subthreshold characteristics contained in Etameas and its new value is displayed in the parameter list. V_{to} , L_d and γ are all used in the extraction of η . Threshold voltage is plotted against drain voltage to illustrate the evaluation of this parameter.

DELTAW - Δ_w is extracted from the data in Wdelmeas using all the parameters extracted so far and its new value is displayed. Transistor gain is plotted against mask width in order to illustrate the extraction process.

DELTANFS - δ is found from the threshold voltage of a device at nonzero substrate bias and the extrapolation to find this threshold voltage is plotted. The data in Delmeas and the parameters extracted so far are used. Using δ and the other parameters, N_{fs} is extracted from the first characteristic in the Etameas array. The new values of both δ and N_{fs} are entered into the parameter list.

MU0THETA - Using the first characteristic contained in the Mobmeas array and the parameters extracted so far, mobility is found and plotted as a function of gate voltage. μ_o and θ are found from this data and their values displayed. μ_o over Mobility is also plotted against V_g to show how the parameters are evaluated.

VMAX - v_{max} is found from the variation of mobility with drain bias which is calculated from the data in Vmaxmeas. All the other parameters are used except κ which has not yet been evaluated. Two graphs are plotted; one showing mobility as a function of drain voltage and another showing the same data on different axes after an iteration has taken place to derive the parameter. v_{max} is updated on the parameter list.

KAPPA - κ is determined from the saturation characteristic in Kappameas after an almost complete implementation of the SPICE Level 3 Model. A graph is plotted of κ against drain bias to illustrate the extraction and the new κ value is displayed.

END INDIV - This option returns the program to the main EXTRACT3 menu.

Store and Read Pars

After the extraction has taken place, it is often useful to store the resulting set of parameters on disc. The STORE menu will be described below and the corresponding selections on the READ menu to reload those parameters will be mentioned.

LEVEL1 - This option is used to store a complete set of level 1 parameters including *Type*, *Dep* and t_{ox} on disc in a single record file. The order of storage is:

Type, *Dep*, t_{ox} , V_{io} 1, γ 1, L_d 1, Δ_w 1, μ_o 1 and θ 1.

The parameters can be read back in by choosing LEVEL 1 in the READ menu.

LEVEL3 - This option is used to store a complete set of level 3 parameters in a single record file on disc. The order of storage is:

Type, *Dep*, t_{ox} , x_j , N_{fs} , V_{io} , γ , L_d , Δ_w , μ_o , θ , v_{max} , η , δ and κ .

The parameters can be read back in by choosing LEVEL 3 in the READ menu.

LEVEL1&3 - This option is used to store a complete set of level 1 and level 3 parameters in a single record file on disc. The order of storage is:

Type, *Dep*, t_{ox} , V_{io} 1, γ 1, L_d 1, Δ_w 1, μ_o 1, θ 1, x_j , N_{fs} , V_{io} , γ , L_d , Δ_w , μ_o , θ , v_{max} , η , δ and κ .

The parameters can be read back in by choosing LEVEL1&3 in the READ menu.

TEMPSAVE - This key causes all the d.c. parameters for the level 1 and level 3 models; the wafer number; the chip number; the date and the time to be placed in the D array in the common block Datas. The number of variables stored, N_v , is 25. The count of the number of parameter sets held in the array, N_o , is incremented. Up to 50 sets of data can be held in the array. This function is used in conjunction with COMPLETE (see below).

COMPLETE - This key instructs the program to store all the information saved in the D array by TEMPSAVE on disc in a form suitable for Hewlett-Packard's Basic Statistics and Data Manipulation Software. The count of the number of sets of parameters held, N_o , is reset to 0.

Simulate

After the parameters have been extracted, their accuracy can be tested by measuring and simulating the same characteristics. These can then be compared and the errors between the two characteristics can be calculated. The simulate menu options are described below:

GRAPH SETUP - This option leads to another menu (see below) which can be used to change the type of graph and the voltage ranges over which the measured and simulated currents are to be compared.

PINS/DIMS - This option allows the user to alter the pins and dimensions of the device to be used for the measurement. It uses the same routine as the measurement section and a complete description of the PINS/DIMS menu can be found in the section describing the measure options.

CHARACTERISTIC - Again this leads to another menu which allows the user to MEASURE a characteristic, STORE a characteristic or READ a characteristic. The most recently measured or read characteristic will be stored in the Idmeas array. When a characteristic is measured, the voltage ranges are as defined in GRAPH SETUP and the device is as defined in PINS/DIMS. This information is stored with the contents of Idmeas when STORE CHARAC is called. When a characteristic is read, all this information (device sizes and graph variables) is overwritten.

CHANGE PARS - A list of parameters appears on the screen when this option is selected. By entering the number beside a particular parameter, the user is prompted for a new value of that parameter. The new value is entered in the parameter list. Entering 22 returns the program to the simulate menu.

SIMUL LEVEL 1 - This option instructs the program to simulate a device with the dimensions described in PINS/DIMS over the voltage ranges defined in GRAPH SETUP using the SPICE level 1 model. The device is simulated using the parameters which are currently held in the common blocks Pars1 and Lev1. The most recently simulated characteristic, from either the level 1 or level 3 SPICE model is held in the array Idsim.

SIMUL LEVEL 3 - This option instructs the program to simulate a device with the dimensions described in PINS/DIMS over the voltage ranges defined in GRAPH SETUP using the SPICE level 3 model. The device is simulated using the parameters which are currently held in the common blocks Pars1, Pars2, Pars3 and Pars4. The most recently simulated characteristic, from either the level 1 or level 3 SPICE model is held in the array Idsim.

PLOT - This option plots the latest measured or simulated characteristic. The user choses whether to use the screen or a plotter. If the device dimensions or any of the graph variables have been changed since the last measurement or simulation was made then the program will refuse to plot.

OVERPLOT - This option plots the latest simulated or measured characteristic on top of the previous graph whether on the screen or a plotter. If the current graph is not of the characteristic which has just been measured then the program automatically resorts to the PLOT option. Up to four characteristics can be superimposed. If this limit is exceeded the program automatically resorts to the PLOT option and makes a new graph.

ERRORS - This option produces a set of error figures between the most recently simulated and measured characteristics. The program will only carry out the error calculation if the most recently specified graph has been both measured and simulated for the current device. Average percentage errors, r.m.s. errors and absolute maximum errors will be displayed for each characteristic, as well as an overall r.m.s. error for all the characteristics.

END SIMULATE - This option returns the user to the main menu.

PARAMEX2B - Quick Start Guide

Insert program disc and type

LOAD "PARAMEX"

(LOAD "PARAMEXCT" for the curve tracer rather than the 4062B system).

If using the 4062B then press k0 : LINK TIS. If using the HP4145 curve tracer then the HPIB address of the instrument must be set in lines 135 and 140 of the program.

Then run the program.

The title will appear with the following menu on the softkeys below:

SET PARS | READ MEAS | MEASURE | STORE MEAS | EXTRACT 1
EXTRACT 3 | STORE PARS | READ PARS | SIMULATE | END

First of all, select 'SET PARS' in order to set up the instruments and input parameters. A list of the instrument settings and input model parameters will be on the screen and these can all be selected and modified using the softkeys. Just leave the SMU's on their medium integration time for the present. *Type* is 1 for NMOS and -1 for PMOS and *Dep* is 0 for enhancement and 1 for depletion. Enter also the approximate junction depth and oxide thickness. Having set these variables, leave the section using 'END SETPARS'.

Next the characteristics from which the parameters will be extracted have to be measured. Select 'MEASURE' from the softkeys in order to receive the following menu:

PINS/DIMS | VT | LD | DELTAW | ETA
DELTA | MU0 | VMAX | KAPPA | END MEAS

Select 'PINS/DIMS' in order to set up the pins and dimensions of the device to be measured. Set up the appropriate connections and the transistor size. (If a switching matrix is not used then the SMU's must be connected as follows: SMU1-drain, SMU2-gate, SMU3-source and SMU4-substrate.) Exit from this section by hitting 'END PINDIM'.

Now press 'VT' and the set of device characteristics will be plotted on the screen when they have been measured. Do the same for 'ETA', 'DELTA', 'MU0', 'VMAX' and 'KAPPA'. The last two measurements: 'LD' and 'DELTAW' require devices of different sizes and so they do not use device dimensions and pins defined in 'PINS/DIMS'. Press 'LD'. The user will be prompted for the number of devices of the same width but of different lengths which are to be used. The transistor sizes and pin numbers will be entered before each is measured. When all the devices have been measured, the characteristics will be plotted. 'DELTAW' is now measured in a similar way using devices of the same length but of different widths. Having completed all the measurements, press 'END MEAS' to return to the main menu.

The next stage is to extract parameters for either the level 1 or level 3 MOSFET model in SPICE. Select 'EXTRACT 3' for level 3 parameters. There is an order in which the parameters should be extracted. V_{to} is independent of the value of all the other parameters, then L_d is dependent only on V_{to} , γ is only dependent on L_d and V_{to} and so on right through to κ .

Select 'INDIVIDUAL'. Go through pressing the softkeys in order k0 up to k8. After each key is pressed, a graph or two will be plotted illustrating some aspect of the extraction procedure and the value will be updated on the on-screen parameter list. 'SEQUENTIAL' would automatically go through extracting each parameter in turn and 'FAST SEQ' does the same without plotting any graphs. Press 'END INDIV' to return to the EXTRACT 3 menu and 'END LEV3EX' to return to the main menu.

Finally it is useful to test the accuracy of the resulting parameters. Enter 'SIMULATE' from the main menu to obtain:

```
GRAPH SETUP| PINS DIMS | CHARACTERIS| CHANGE PARS| SIMUL LEV 1
SIMUL LEV 3 | PLOT      | OVERPLOT  | ERRORS    | END SIMULAT
```

Hit 'CHARACTERISTIC' and then 'MEASURE' on the menu that results. The characteristic defined in 'GRAPH SETUP' on the previous menu will be measured on the device described in 'PINS DIMS'. In order to plot this characteristic, press 'END CHARAC' and then 'PLOT'. The user will be given the option of using the screen or a plotter. Then press 'SIMUL LEVEL 3' to simulate the same characteristic using the level 3 MOSFET model. When this is complete this characteristic can be plotted on top of the measured one by using 'OVERPLOT'. Up to four sets of characteristics can be plotted on one graph. An assessment of the errors can then be carried out by hitting 'ERRORS'. This will list the errors between the latest simulated and measured device characteristics.

This concludes a brief run through of the main capabilities of the program and hopefully familiarises the user with the basic outline: Measure, Extract and Simulate. At various stages during the program, measurements and parameters may be stored or read in from disc. Most of these are self-explanatory, however within the 'STORE PARS' menu are two options: 'TEMPSAVE' and 'COMPLETE'. These are used to save a complete set of level 1 and level 3 parameters, the wafer number, the chip number and the date and time in a form suitable for Hewlett-Packard's Basic Statistics and Data Manipulation Software. After a complete set of parameters have been extracted use 'TEMPorary SAVE' to put the information into a buffer and when all the required sets of parameters have been temporarily saved in this way, 'COMPLETE' is used to empty this information down into a file on disc.

The characteristic used to test the accuracy of the parameters can be altered using 'GRAPH SETUP' in the 'SIMULATE' menu and there is also an option to investigate the effects of changing parameters using 'CHANGE PARS'.

APPENDIX D - Derivation of the Switching Point of CMOS Inverter

A CMOS inverter is shown in figure D1 and the NMOS and PMOS devices have been given the labels T1 and T2 respectively. The gain and threshold voltage of the NMOS device are $Beta_n$ and V_{TN} respectively and similarly $Beta_p$ and V_{TP} are the gain and threshold of the PMOS.

Consider various values of input voltage:

$$\begin{aligned}
 V_{IN} \leq V_{TN} \quad T1 \text{ OFF} \quad T2 \text{ ON} \quad V_{OUT} &= V_{DD} \\
 V_{IN} \geq V_{DD} - V_{TP} \quad T1 \text{ ON} \quad T2 \text{ OFF} \quad V_{OUT} &= 0 \\
 V_{TN} \leq V_{IN} \leq V_{OUT} + V_{TN} \quad T1 \text{ ON and saturated.} \\
 V_{OUT} - V_{TP} \leq V_{IN} \leq V_{DD} - V_{TP} \quad T2 \text{ ON and saturated.}
 \end{aligned}$$

Assuming currents are always equal then when T1 is in saturation

$$Beta_n [V_{IN} - V_{TN}]^2 = Beta_p \left\{ 2 [V_{DD} - V_{IN} - V_{TP}] (V_{DD} - V_{OUT}) - (V_{DD} - V_{OUT})^2 \right\} \quad D1$$

and when T2 is in saturation

$$Beta_p [V_{DD} - V_{IN} - V_{TP}]^2 = Beta_n \left\{ 2 [V_{IN} - V_{TN}] V_{OUT} - V_{OUT}^2 \right\} \quad D2$$

Switching occurs when both devices are in saturation i.e. when

$$Beta_n [V_{IN} - V_{TN}]^2 = Beta_p [V_{DD} - V_{IN} - V_{TP}]^2 \quad D3$$

Rearranging this equation gives the switching voltage

$$V_{IN} (switching) = \frac{\left(\frac{Beta_p}{Beta_n} \right)^{\frac{1}{2}} [V_{DD} - V_{TP}] + V_{TN}}{1 + \left(\frac{Beta_p}{Beta_n} \right)^{\frac{1}{2}}} \quad D4$$

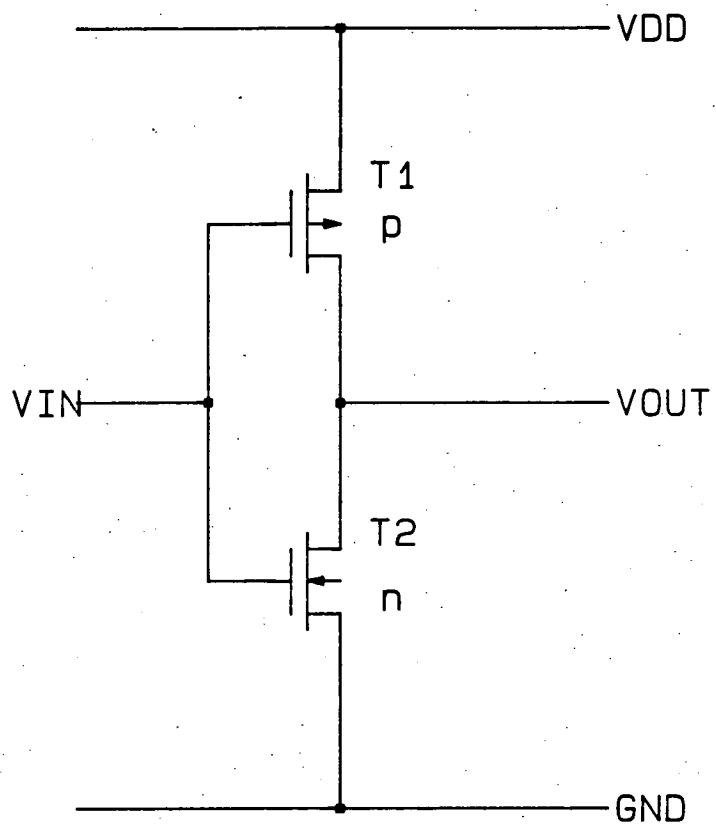


Figure D1 CMOS Inverter

APPENDIX E - Numerical Differentiation and Inverse Interpolation

The algorithms used in numerical differentiation and Inverse Interpolation can be derived using the Lozenge Diagram (see figure E1).⁷⁴ In figure E1 the symbols have the following definitions:

$$\left(\frac{u}{r}\right) = \frac{u(u-1)(u-2)\cdots(u-r+1)}{r!} \quad \text{E1}$$

and

$$\Delta y_n = y_{n+1} - y_n \quad \text{E2}$$

The difference between the independent variable values is called h and it is defined by

$$h = x_n - x_{n-1} \quad \text{E3}$$

In order to find the value of y corresponding to a particular x (interpolation) where $x = x_0 + uh$, three different formulae can be used:

(i) The Newton-Gregory forward difference formula

$$y = y_0 + \left(\frac{u}{1}\right) \Delta y_0 + \left(\frac{u}{2}\right) \Delta^2 y_0 + \cdots + \left(\frac{u}{r}\right) \Delta^r y_0 \quad \text{E4}$$

(ii) The Newton-Gregory backward difference formula

$$y = y_0 + \left(\frac{u}{1}\right) \Delta y_{-1} + \left(\frac{u}{2}\right) \Delta^2 y_{-2} + \cdots + \left(\frac{u+r-1}{r}\right) \Delta^r y_{-r} \quad \text{E5}$$

and (iii) The Stirling central difference formula

$$y = y_0 + \frac{1}{2} \left(\frac{u}{1}\right) [\Delta y_{-1} + \Delta y_0] + \frac{1}{2} \left[\left(\frac{u+1}{1}\right) + \left(\frac{u}{2}\right) \right] \Delta^2 y_{-1} + \frac{1}{2} \left(\frac{u+1}{3}\right) [\Delta^3 y_{-1} + \Delta^3 y_{-2}] + \frac{1}{2} \left[\left(\frac{u+2}{4}\right) + \left(\frac{u+1}{4}\right) \right] \Delta^4 y_{-2} + \cdots \quad \text{E6}$$

$$y = y_0 + \frac{u(\Delta y_{-1} + \Delta y_0)}{2} + \frac{u^2}{2} \Delta^2 y_{-1} + \frac{u(u^2-1)}{3!} \frac{(\Delta^3 y_{-2} + \Delta^3 y_{-1})}{2} + \frac{u^2(u^2-1)}{4!} \Delta^4 y_{-2} + \cdots \quad \text{E7}$$

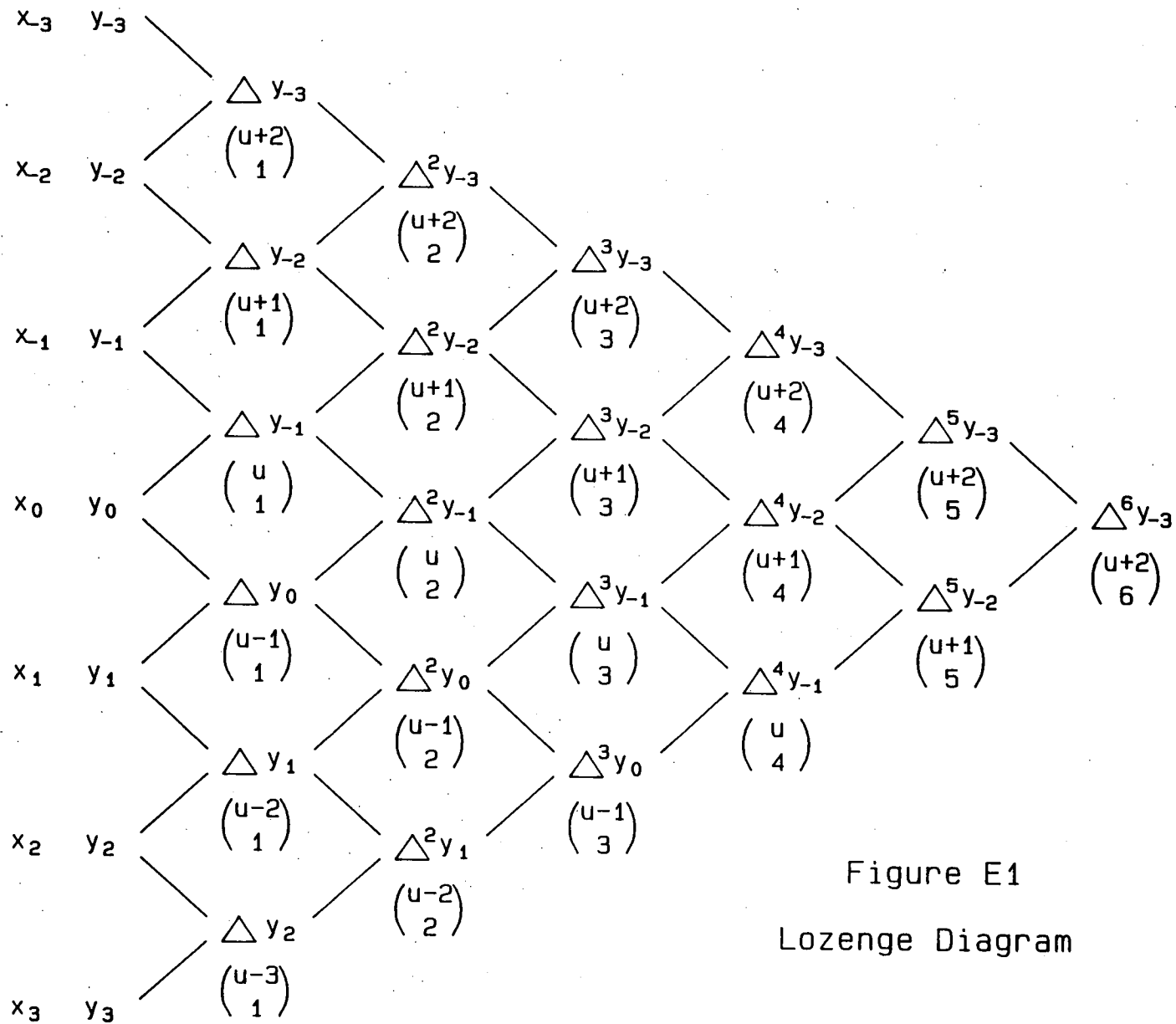


Figure E1
Lozenge Diagram

The Stirling formula using three points is therefore

$$y = y_0 + \frac{u}{2} [y_1 - y_{-1}] + \frac{u^2}{2} [y_1 - 2y_0 + y_{-1}] \quad \text{E8}$$

If a value x corresponding to a known value y has to be found then the unknown in the above equation is u . In fact, it is a quadratic in the variable u which can be solved.

$$\frac{[y_1 - 2y_0 + y_{-1}]}{2} u^2 + \frac{[y_1 - y_{-1}]}{2} u + (y_0 - y) = 0 \quad \text{E9}$$

Inverse interpolation can then be completed by calculating

$$x = x_0 + uh \quad \text{E10}$$

The derivative $\frac{dy}{dx}$ can also be derived

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx} \quad \text{E11}$$

$$u = \frac{(x - x_0)}{h}$$

$$\Rightarrow \frac{du}{dx} = \frac{1}{h} \quad \text{E12}$$

The Stirling formula using three points E8 can be differentiated:

$$\frac{dy}{du} = \frac{[y_1 - y_{-1}]}{2} + u [y_1 - 2y_0 + y_{-1}] \quad \text{E13}$$

Hence

$$\frac{dy}{dx} = \frac{1}{2h} [y_1 - y_{-1}] + \frac{u}{h} [y_1 - 2y_0 + y_{-1}] \quad \text{E14}$$

APPENDIX F - Publications

Accurate Physical Parameter Extraction for Small Geometry Devices

Mr. A. Gribben
Prof. J.M. Robertson
Dr. A.J. Walton

Edinburgh Microfabrication Facility,
Dept of Electrical Engineering,
King's Buildings,
Mayfield Road,
EDINBURGH.

INTRODUCTION

This paper outlines techniques for extracting MOSFET parameters which link the models to the fabrication process. It will be demonstrated that these parameters maintain their physical significance and hence can be used for process control as well as circuit simulation. The accuracy with which device characteristics can be simulated is then demonstrated and this will be compared with the accuracy resulting from the use of TECAP. The variation of parameters with device geometry, with position on the wafer and with process have been investigated and the results are presented.

PARAMETER EXTRACTION

Device models used in circuit simulation packages require certain input parameters in order to link the mathematical equations to particular devices. The Berkley memo "The Simulation of MOS using SPICE 2"[1] gives some guidelines on how to go about this. Other authors have written about the extraction of particular parameters e.g. Takacs et al[2] and Moll et al[3] but a detailed thorough description of a complete extraction process is not readily available. Wright[4] does describe a complete extraction process but then the parameters obtained are fed into a statistical computer package to obtain a good fit.

Most software packages which are commercially available to measure these parameters use numerical optimisation. Measured values are input and after a Jacobian iteration or similar, the correct parameters result. Some programs optimise all the parameters for all bias voltages at once e.g. SUXES produced by Stanford University and others e.g. TECAP [5] written by Hewlett-Packard are capable of optimising particular parameters in specific regions of operation separately. As a result of both of these approaches, the meaning of the parameters is not taken into account during the extraction and since there is interaction between them, there is no precise definition of the measured parameters.

The philosophy adopted here in extracting parameters for the SPICE 2 level 3 MOSFET model is to avoid using any numerical optimisation and extract parameters separately in sequence. Also each parameter is extracted by the same technique and from the same set of measurements each time, therefore the parameters can be used for process control as well as circuit simulation. The techniques for doing this will be described with graphs showing the measurement of these parameters for an NMOS enhancement $5\mu\text{m} \times 5\mu\text{m}$ device.

Non-Electrical Parameters

Two parameters are not calculated from measured current against voltage device characteristics. The gate oxide thickness per unit area, T_{ox} can be found from a measurement of the capacitance of an MOS capacitor in accumulation.

$$T_{ox} = \frac{\epsilon_{OX} \epsilon_0 \text{Area}}{C_{OX}} \quad (1)$$

Alternatively, it may be estimated from process specifications or physically measured in process. The junction depth, X_j , can be measured by profiling or estimated from process simulation.

Threshold Voltage, V_{th}

The first electrical parameter to be extracted is the threshold voltage, V_{th} . Rearranging the first order the equation for current in the linear region of operation derived by Ihantola and Moll[6] gives

$$\frac{I_d}{\beta V_d} = V_g - V_{th} - \frac{V_d}{2} \quad (2)$$

On a $V_g:I_d$ graph, the intercept on the V_g -axis i.e. when $I_d=0$ is $V_{th}+V_d / 2$. Having measured the $V_g:I_d$ characteristic in the linear region, a search is made for the two adjacent points on the characteristic between which there is greatest slope. This should be the portion of the graph just above threshold where the mobility of the carriers and hence the beta of the transistor is greatest. The straight line is extrapolated and the intercept with the gate voltage axis is $V_{to} + V_d / 2$ (figure 1).

Diffusion Length , L_d

In the processing of wafers, the source and drain regions are implanted or diffused into areas defined by the gate. Inevitably there is some sideways diffusion under the gate which causes a reduction from the mask length L_m , of $2L_d$.

To find the diffusion length L_d , the betas of different length transistors are found. The drain voltage is kept low so that there is no depletion region around the drain which would affect the measured lateral diffusion. The gate voltage should be large enough to bias the device in the linear region but no higher. This enables the maximum β to be measured and so reduces the influence of parasitic source and drain contact resistances on the measurement.[2]

The expression for transistor gain β can be manipulated to obtain

$$\frac{1}{\beta} = \frac{L_m}{\mu_{eff} C_{ox} W_m} - \frac{2 L_d}{\mu_{eff} C_{ox} W_m} \quad (3)$$

If $1 / \beta$ is plotted against L_m then the intercept of the best fit straight line with the L_m axis is $2L_d$ (figure 2).

Substrate Bias Coefficient , γ

In order to extract the parameter γ , the same $V_g:I_d$ curve is measured as for V_{to} at several different values of substrate bias. From these, the threshold voltages at different substrate biases are obtained in a similar fashion to the extraction of V_{to} (figure 3). This leaves three equations to be solved:

$$N_{sub} = \frac{\gamma^2}{2 q \epsilon_{si} \epsilon_0} C_{ox}^2 \quad (4)$$

$$\phi = \frac{2 k T}{q} \ln \left\{ \frac{N_{sub}}{N_i} \right\} \quad (5)$$

and

$$V_{th} = V_{to} + \gamma F_s (\phi - V_b)^{1/2} \quad (6)$$

These equations are all dependent upon one another. The short channel factor, F_s is dependent upon substrate concentration and substrate bias.

An estimate is made of γ and the following procedure is repeated until a value satisfying the above equations is found. N_{sub} , X_d and ϕ are all derived from γ . $F_s (\phi - V_b)^{1/2}$ is then evaluated for each substrate bias. A new γ is then determined as the slope of the best fit straight line between the values of $F_s (\phi - V_b)^{1/2}$ and V_{th} (figure 4).

Drain Feedback Coefficient , η

Threshold voltage is also dependent upon the drain voltage through the coefficient σ which is related to the parameter, η . This effect is known as static feedback and its magnitude varies considerably with the length of device. From the threshold voltage equation, it can be seen that σ is minus the slope of the $V_{th}:V_d$ line.

$$V_{th} = V_{fb} + \phi - \sigma V_d + \gamma F_s(\phi - V_b)^{1/2} + F_n(\phi - V_b) \quad (7)$$

The parameter η theoretically independent of device geometry is deduced from

$$\eta = \frac{\sigma C_{ox} L^3}{8.15E-22} \quad (8)$$

To evaluate η , threshold voltage must be measured at different drain voltages. Gate voltage against drain current characteristics at different drain voltages are measured (figure 5).

In order to evaluate V_{th} at each value of drain voltage, first of all the leakage current is found on the lowest drain voltage curve where $V_g = V_{th}$. If the threshold voltage lies between two gate voltage points, the leakage current is estimated using linear interpolation. The reverse process to the one outlined above for deducing the leakage current, is used to find the gate voltages (equal to the threshold voltages) corresponding to that leakage current on the gate voltage characteristics at different drain voltages.

V_{th} is plotted against V_d and σ and η are evaluated as described previously (figure 6).

Width Reduction , Δw

As well as the reduction in channel length, channel width is reduced below the mask width during processing. In this case, the difference is caused by the bird's beak effect in the field oxide and also the encroachment of the field implant into the channel region.

The β values for various width devices have to be found. The drain voltage is kept low to keep the depletion region small. The gate voltage is chosen, just as for L_d , to maximise β . [2]

Expanding the equation for β gives

$$\beta = \mu_{eff} C_{ox} \frac{W_m}{L_m - 2 L_d} - \mu_{eff} C_{ox} \frac{2 \Delta w}{L_m - 2 L_d} \quad (9)$$

reveals that $2 \Delta w$ is the intercept on the W_m axis when plotting β against W_m (figure 7).

Narrow Channel Factor , δ

This parameter is used to take account of threshold of a narrow device at non-zero substrate bias. Therefore the $V_g:I_d$ characteristic of a narrow device is measured with a substrate voltage applied. The threshold voltage is extracted using linear regression.

The threshold voltage predicted using the model without any narrow channel correction is evaluated using

$$V_{th} = V_{to} - \gamma \phi^{1/2} + \gamma F_s(\phi - V_b)^{1/2} - \sigma V_d \quad (10)$$

The difference between the measured threshold voltage and the one calculated by the model is equated to the narrow channel term and leads to

$$\delta = \frac{4 \text{ Cox } W (V_{th}(\text{meas}) - V_{th}(\text{calc}))}{2 \pi \epsilon_{si} \epsilon_0 (\phi - V_b)} \quad (11)$$

δ is then calculated from the expression above.

Fast State Density , Nfs

Nfs is a measure of the slope of the subthreshold, gate voltage:drain current characteristic. The form of the drain current expression is

$$I_d = I_0 \exp \left[\frac{q (V_g - V_{on})}{N k T} \right] \quad (12)$$

where I_0 is a base current given by the expression for current in the linear region with $V_g = V_{on}$. This is modified by the exponential term which varies with V_g . Currents are measured at two gate voltages in the middle of the subthreshold region. These two points are V_{g1} , I_{d1} and V_{g2} , I_{d2} . Then

$$N = \frac{q (V_{g1} - V_{g2})}{k T \ln \left[\frac{I_{d1}}{I_{d2}} \right]} \quad (13)$$

Then

$$N_{fs} = \left[N - 1 - \frac{Q_b}{2 (\phi - V_b) \text{ Cox}} \right] \frac{\text{Cox}}{q} \quad (14)$$

Two points are chosen from the middle of the subthreshold region on the low drain voltage characteristic as measured for η (figure 5). The procedure outlined above yields the parameter, δ .

Carrier Mobility , μ_0 and θ

Carrier mobility varies both with drain voltage and with gate voltage. The model assumes that carrier mobility is a maximum, μ_0 at threshold voltage and that there is a gradual degradation of mobility as gate voltage increases. This degradation is modeled by θ . The lesser effect, carrier mobility reduction with increasing drain voltage, is modeled by V_{max} and its extraction will be described later. Mobility parameters are highly dependent upon the length of device.

Firstly, the values of mobility at different gate voltages are found. The gate voltage:drain current characteristic is measured at very low V_d . The parameters extracted so far are used in the model to provide V_{th} and F_b . Mobilities can then be calculated using the expression:

$$\mu_{eff} = \frac{I_d}{\text{Cox} \frac{W}{L} \left[V_g - V_{th} - \frac{1+F_b}{2} v_d \right] v_d} \quad (15)$$

These mobilities (figure 8) are taken to be the surface mobilities since the drain voltage was kept low. Surface mobility is defined as the carrier mobility before any reduction due to the drain voltage takes place.

Rearranging the equation governing mobility as a function of gate voltage leads to

$$\frac{\mu_0}{\mu_s} = \theta V_g - \theta V_{th} + 1 \quad (16)$$

If the best fit straight line is fitted on the graph of $1 / \mu_s : V_g$ then $1 / \mu_s = 1 / \mu_0$ when $V_g = V_{th}$. Theta is the slope of the graph μ_0 / μ_s against V_g (figure 9).

Maximum Carrier Velocity , Vmax

Drain current is measured as a function of drain voltage for a high gate voltage so that most of the measured characteristic lies in the linear region of operation. Mobility is calculated from the drain current values in the same way as it was calculated for the evaluation of μ_0 and θ using all the parameters extracted so far (figure 10).

Rearranging the expression governing drain modulation of mobility in the linear region gives

$$\frac{V_d}{L} = \left[\frac{1}{\mu_{eff}} - \frac{1}{\mu_s} \right] V_{max} \quad (17)$$

At low drain voltage, the effective carrier mobility μ_{eff} is approximately the surface mobility, μ_s . V_{max} is the slope of the line obtained when plotting V_d / L against $1 / \mu_{eff} - 1 / \mu_s$ for $V_d < V_{dsat}$ (figure 11).

Saturation Slope Coefficient , κ

κ is the parameter which characterises the slope of the drain voltage : drain current curve in saturation. As the voltage on the drain is increased with respect to the gate, the point is reached where carrier inversion cannot be sustained at the drain end of the channel. This effect which is called channel pinch-off, combined with carrier velocity saturation explains why at a certain drain voltage the device current stops increasing at the same rate. [7] As drain voltage further increases, the length of channel which is no longer inverted; L_{del} , increases and thus the actual length of the channel is reduced. Conduction continues because of the high parallel electric field across the pinched off region. There is a slight increase in current due to the shortening of the channel. This slight increase is much greater for a shorter channel device where the reduction in length is much more significant.

The characteristic of a device in saturation is measured by measuring the drain current at a medium gate voltage over a range of high drain voltages. The model is then implemented with the parameters extracted to find V_{dsat} and I_{dsat} . I_{dsat} and the actual measured drain current I_d are compared to find the change in length, L_{del} using

$$L_{del} = L \left[1 - \frac{I_{dsat}}{I_d} \right] \quad (18)$$

If as in some cases, I_{dsat} is greater than I_d then for that portion of the measured curve, $L_{del}=0$ since a negative value of L_{del} is unrealistic. κ for various drain values is calculated (figure 12) from the L_{del} values using

$$\kappa = \frac{L_{del}^2}{X_d^2 (V_d - V_{dsat})} \quad (19)$$

It is found that κ almost becomes a constant at higher drain voltages. Therefore, the value of κ found at the largest normal operating V_d is chosen as the parameter.

The Parameters

The parameters for a $5\mu\text{m} \times 5\mu\text{m}$ NMOS enhancement device obtained from the graphs illustrating the extraction procedure are as follows:

Tox	699E-10	m
Xj	0.8E-6	m
Nfs	6.34E+14	1/(m ²)

Vto	1.01	V
γ	1.70	1/(V ^½)
Ld	1.06E-6	m
Δw	1.05E-6	m
μo	0.061	(m ²)/(Vs)
θ	0.026	1/V
Vmax	2.96E+5	m/s
η	0.021	
δ	0.093	
K	0.523	

The performance of the extraction techniques for short devices was tested using a 30μm X 1.5μm device. The parameters obtained were

Tox	350E-10	m
Xj	0.5E-6	m
Nfs	5.70E+14	1/(m ²)
Vto	0.82	V
γ	1.41	1/(V ^½)
Ld	0.29E-6	m
Δw	0.65E-6	m
μo	0.046	(m ²)/(Vs)
θ	0.099	1/V
Vmax	1.98E+5	m/s
η	0.046	
δ	0.0	
K	0.015	

RESIMULATION of CHARACTERISTICS

These parameter sets were used to simulate two sets of device characteristics:

- Vd 0 to 5 V
Vg 1,2,3,4 and 5 V
Vb 0 V
- Vd 5 V
Vg 0 to 5 V
Vb 0, -1, -2, -3 and -4 V

The simulated characteristics are plotted with the actual measured characteristics in figures 13 and 14. The solid lines are the measured characteristics and the dotted lines are the simulated ones. In figure 13, the average percentage errors in the Vg=3, Vg=4 and Vg=5 are 0.8%, 0.9% and 1.4% respectively. The accuracy of the parameters in saturation is confirmed by the second set of characteristics where the absolute maximum error for Vb=0 is 0.52μA and for Vb=-3 is 2.8μA.

For the short channel device, the lateral diffusion Ld, at each end of the channel is about 0.3μm which means that the effective channel length for this device is less than 1μm. The first set of characteristics set out above was used to confirm the validity of the parameters. In figure 15, the average percentage errors in the Vg=3, Vg=4 and Vg=5 characteristics are 3.7%, 3.3% and 2.1% respectively.

Comparison with TECAP

The Transistor Electrical Characterisation and Analysis Program (TECAP) written by Hewlett-Packard uses numerical optimisation to obtain parameters from measured characteristics. The program is used in the HP Pascal 3.0 operating environment and will run on a variety of combinations of HP instruments. The program is menu driven from a main menu and a set of submenus resulting in a very user friendly interface between program and user.

The suggested procedure is as follows. First of all, approximate values for some parameters e.g. L_d and Δw are set up. Then a set of linear region characteristics are measured to obtain the primary parameters V_{to} , μ_0 , θ and γ . Having measured the characteristics, the user instructs the program to hold all the other parameters fixed while the error between the measured and simulated characteristics is minimised by altering these parameter values. Figures 16 and 17, show the comparison of measured and simulated linear characteristics before and after optimisation. N_{fs} is found similarly from measuring a subthreshold characteristic and η and κ from a saturation characteristic. By measuring characteristics on devices of different lengths and widths, values for effective lengths and widths can be found. When the transistor characteristic is simulated, a fit with similar errors to those obtained by using the Parameter extraction routines outlined above is obtained (figure 18).

Although the program is very user friendly and a good fit can be produced, it needs a fairly high degree of expertise to obtain physically realistic parameters. It is important to impose good physical limits on the parameters during optimisation and optimise parameters over the measured current regions where they have most influence. The initial estimates of L_d and Δw will affect the values of μ_0 and θ which will lead to changes in η and κ . It may be necessary to repeat the primary parameter extraction when L_d and Δw have been found later in the extraction procedure. Also the numerical optimisation route for determining effective width and length will only work on fairly large devices where threshold voltages and mobilities are constant from one device to the next.

The program has capabilities beyond extracting d.c. parameters for the SPICE type model. It is able to measure C-V parameters and with a little extra software parameters for other MOS models. Finally parameters can also be extracted for bipolar devices where approximate physical parameter extraction is performed before numerical optimisation.

PARAMETER VARIATIONS (i)with geometry, (ii)across a wafer and (iii)with process.

Parameter Variation with Geometry

A couple of examples of geometry effects are demonstrated here. Firstly, the measurement of η and secondly the variation of mobility with drain bias.

The subthreshold characteristics of a $100\mu\text{m} \times 100\mu\text{m}$ device, a $3.5\mu\text{m} \times 3.5\mu\text{m}$ device and a $2.5\mu\text{m} \times 2.5\mu\text{m}$ device are shown in figures 19, 20 and 21 respectively. The shift in threshold with drain bias is very small for the large device, noticeable for the $3.5\mu\text{m}$ device and very large for the small device. The graphs illustrating the extraction of η for the large and small devices are shown in figures 22 and 23 respectively. From figure 22 it is seen that there is virtually no change in threshold voltage with drain bias whereas for the small device the threshold shifts by approximately 0.3V as the drain voltage is increased from 1V to 5V.

Secondly the variation of mobility with drain bias which is a strong function of the channel length was examined. For the $100\mu\text{m}$ device, figure 24 shows that there is only a very slight decrease in carrier mobility with drain bias.

The 3.5 μ m device however, (figure 25) exhibits a decrease of about 30% in the mobility.

Parameter Variations across a Wafer

Some study has been made into the variation of parameters on a single wafer. Two examples of the results are shown in figures 26 and 27. One shows the variation of V_{to} and the other the variation of L_d . For V_{to} , there is an area on the left, where the values are higher than average (above 1.058V) and V_{to} gradually decreases to a low spot on the right where V_{to} values are below 1.028V.

The wafer map of L_d also shows that there is a gradual change in the parameter over the wafer. The low L_d values, less than 1.04 μ m occur on the left and there are some L_d 's greater than 1.19 μ m on the extreme right. This shows that there is a change in the value of L_d by over 15% over the wafer.

Parameter Variation with Process

Parameters are measured on a batch of 24 wafers, each wafer resulting from a different process. Figures 28 and 29 show how V_{to} and γ vary with changes in the process. Three sites are measured on each wafer and this leads to the slight spread of the parameters plotted for each wafer. The process is a p-well CMOS process so increasing the substrate implant decreases the threshold voltages of the n-channel devices. The opposite effect results from increasing the tub implant dose. Threshold voltage also goes up as oxide thickness increases. In this batch the thickness increases from 700A through 800A and 900A to 1000A.

γ is proportional to the oxide thickness and the square root of substrate concentration. Hence the trend for the change in γ with the process changes used here is the same as for V_{to} .

References

1. Andrei Vladimirescu and Sally Liu, "The Simulation of MOS Integrated Circuits using SPICE2," UCB/ERL M80/7 (February 1980).
2. Dezsoe Takacs, Wolfgang Muller, and Ulrich Schwabe, "Electrical Measurement of Feature Sizes in MOS Si^2 -Gate VLSI Technology," IEEE Trans. Electron Devices, Vol. ED-27, (8) pp. 1368-1373 ().
3. K. L. Peng, S. Y. Oh, M. A. Afromowitz, and J. L. Moll, "Basic Parameter Measurement and Channel Broadening Effect in the Submicrometer MOSFET," IEEE Trans. Electron Devices, Vol. EDL-5, (11) pp. 473-475 (November 1984).
4. G. T. Wright and H. M. Gaffur, "Pre-Processor Modeling of Parameter and Geometry Dependence of Short and Narrow MOSFET's for VLSI Circuit Simulation, Optimization, and Statistics with SPICE," IEEE Trans. Electron Devices, Vol. ED-32, (7) pp. 1240-1245 (July 1985).
5. Ebrahim Khalily, Peter H. Decher, and Darrell A. Teegarden, TECAP2: An Interactive Device Characterisation and Model Development System.
6. H. K. J. Ihantola and J. L. Moll, "Design Theory of a Surface Field-Effect Transistor," Solid-State Electronics, Vol. 7, pp. 423-430 (1964).

7. D. Frohman-Bentchkowsky and A. S. Grove, "Conductance of MOS Transistors in Saturation," IEEE Trans. Electron Devices, Vol. ED-16, (1) pp. 108-113 (January 1969).

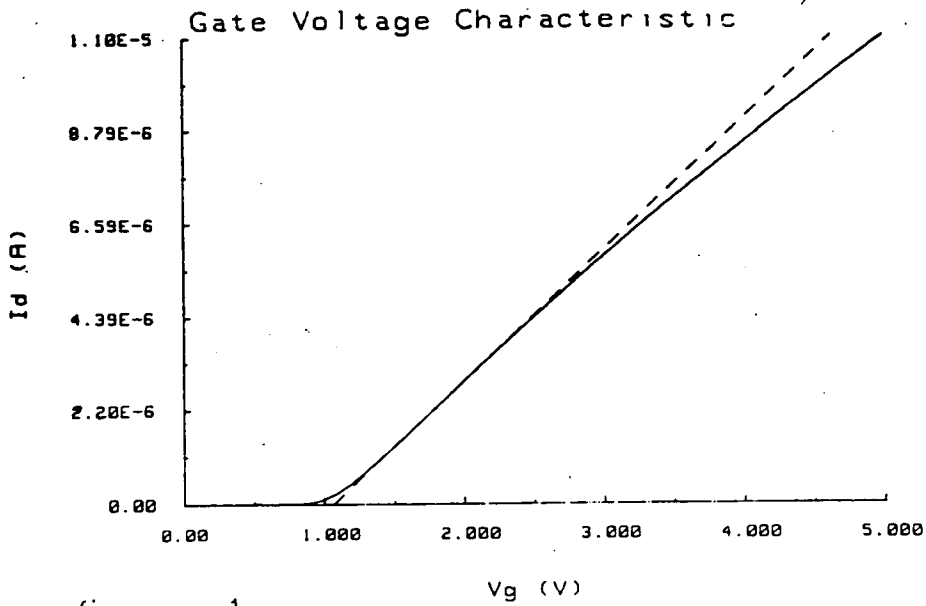


figure 1

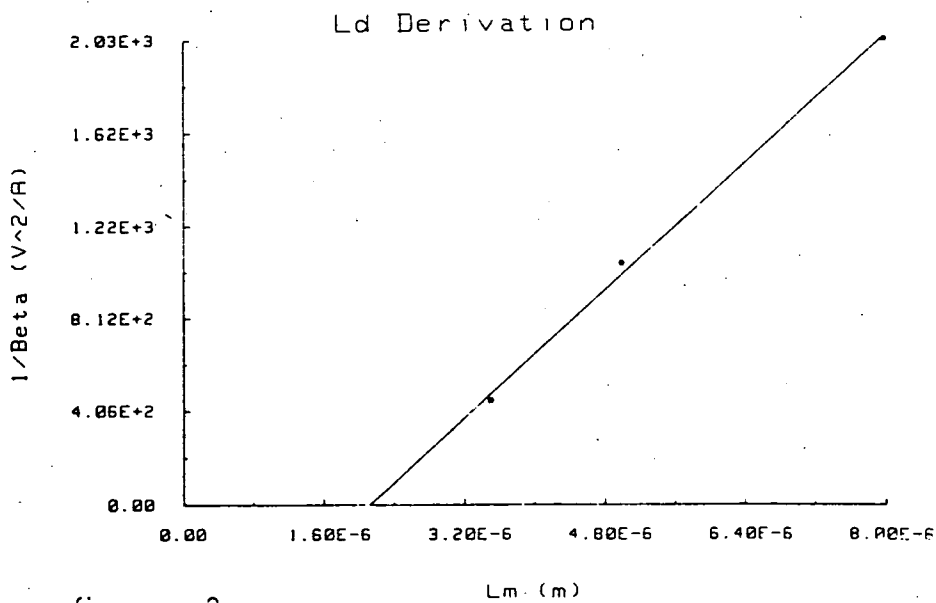


figure 2

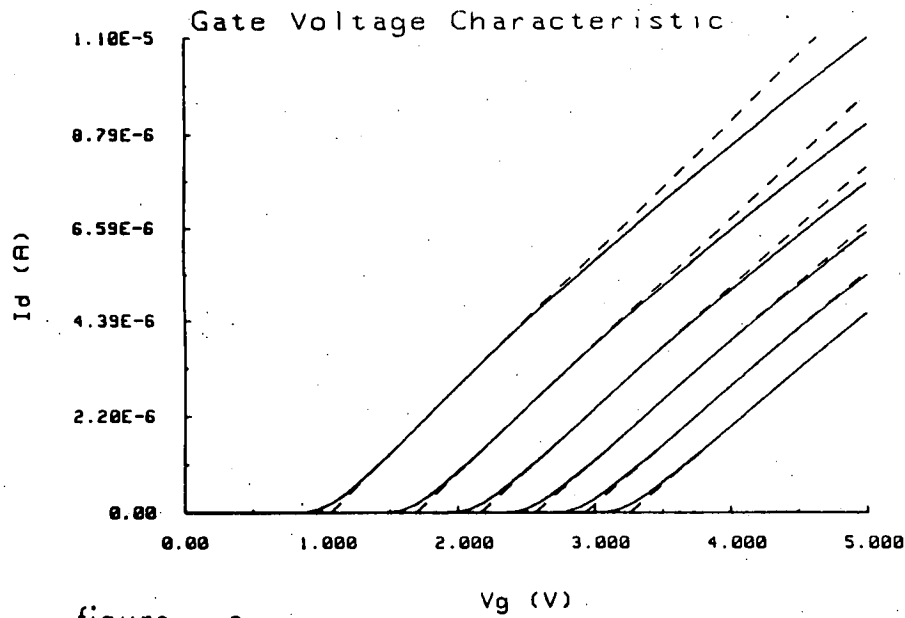


figure 3

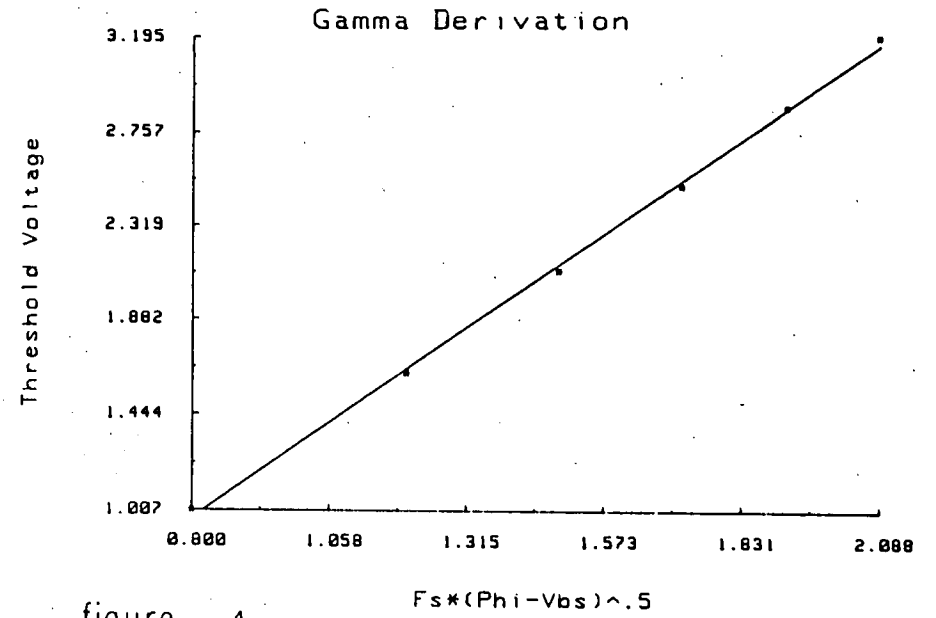


figure 4

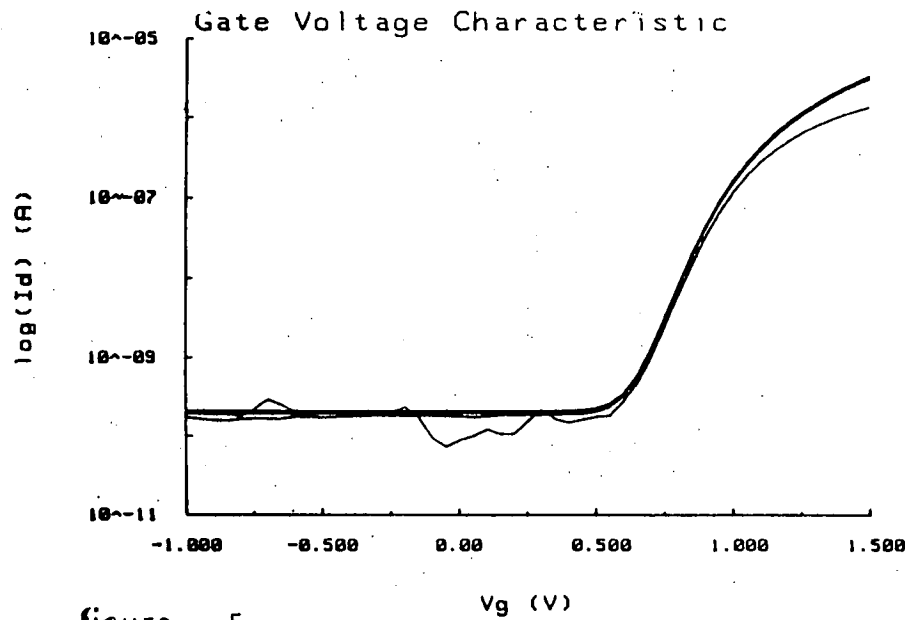


figure 5

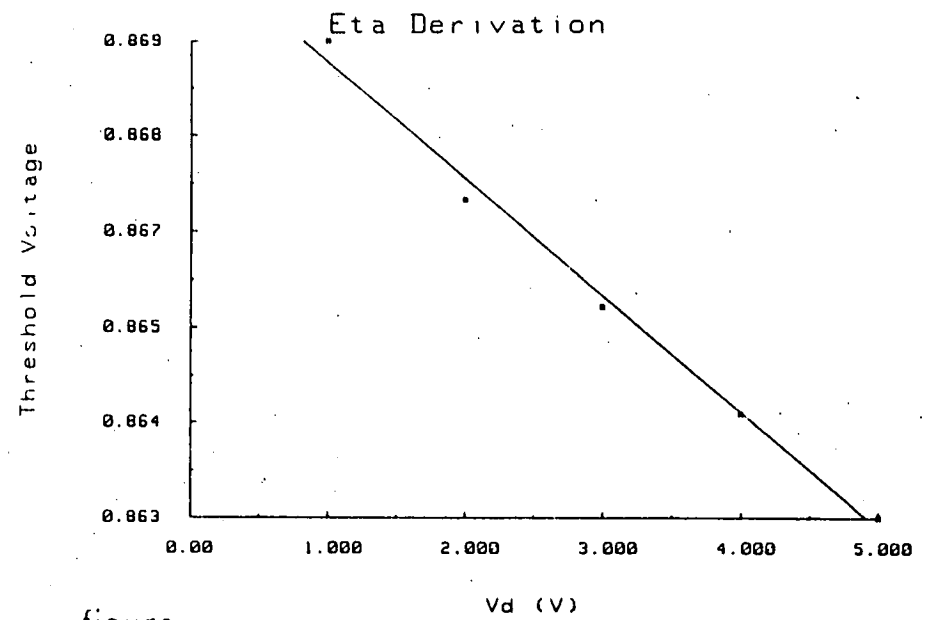


figure 6

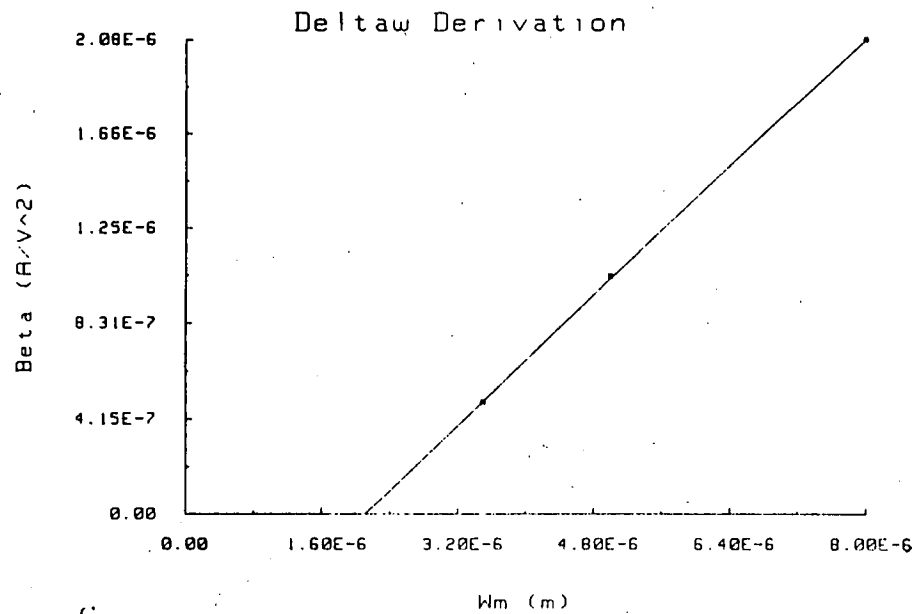


figure 7

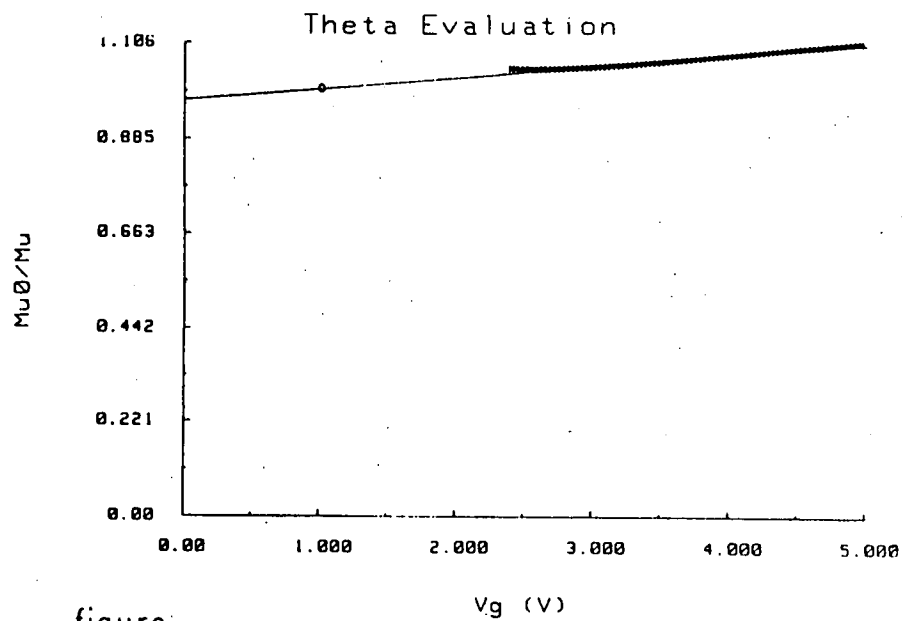


figure 9

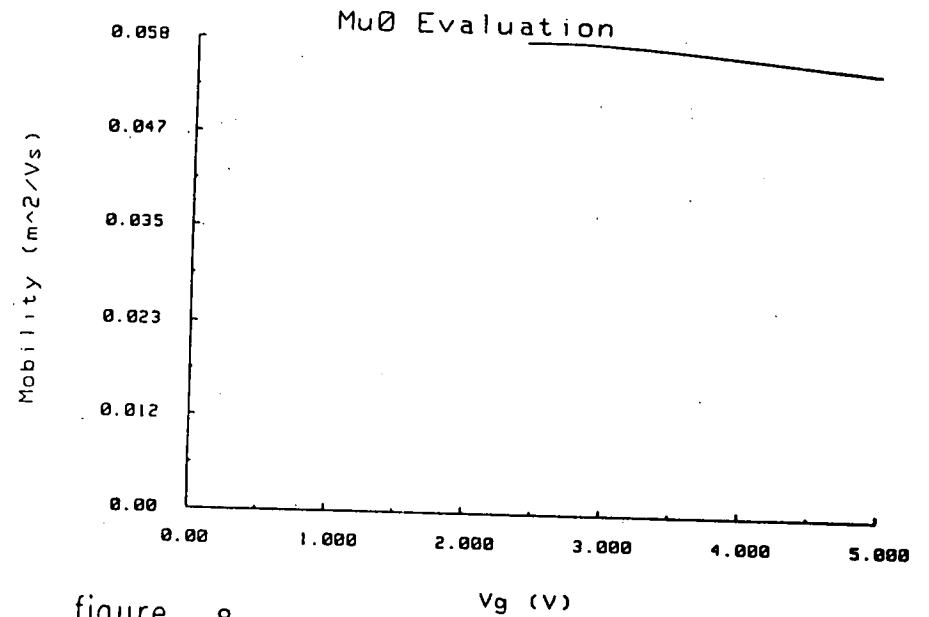


figure 8

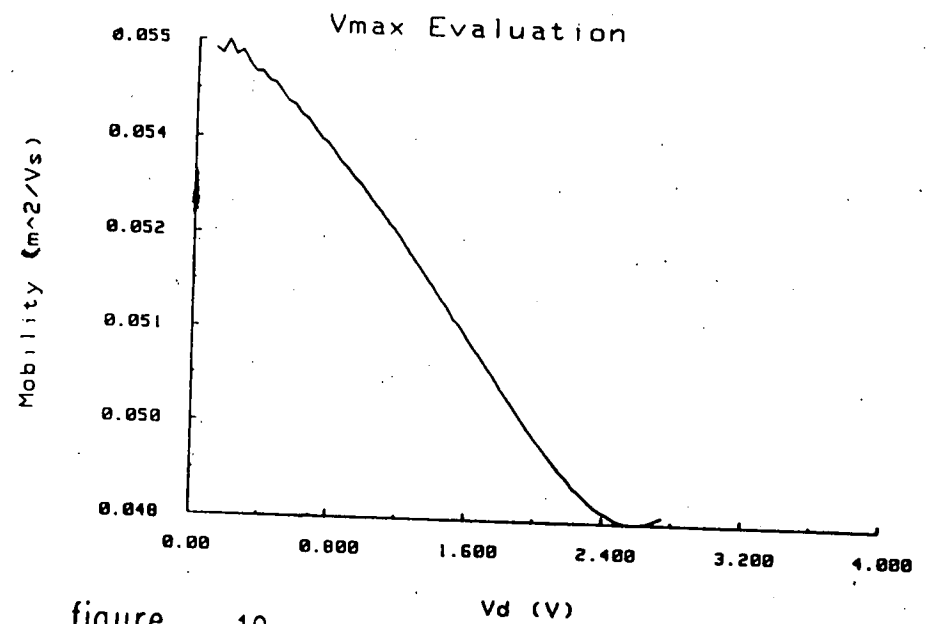


figure 10

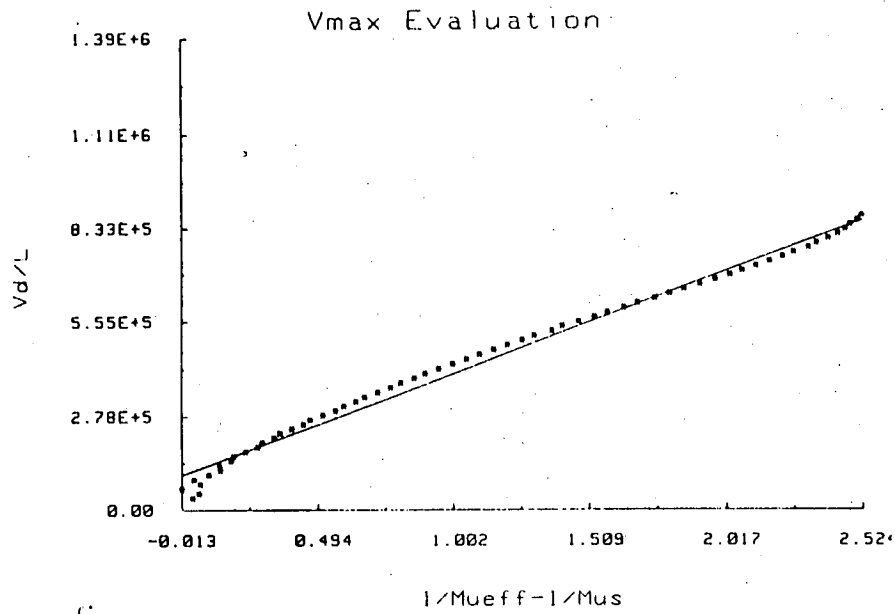


figure 11

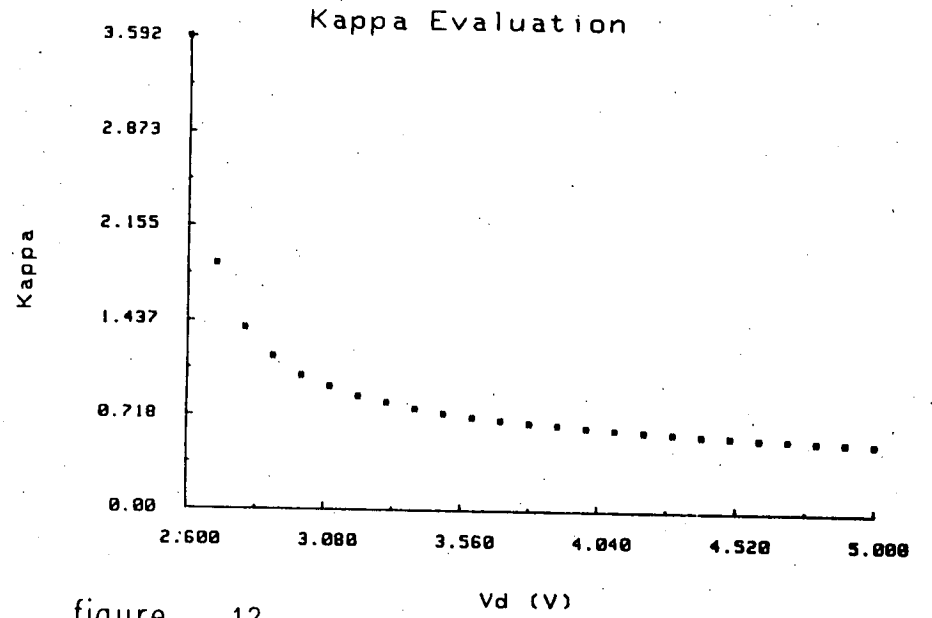


figure 12

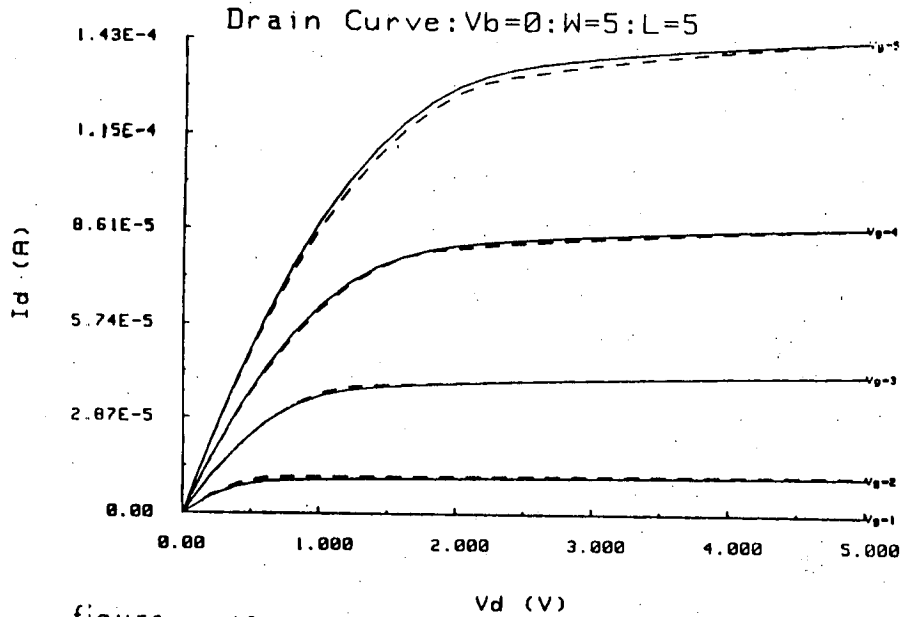


figure 13

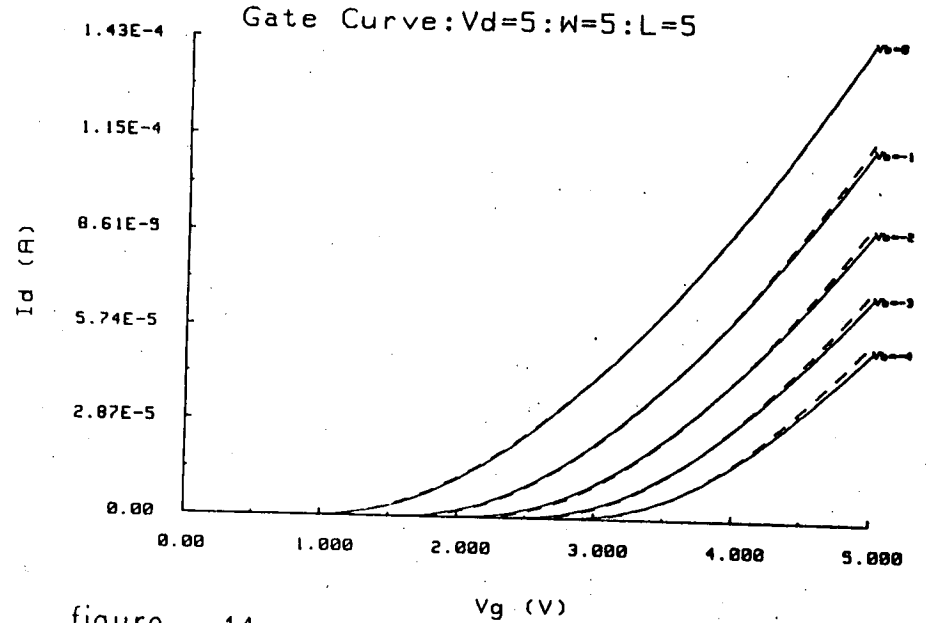


figure 14

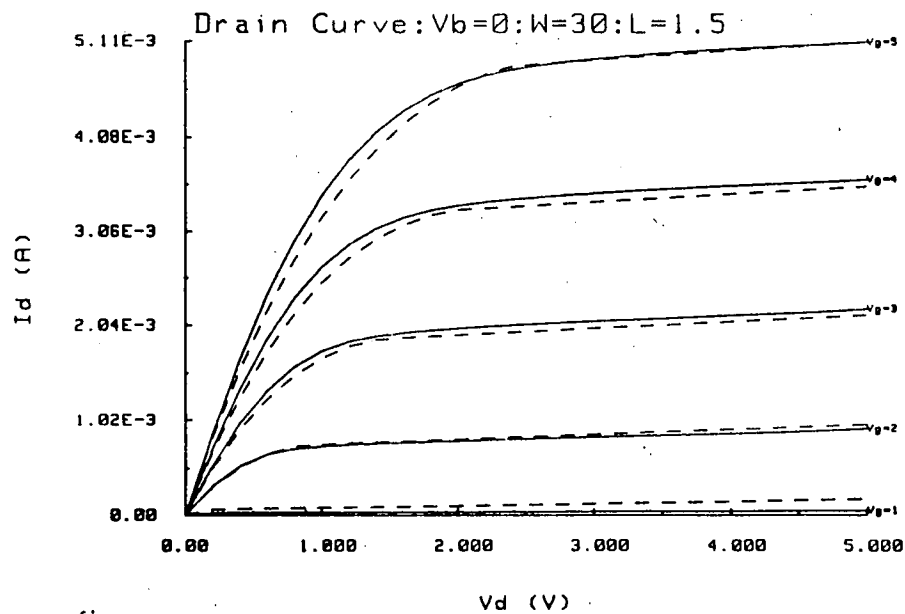
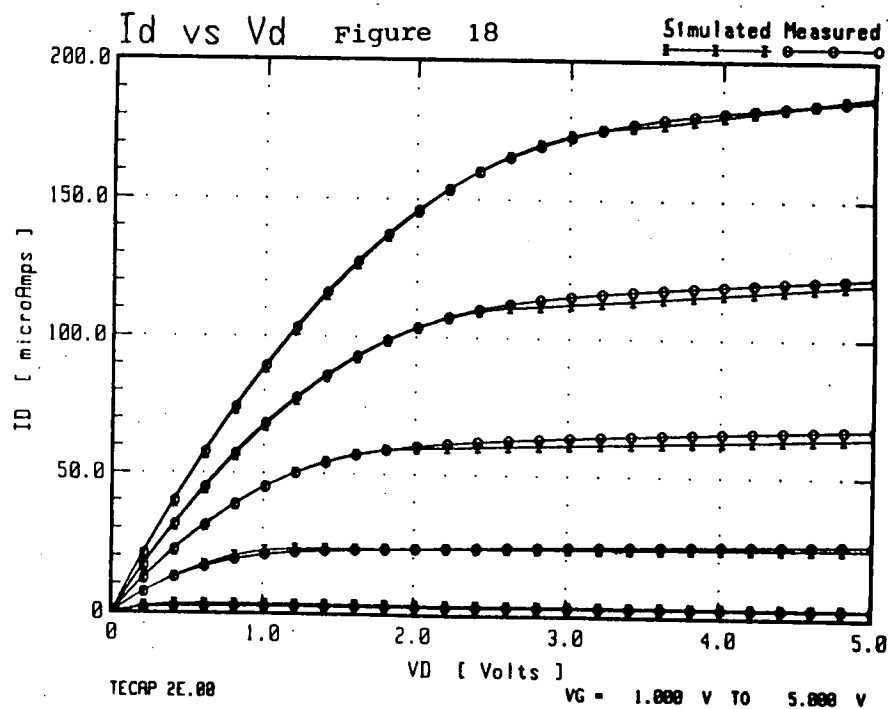
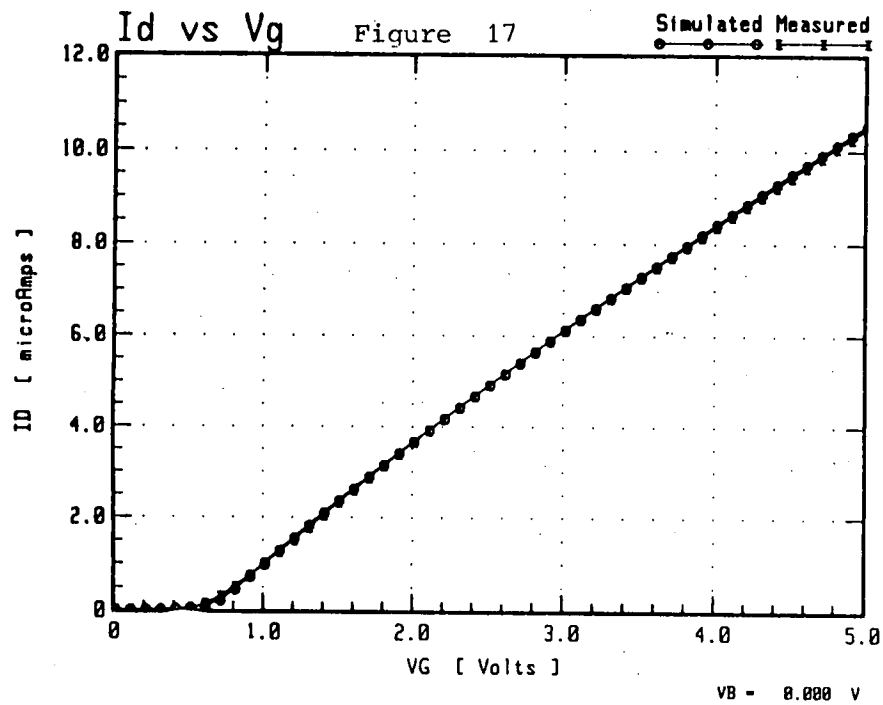
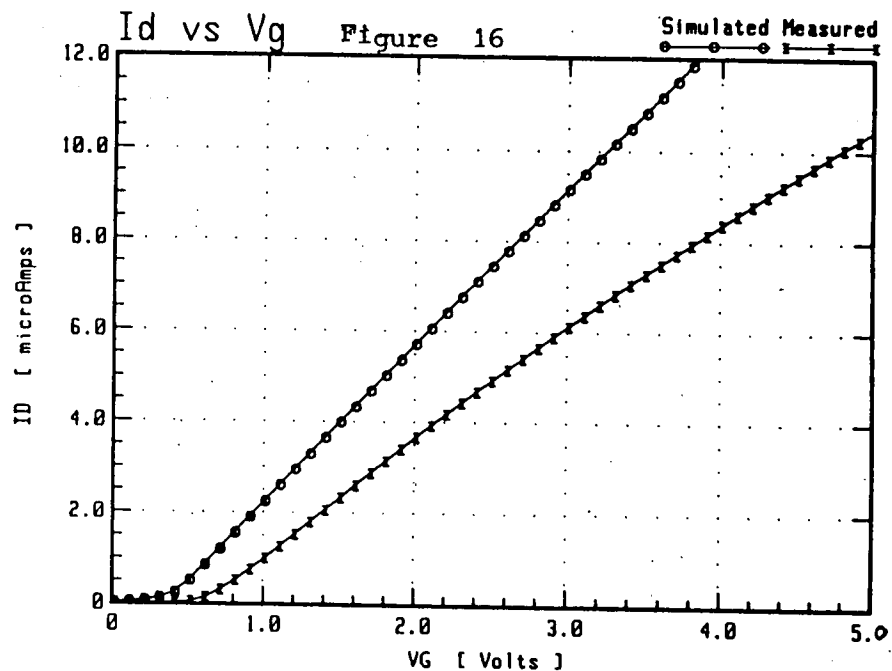


figure 15



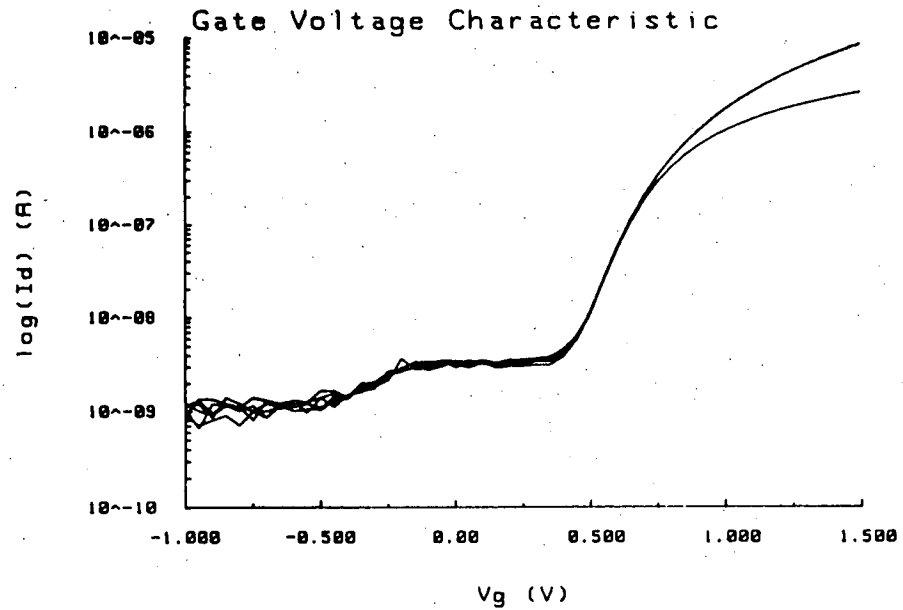


figure 19 - 100um X 100um Device

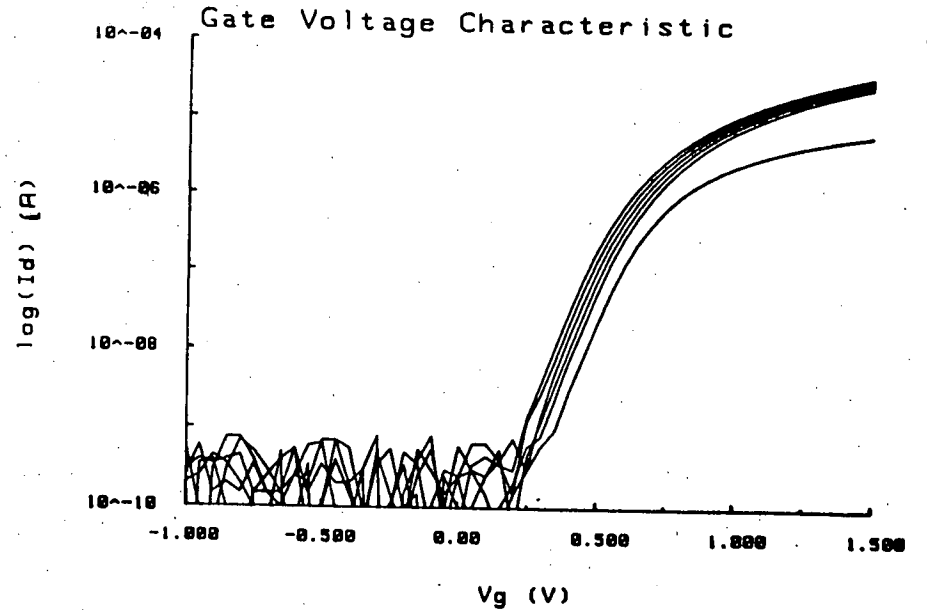


figure 20 - 3.5um X 3.5um Device

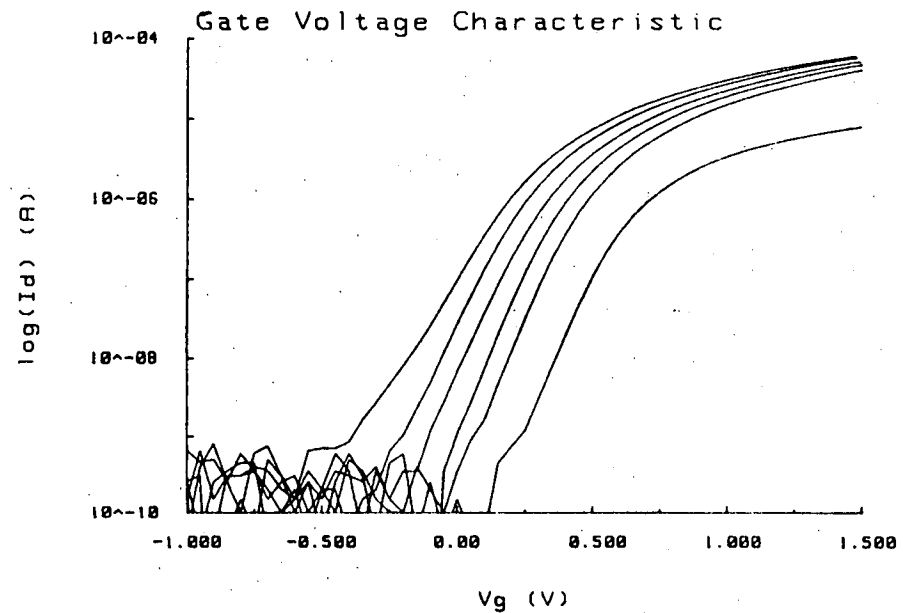


figure 21 - 2.5um X 2.5um Device

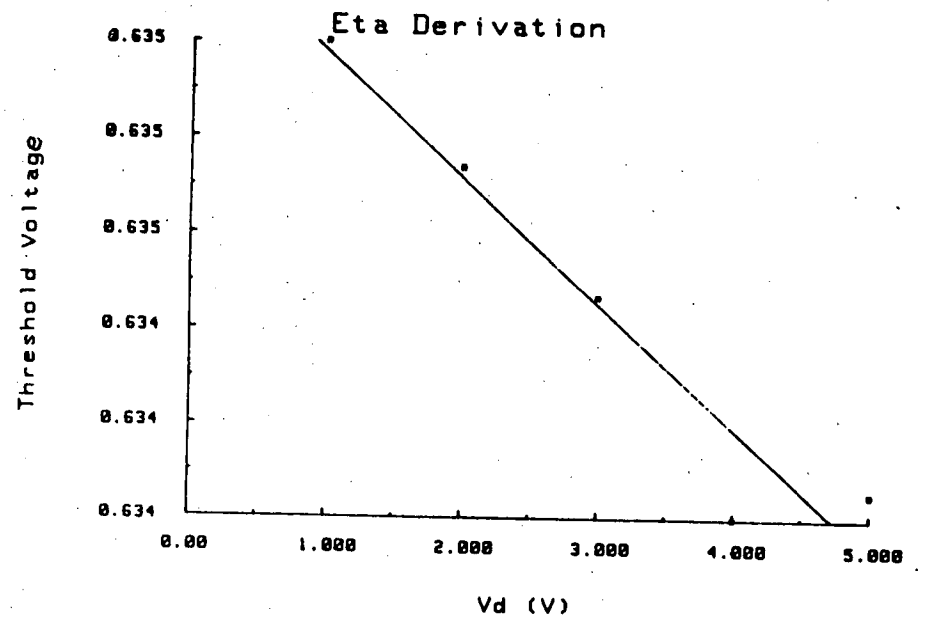


figure 22 - 100um X 100um Device

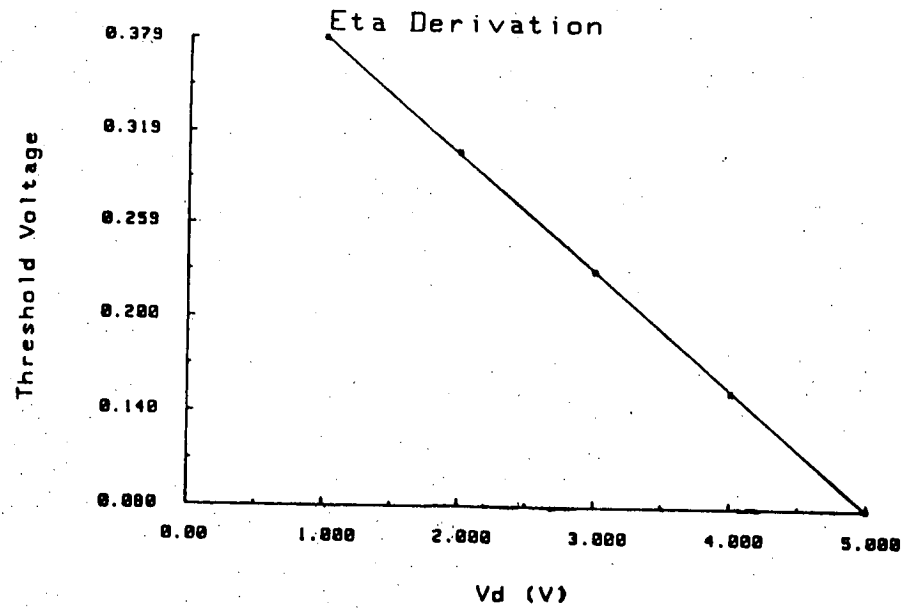


figure 23 - 2.5um X 2.5um Device

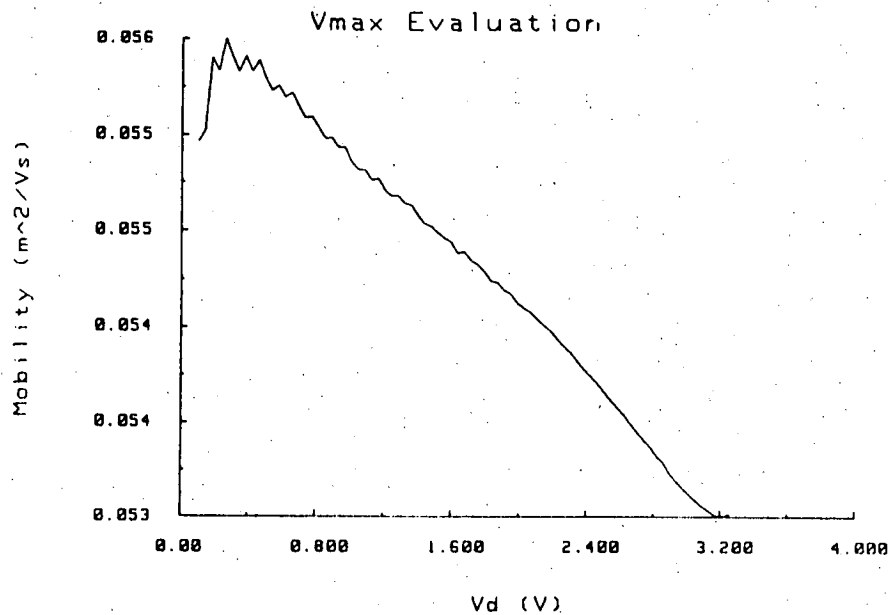


figure 24 - 100um X 100um Device

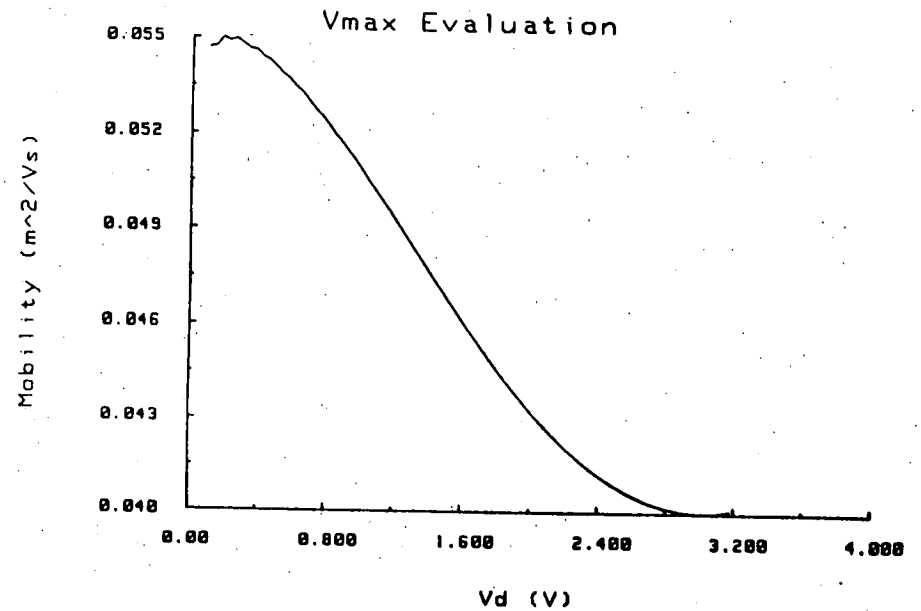


figure 25 - 3.5um X 3.5um Device

Figure 26
Variation of V_{t0} across a Wafer

A ≥ 1.073
B ≥ 1.058
C ≥ 1.043
D ≥ 1.028
E < 1.028

V

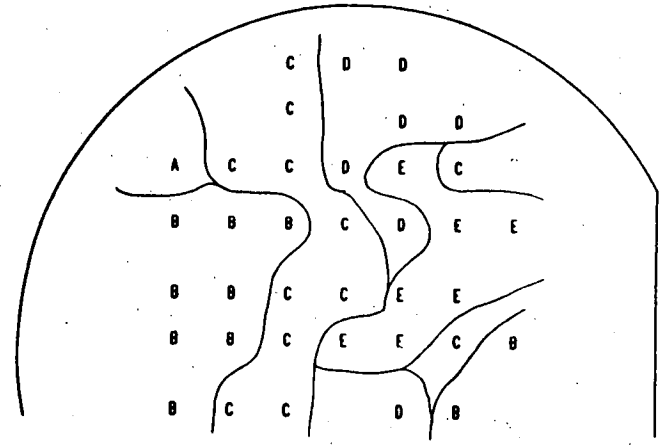


Figure 27
Variation of L_d across a Wafer

A ≥ 1.19
B ≥ 1.14
C ≥ 1.09
D ≥ 1.04
E < 1.04

A

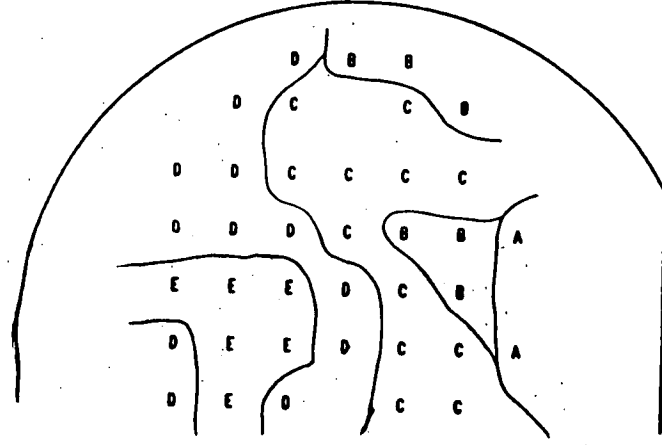


Figure 28 Parameters V Process

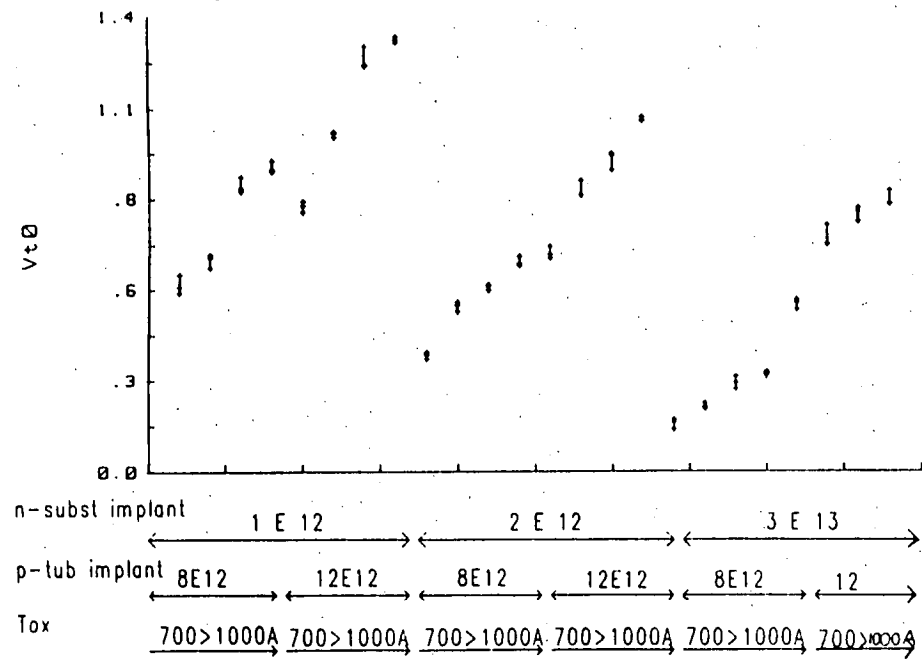
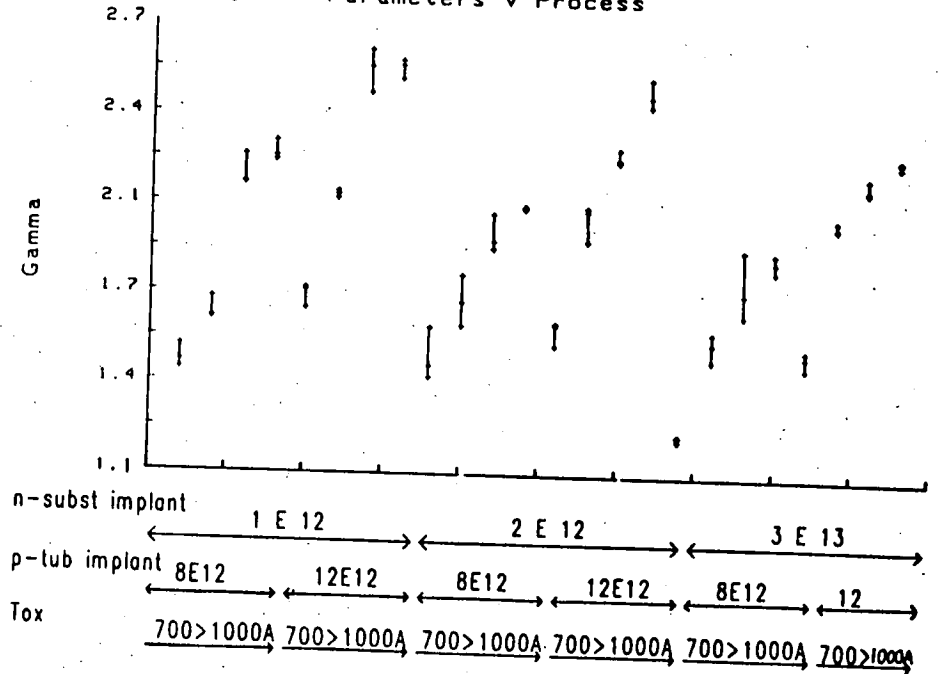


Figure 29 Parameters V Process



PARAMETRIC TESTING TO LINK DESIGN AND FABRICATION

A. Gribben, J.M. Robertson and A.J. Walton

Parametric test is currently carried out at the end of the wafer fabrication process to determine whether or not the process is within specification. These parameters are measured on drop-in test die and an assessment is then made as to whether the wafer will yield working circuits. Typically, sheet resistances, oxide capacitances, contact resistances and threshold voltages are monitored. Usually, only a value well out of specification indicates a non-working wafer and the precise effect on the operation of the circuits, in terms of output drive currents and speed of switching is unknown. By measuring parameters for the SPICE model, they can be used for both process control and circuit simulation, since they physically represent different aspects of device operation and can be used in the model to accurately simulate devices. Investigation into how these device model parameters vary with process variables such as implant doses and oxide thickness, provides a link between design and fabrication. As device geometries are scaled down, smaller process variations have larger effects on device characteristics and this link becomes increasingly important when optimum performance is sought.

Although many companies have their own circuit simulation programs, SPICE may be considered the industry standard integrated circuit simulation program and the level 3 MOSFET model [1] is the most widely used model for simulating small geometry MOS transistors. Despite the fact that SPICE has been in existence for many years, only recently have commercial software packages been readily available for extracting device parameters. All these packages use numerical optimisation. Measured values are input and after a Jacobian iteration or similar, the parameters are derived. Some programs optimise all the parameters at once e.g. SUXES produced by Stanford University and others e.g. TECAP [2] written by Hewlett-Packard are capable of optimising particular parameters in specific regions of operation separately. As a result of both of these approaches, the meaning of the parameters is not taken into account during the extraction and since there is interaction between them, there is no precise definition of the measured parameters.

The philosophy adopted here in extracting parameters for the SPICE 2 level 3 MOSFET model is to avoid using any numerical optimisation and extract parameters separately in sequence. Each parameter is extracted by the same technique and from the same set of measurements each time, therefore they can be used for process control as well as circuit simulation.

Two process parameters and eleven electrical parameters are used as input to the SPICE 2 level 3 MOSFET model and the aspects of device operation which they characterise are as follows:

Professor J.M. Robertson and Dr A.J. Walton are with the Edinburgh Microfabrication Facility, Dept of Electrical Engineering, University of Edinburgh.

Mr A. Gribben, formerly with the Edinburgh Microfabrication Facility is now with Analog Devices, Newbury, Berkshire.

t_{ox} - oxide thickness
 x_j - junction depth
 V_{to} - threshold voltage
 γ - substrate bias coefficient on threshold
 L_d - diffusion length under the gate
 η - drain voltage effect on threshold
 $\Delta\omega$ - field oxide encroachment into the channel
 θ - threshold shift with device width
 n_{fs} - subthreshold slope
 μ_0 - low field mobility
 θ - mobility degradation with gate bias
 V_{max} - mobility degradation with drain bias
 K - saturation slope

Two examples of parameter extraction are given in figures 1 and 2. Figure 1 shows the variation in threshold with drain bias and figure 2 shows how mobility varies with gate bias.

These techniques have been implemented on a HP controller and HP4145 Semiconductor Parameter Analyzer. The resulting parameters can be used to accurately simulate device operation at small geometries as shown in figure 3. Using the above approach parameter variations across a wafer and their relationship with the process have been examined.

The measurements and extraction program makes over 1000 measurements and it takes a few minutes to obtain a complete set of parameters. This approach is too time-consuming for an end of process test on every chip; this is also true of the numerical optimisation extraction programs whose mathematical manipulations take significantly longer. An advanced extraction program has been devised which requires only 27 measurement points and a parallel test system is being developed. The usual sophisticated parametric test system, including controller, instruments and switching matrix, is replaced by some simple circuitry which obtains the measurement values. These values are manipulated to obtain SPICE parameters while the prober moves to the next site. This approach makes SPICE parameter extraction feasible at every site in a production environment.

References

1. A. Vladimirescu and S. Liu, "The Simulation of MOS Integrated Circuits Using SPICE 2", UCB/ERL M80/7 (February 1980).
2. E. Khalily, P.H. Decher and D.A. Teegraden, "TECAP 2 : An Interactive Device Characterisation and Model Development System", Hewlett-Packard Report 1984.

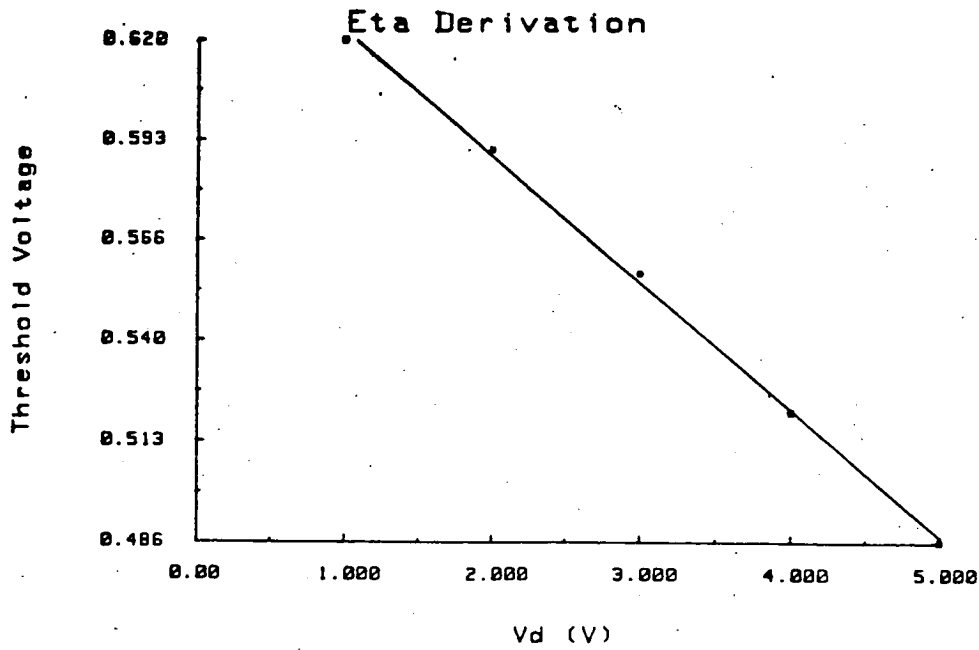


Figure 1
Variation of
Threshold with
Drain Bias

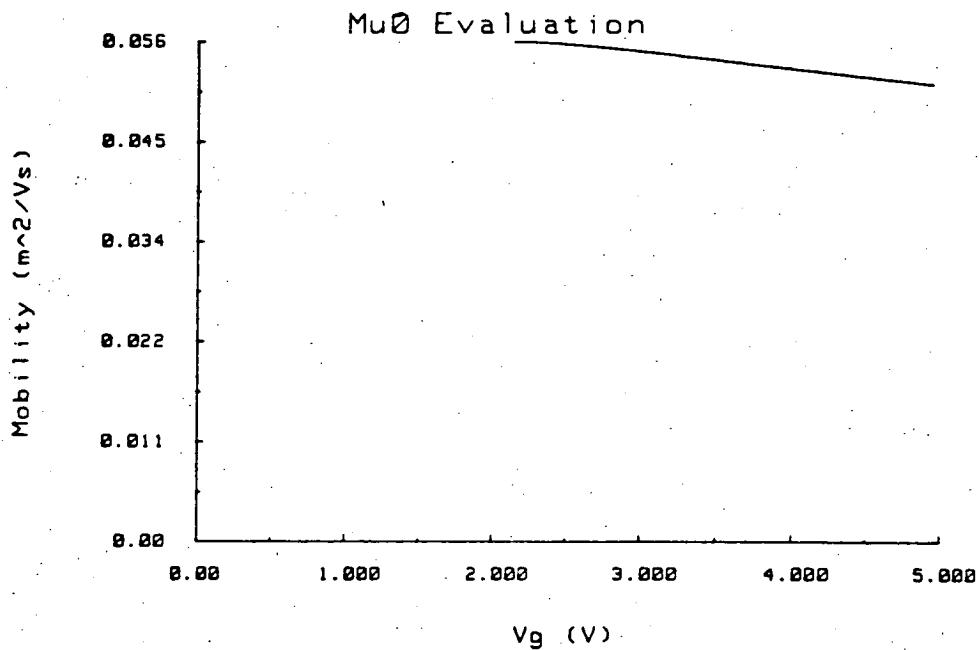


Figure 2
Relationship
Between Mobility
and Gate Voltage

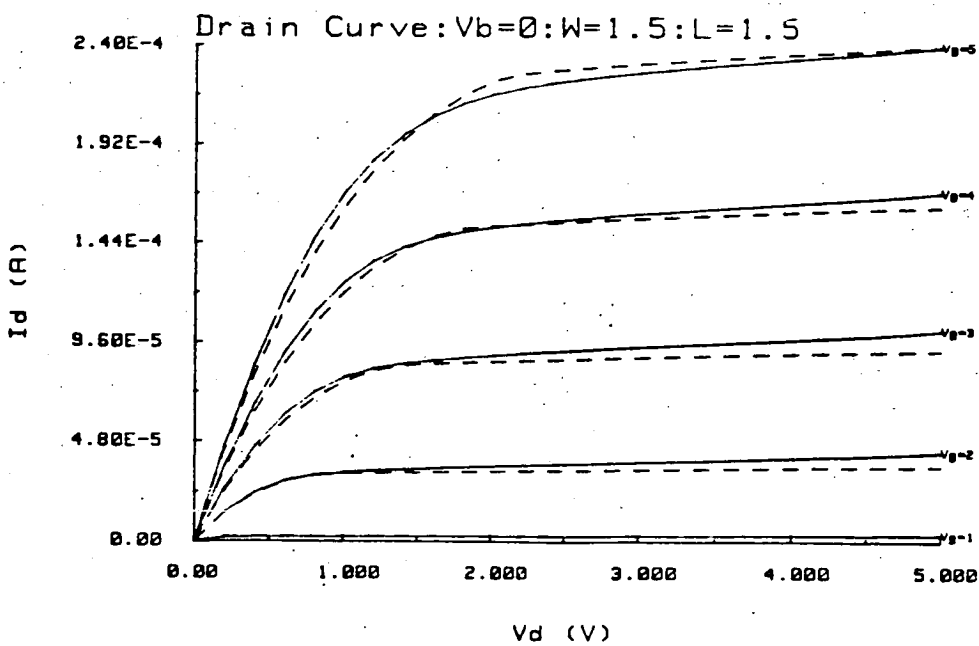


Figure 3
Comparison of
Measured and
Simulated
Operation

Realistic worst-case parameters for circuit simulation

P. Tuohy
A. Gribben
A.J. Walton
J.M. Robertson

Indexing terms: Simulation, Design, Transistors

Abstract: The spread in device performance due to random variations in the manufacturing process is usually characterised for the designer by a set of parameters representing typical transistors along with the best and worst cases. The success of a design is obviously dependent on the accuracy of these parameters and it is shown that obtaining a worst and best case set by combining the worst and best values of individual parameters gives unrealistic results. An alternative technique is proposed which results in an accurate parameter set using a single measurement to characterise transistor performance. Both methods have been used to obtain parameter sets for n- and p-channel devices.

1 Introduction

An IC design is usually first verified by simulation but the accuracy of the results depends upon the validity of the circuit model [1]. The model is linked to a specific manufacturing process through its parameters and these are obtained from transistor measurements. The model considered in this work is that implemented in level-3 SPICE [2] although the techniques discussed are more generally applicable.

Random fluctuations [3, 4] in the manufacturing process result in variations in transistor performance which must be accounted for at the design stage by simulating the circuit using sets of parameters which represent best- and worst-case performance. The accuracy of these parameters is critical to the success of the design as illustrated by Table 1. If the simulated spread in transistor performance is less than the actual spread, products will be manufactured which may not yield. Conversely, if the simulated spread is greater than the actual spread, the design task becomes unnecessarily difficult and time consuming [5, 6]. To accommodate large variations, the circuit performance often has to be compromised or the layout becomes more complex and as a consequence occupies a greater area.

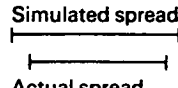
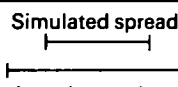
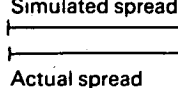
Paper 5590I (E3), first received 19th June and in revised form 5th August 1987

A.J. Walton and J.M. Robertson are, and P. Tuohy and A. Gribben were formerly, with the Edinburgh Microfabrication Facility, Department of Electrical Engineering, The King's Buildings, University of Edinburgh, Edinburgh EH9 3JL, United Kingdom

P. Tuohy is now with Motorola Ltd., Kelvin Industrial Estate, Colevilles Road, East Kilbride, United Kingdom

A. Gribben is now with Analog Devices, Rothwell House, Pembroke Road, Newbury, Berks RG13 1BX, United Kingdom

Table 1: Simulated spread of SPICE parameters

Case 1:	Simulated spread 	Manufactured device works every time
	Actual spread	Nonoptimum device performance and chip area
Case 2:	Simulated spread 	Manufactured device will not work for all processing conditions
	Actual spread	Maximum possible yield < 100%
Case 3:	Simulated spread 	Manufactured device works every time
	Actual spread	Optimum device performance, chip area and design effort

2 Simulation parameter sets from best and worst case individual transistor parameters

Using 163 sites on a wafer, full sets of level-3 SPICE parameters were extracted using PARAMEX [7]. Some statistical analysis was performed to determine mean, maximum, minimum and standard deviations (σ) for each of the parameters. To exclude any bogus data points, the extremes of each distribution were calculated as the mean $\pm 3\sigma$ and a summary of these measurements is presented in Table 2.

The contribution of each parameter to the transistor current was then classified as either positive or negative, such that, a positive sign represents an increase in source-drain current whereas a negative sign indicates the reverse. These data, together with the maximum and minimum values, are given in Table 3. These parameters, when used in the SPICE model, result in a spread in the transistor drive current of $\pm 37\%$ as illustrated in Fig. 1. The measured parameters for each of the 163 sites were also used to calculate the same characteristics and in this case the extremes of the distribution were only $\pm 14\%$ as shown in Fig. 1. Obviously using the individual best- and worst-case parameters clearly leads to a substantial overestimate of the actual process spread which, if used in the design process would result in over-conservative design.

3 Correlations between extracted transistor parameters

The reason for the overestimate of process spread is the interdependence of some transistor parameters. Only if the parameters correlate so that the worst case for each parameter occurs at the same site will the above approach provide a realistic result. If the parameters are completely independent, the probability of worst-case

Table 2: SPICE parameters summary from data measured at 163 sites

Parameter	Mean	σ	Maximum	Minimum	+3 σ	-3 σ
t_{ox} , nm	62.4	0.65			64.4	60.5
V_{to} , V	0.506	0.022	0.572	0.449	0.573	0.438
γ , $\sqrt{(V)}$	1.39	0.033	1.46	1.25	1.49	1.29
μ_o , $m^2 V^{-1} s^{-1}$	0.053	0.0025	0.059	0.048	0.060	0.046
θ , V^{-1}	0.027	0.0023	0.034	0.021	0.034	0.020
V_{max} , $\times 10^5 ms^{-1}$	1.80	0.06	1.96	1.60	1.98	1.62
N_{is} , $\times 10^{15} m^{-2}$	1.70	0.35	2.63	1.05	2.76	0.64
L_{del} , μm	1.05	0.04	1.19	0.94	1.18	0.93
ΔW , μm	0.85	0.04	0.98	0.75	0.97	0.73
δ	0.216	0.024	0.278	0.161	0.291	0.141
η	0.041	0.006	0.068	0.028	0.058	0.023
κ	0.301	0.020	0.359	0.236	0.361	0.241

Table 3: Best and worst case SPICE parameters measured from 163 sites

Parameters	Contribution to current	Best-case parameter	Worst-case parameter
t_{ox} , nm	negative	60.5	64.4
V_{to} , V	negative	0.438	0.573
γ , $\sqrt{(V)}$	negative	1.29	1.49
μ_o , $m^2 V^{-1} s^{-1}$	positive	0.060	0.046
θ , V^{-1}	negative	0.020	0.034
V_{max} , $\times 10^5 ms^{-1}$	positive	1.98	1.62
N_{is} , $\times 10^{15} m^{-2}$	positive	2.76	0.64
L_{del} , μm	positive	1.18	0.92
ΔW , μm	negative	0.73	0.97
δ	negative	0.141	0.291
η	positive	0.058	0.023
κ	positive	0.361	0.241

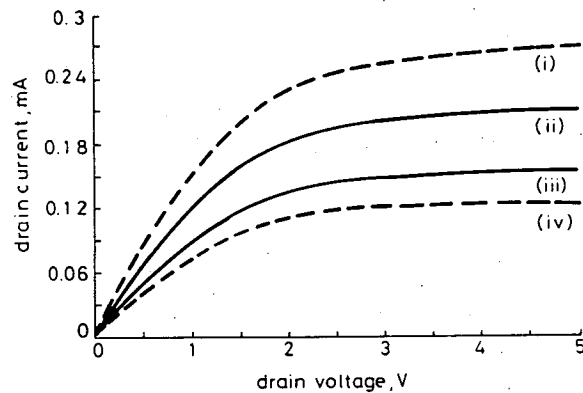


Fig. 1 Simulated characteristics for n-channel devices ($W = 5 \mu m$, $L = 5 \mu m$)

- (i) best case from best individual parameters
- (ii) best case from parameters with highest values of I_{drive}
- (iii) worse case from parameters with lowest values of I_{drive}
- (iv) worst case from worst individual parameters

parameters coinciding is very remote and the actual spread will be very much less than predicted. Correlations can exist between the parameters so that when one of them has a positive effect on current, another always makes a negative contribution. Hence the actual spread

can be less than that of independent parameters and an allowance must be made for this factor to obtain a realistic best and worst case set.

Table 4 shows the interdependence of level-3 SPICE parameters obtained by calculating a correlation factor for each pair of parameters using the same data as that used to derive Table 2. The values 0, 1 and -1 represent no correlation, perfect correlation and perfect negative correlation, respectively. The correlations between the major parameters are detailed in Table 5 together

Table 5: SPICE parameters with the most significant correlations and their contribution to current

Parameter	Contribution to current	Correlation factor	Contribution to spread
T_{ox}, L_d	opposite	0.55	
V_{to}, θ	same	-0.40	
θ, μ_o	opposite	0.55	decrease spread
$\Delta W, \theta$	same	-0.42	
L_d, κ	same	-0.90	
T_{ox}, κ	opposite	-0.50	increase spread
V_{max}, μ_o	same	0.42	

with the sense of each parameter's contribution to current. The fact that μ_o correlates positively with θ means that the extreme cases, i.e. μ_o low, θ high and μ_o high, θ low, are less likely to occur. Five of the six major correlations act to reduce the actual spread when compared with the independent parameter case. The correlations may have physical meaning or may be a consequence of the extraction techniques but they clearly indicate that actual worst-case parameter sets are not easily determined by combining the extreme values of the individual parameters.

4 Realistic parameter sets for simulation

The method which is proposed to obtain a realistic parameter set for simulation does not depend upon the combination of individual parameter values. Transistors with the best and worst performance must be first identi-

Table 4: Correlation between SPICE parameters for measurements at 163 sites on the same wafer

	V_{to}	γ	μ_o	θ	V_{max}	N_{is}	L_{del}	ΔW	η	κ	δ
t_{ox}	0.13	0.21	0.25	0.25	-0.41	0.55	0.34	0.34	-0.50		
V_{to}		0.62	-0.24	-0.40	0.13		-0.17	0.32	0.30	0.29	-0.28
γ			-0.50	-0.60	0.16			0.39	-0.49	0.23	-0.23
μ_o				0.55	0.42			0.22	-0.27	-0.10	
θ					0.21			-0.44	0.46	-0.18	0.35
V_{max}						-0.13	-0.11	0.22	-0.27		
N_{is}											
L_{del}								0.30		-0.90	0.75
ΔW										0.10	
η										-0.25	0.19
κ											-0.10

Table 6: Maximum and minimum measured parameters and their associated site numbers for the n -channel devices

Parameter	Site number with max value	Maximum measured value	Measured value at site 17	Site number with min value	Minimum measured value	Measured value at site 162
I_{drive}, A	17	2.09×10^{-4}	2.09×10^{-4}	162	1.67×10^{-4}	1.67×10^{-4}
β_o, AV^{-2}	29	3.93×10^{-5}	3.83×10^{-5}	162	3.04×10^{-4}	3.04×10^{-4}
I_{dsat}, A	17	1.97×10^{-4}	1.97×10^{-4}	162	1.58×10^{-4}	1.58×10^{-4}
β, AV^{-2}	29	3.40×10^{-5}	3.36×10^{-5}	162	2.74×10^{-5}	2.74×10^{-5}

Table 7: Maximum and minimum measured parameters and their associated site numbers for the p -channel devices

Parameter	Site number with max value	Maximum measured value	Measured value at site 80	Site number with min value	Minimum measured value	Measured value at site 141
I_{drive}, A	80	7.19×10^{-5}	7.19×10^{-5}	141	5.58×10^{-5}	5.58×10^{-5}
β_o, AV^{-2}	80	1.09×10^{-5}	1.09×10^{-5}	141	8.39×10^{-6}	8.39×10^{-6}
I_{dsat}, A	80	7.02×10^{-5}	7.02×10^{-5}	3	5.12×10^{-5}	5.41×10^{-6}
β, AV^{-2}	80	9.64×10^{-6}	9.64×10^{-6}	3	7.18×10^{-6}	7.71×10^{-6}

fied, and the parameters only measured for these devices. To identify the appropriate transistors, a single parameter which represents transistor performance is required and several were investigated.

(a) I_{drive} = drain current for gate and drain held at +5 V (-5 V for p -channel).

(b) I_{dsat} = saturation drain current

$$(c) \beta_o = \mu_o C_{ox} \frac{W}{L} \quad (1)$$

$$(d) \beta = \frac{\beta_o}{1 + \theta(V_{GS} - V_{i0})}, V_{GS} = 5 V \quad (2)$$

Each of the above parameters was evaluated at all 163 sites and Tables 6 and 7 give the site numbers of the maximum and minimum measured values for the n and p -channel devices, respectively. Table 8 shows the mean and spread of the n -channel parameters and Table 9 gives

Table 8: Performance parameters, data from 163 sites

Parameter	Mean	σ	+3 σ	-3 σ
$I_{drive}, \mu A$	185.0	8.03	209.1	160.9
$\beta, \mu AV^{-2}$	33.8	1.61	38.63	28.97
$I_{dsat}, \mu A$	174.4	7.51	196.9	151.9
$\beta_o, \mu AV^{-2}$	30.1	1.21	33.73	26.47

Table 9: Correlation matrix

	I_{drive}	I_{dsat}	β_o	β
I_{drive}		0.999	0.967	0.956
I_{dsat}	0.999		0.965	0.954
β_o	0.967	0.965		0.989
β	0.956	0.954	0.989	

their correlation factors, and it can be observed that all four parameters show strong correlations.

I_{drive} was picked as the best measure of performance for the following reasons. It is the only parameter which depends on all of the level-3 SPICE parameters and this is particularly important for small geometry devices where the influence of V_{max} on device performance becomes more significant [7]. (Only the parameters I_{drive} and I_{dsat} depend on V_{max} .) Also, I_{drive} is easily measured as well as relating closely to transistor operation in typical IC circuit designs as illustrated in Fig. 2.

To obtain realistic parameter sets, with the minimum number of measurements, the following method is proposed:

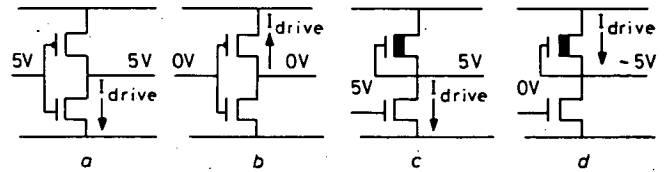


Fig. 2 I_{drive} during typical transistor operation

- a CMOS, n -channel
- b CMOS, p -channel
- c NMOS, enhancement
- d NMOS, depletion

(i) I_{drive} , V_{i0} and β_o should be determined. β_o can be extracted using the level-1 SPICE equation from the V_{i0} measurement:

$$I_D = \beta_o \left((V_g - V_{i0}) V_D - \frac{V_D^2}{2} \right) \quad (3)$$

(ii) exclude data for transistors with I_{drive} values outside the $\pm 3\sigma$ range

(iii) find maximum and minimum values of I_{drive} from the remaining datapoints

(iv) check that β_o for these transistors is within $\pm 5\%$ of its extreme values. If not choose the 'next worse' transistors

(v) extract model parameters from these transistors

(vi) adjust μ_o so that I_{drive} for the extracted best and worst case parameters is exactly $I_{drive} \pm 3\sigma$. (This adjustment should be small.)

The extracted values of V_{i0} in the best and worst parameters sets will be close to -3σ and $+3\sigma$, respectively, for the V_{i0} distribution. As V_{i0} is uniquely significant to circuit operation the -3σ and $+3\sigma$ values should be substituted into the best and worst parameter set to ensure the full spread of transistor characteristics is reflected by the parameter set. It should be noted that the inclusion of β_o in the extraction procedure is to ensure that the chosen transistor is behaving in the expected manner.

The above method has been used to extract realistic worst-case parameter sets for both n - and p -channel transistors fabricated using silicon-gate CMOS technology. Their parameter sets are listed in Tables 10 and 11 and the characteristics illustrated in Figs. 1 and 3. The worst and best case characteristics, calculated from the parameters extracted in the above manner, were found to have excellent agreement with those derived from direct measurements at each of the individual sites.

Table 10: Best and worst case SPICE parameter set for the *n*-channel devices

Parameter	Worst case	Best case
I_{drive} , A	1.61×10^{-4}	2.09×10^{-4}
V_{t0} , V	0.530 ($3\sigma = 0.573$)	0.451 ($3\sigma = 0.438$)
γ , \sqrt{V}	1.46	1.30
μ_o , $m^2 V^{-1} s^{-1}$	0.0495	0.059
θ , V^{-1}	0.0243	0.0305
V_{max} , ms^{-1}	1.82×10^5	1.78×10^5
N_{ts} , m^{-2}	2.34×10^{15}	
L_{del} , m	1.05×10^{-6}	1.07×10^{-6}
ΔW , m	9.16×10^{-7}	8.38×10^{-7}
δ	0.192	0.235
η	0.0283	0.0542
κ	0.318	0.283
t_{ox} , m	6.76×10^{-8}	6.19×10^{-8}

Table 11: Best and worst case SPICE parameter set for the *p*-channel devices

Parameter	Worst case	Best case
I_{drive} , A	5.58×10^{-5}	7.19×10^{-5}
V_{t0} , V	-0.650 ($3\sigma = -0.686$)	-0.594 ($3\sigma = -0.577$)
γ , \sqrt{V}	1.29	1.22
μ_o , $m^2 V^{-1} s^{-1}$	0.0152	0.0177
θ , V^{-1}	0.0199	0.0295
V_{max} , ms^{-1}	0	0
N_{ts} , m^{-2}	5.10×10^{15}	2.94×10^{15}
L_{del} , m	1.01×10^{-6}	1.08×10^{-6}
ΔW , m	8.92×10^{-7}	7.52×10^{-7}
δ	0.265	0.342
η	0.0366	0.0506
κ	0.0409	0.0222
t_{ox} , m	6.76×10^{-8}	6.90×10^{-8}

5 Conclusions

A method of deriving a realistic worst and best case parameter set has been presented. A significant advantage of the proposed technique is that it can be implemented so that only I_{drive} , β_o and V_{t0} need to be monitored at every site. As a consequence full parameter extraction is now only necessary for a small number of selected transistors and hence the associated computation (and measurement) is reduced by more than two orders of magnitude.

In an IC production plant, process control parameters are measured after fabrication using a parametric tester and I_{drive} , V_{t0} and β_o are often routinely monitored. This

makes the approach of using I_{drive} to identify the transistors which should be fully measured to extract the worst-best case parameters a very attractive proposition in the product environment. Once a set of parameters have been derived for a stable process it is then only necessary from the process control viewpoint to monitor I_{drive} and only extract model parameters when a change in the distribution of I_{drive} is observed.

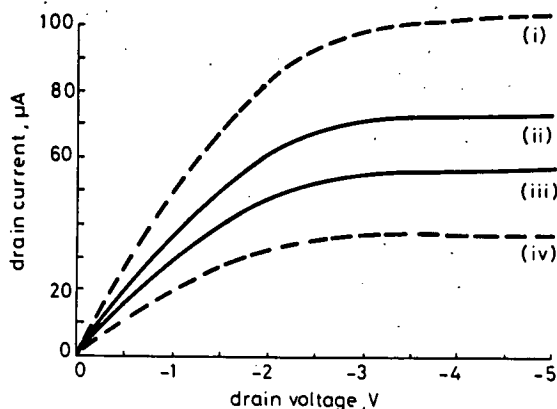


Fig. 3 Simulated characteristics for *p*-channel devices ($W = 5 \mu m$, $L = 5 \mu m$)

- (i) best case from best individual parameters
- (ii) best case from parameters with highest values of I_{drive}
- (iii) worse case from parameters with lowest values of I_{drive}
- (iv) worst case from worst individual parameters

6 References

- 1 WRIGHT, G.T.: 'A simple and continuous MOSFET model', *IEEE Trans.*, 1985, ED-32, (7), pp. 1259-1262
- 2 VLADIMERESCU, A., and LIU, S.: 'The simulation of MOS integrated circuits using SPICE 2'. UCB/ERL M80/7, 1980
- 3 SPENCE, R.: 'Parametric yield maximisation'. Proceedings of Electronic Automation Conference, Wembley, UK, 1987, pp. 477-481
- 4 YANG, P., HOCEVAR, D.E., COX, P.F., MACHALA, C., and CHATTERJEE, P.K.: 'An integrated and efficient approach for MOS VLSI statistical circuit design', *IEEE Trans.*, 1986, CAD-5, (1), pp. 5-14
- 5 SILBURT, A.: 'Linking MOSFET design and fabrication', *Microelectron. Manuf. & Test.*, 1986, 9, (5)
- 6 COX, P., YANG, P., MAHANT-SHETTI, S.S., and CHATTERJEE, P.: 'Statistical modeling for efficient yield estimation of MOS VLSI circuits', *IEEE Trans.*, 1985, ED-32, (2), pp. 471-478
- 7 GRIBBEN, A., ROBERTSON, J.M., and WALTON, A.J.: 'Accurate physical parameter extraction for small geometry devices'. SEMICON, Birmingham, UK, 1986

A REVIEW OF PARAMETRIC TESTING

*A.J. Walton
Department of Electrical Engineering
Kings Buildings
Edinburgh University
Edinburgh
EH9 3JL*

*A. Gribben
Analog Devices
Rothwell House
Pembroke Road
Newbury
Berks
RG13 1BX*

1. INTRODUCTION

Process control chips (PCC) are used to give parametric information for both the process engineer and designer. They may include a range of transistors with varying geometries, MOS capacitors, and a number of specialized devices for measuring many of the macroscopic parameters which are traditionally used for process verification. Process control chips containing these structures are used to monitor the engineering yield of wafers by using them as drop ins in the place of product circuits. Figure 1 shows how pccs may be located as drop ins on a typical wafer.

2. DESIGN CONSIDERATIONS FOR A PROCESS CONTROL CHIP

Early PCCs consisted of test structures on small die with the contact pads situated around the edge of the chip in a similar manner to that found on product circuits. With the trend to larger circuits it became no longer feasible to have all the contact pads in this position because, as the number of test devices increase, so does the number of probe pads. A point can be reached where, it becomes not only very expensive to make up the probe cards but physically impossible. This makes it necessary, for example, to divide the chip into four identical probing patterns and step the probes internally. A more flexible approach is to use a matrix of equally spaced pads with the test devices constructed around them [1]. A 2 x N probe card can then be used to probe any device on the chip and figure 2 shows an example of such a PCC.

One item which needs to be considered is the size of the contact pads. When devices are tested it is important that contact problems with the probes should be kept to an absolute minimum. It is recommended that pads should be at least 80 microns square to give a maximum silicon to pad area ratio while allowing a suitable margin of error [1].

While the use of the 2 x N type structure is very good from the point of view of flexibility it does suffer because testing times are increased due to the extra prober movement within the chip. It is for this reason that many PCCs used as drop ins do not use this approach. To increase packing density the number of probe pads are often reduced by using single pads for a number of devices. This increases the area of silicon available for the product circuits and also speeds up testing times. Both of these factors are important considerations in a production environment.

In recent years there has been a trend to use either test structures located within the chip area or scribe channels [2]. This approach has been as a direct result of the introduction of wafer stepper technology which may require a mask change for exposing the PCC. The time taken changing over masks reduces throughput on a very expensive piece of equipment and as a result the use of dropins becomes less attractive. Figure 3 gives an example of test structures located within scribe channels. This approach has some advantages. It enables wafer mapping of measurements which can be used to investigate uniformity across wafers which is not possible with the drop in shown in figure 1. Secondly in many cases it proves possible to locate test structures without any loss of silicon real estate which maximizes the number of product circuits per wafer.

3. TEST STRUCTURES

The PCC can contain a large number of structures and there now follows a description of some of the more important ones.

3.1 VAN der PAUW DEVICES (GREEK CROSSES)

Van der Pauw [3] developed a theory enabling the measurement of the resistivity of a continuous surface of arbitrary shape. The greek cross [4] structure shown in figure 4 is a refinement and is ideally suited to the PCC. Furthermore it is an easy structure to layout and define photolithographically. The measured sheet resistance is at the heart of the cross and an accuracy of better than 0.1% can be achieved in practice.

To use the greek cross to obtain an approximate sheet resistance (within 1%) current is passed between contacts A and B (I_{AB}) and the potential different between pads D and C (V_{DC}) measured

$$R = \frac{V_{DC}}{I_{AD}} \quad (1)$$

$$R_s = \frac{\pi R}{\ln 2} \quad \Omega/\square \quad (2)$$

The assumptions used are exactly the same but, it can be observed that current is not forced and voltage measured at an infinitely small point. This proves not to be a problem provided that the length of the arms is greater than or equal to the size of the heart of the cross. In this case the error from this approximation will be less than 0.1%. A more accurate measurement is force I_{AB} and measure V_{DC} , and then reverse the current and measure the voltage V_{CD} . The resistance is then calculated as follows.

$$R_{0^\circ} = \frac{V_{DC} - V_{CD}}{I_{AB} - I_{BA}} \quad (3)$$

Current is then forced between B and C and the process repeated to give

$$R_{90^\circ} = \frac{V_{AD} - V_{DA}}{I_{BC} - I_{CB}} \quad (4)$$

The average resistance is

$$R = \frac{R_{0^\circ} + R_{90^\circ}}{2} \quad (5)$$

from which the sheet resistance can be derived.

$$R_s = \frac{\pi R}{\ln 2} \quad \Omega/\square \quad (6)$$

It is possible to evaluate parameters which illustrate the degree of asymmetry of the structure, the degree of offset in the measuring equipment and the linearity of the current voltage characteristic of the conductor. The greek cross can be used with a wafer mapping technique to allow the uniformity of an implant dose to be assessed to within 1%. Measurement currents must be kept low especially for diffused structures (<0.1mA) or silicon will show its positive TCR and surface leakage currents may affect the measurement [5].

Reference [6] describes a structure which may be used for measuring the bulk resistivity of a silicon substrate which may be useful in special situations. (This measurement can be made using a four point probe before processing starts).

3.2 LINE WIDTH MEASUREMENTS

A design [7] which may be used for the measurement of conductor linewidth is shown in figure 5. The greek cross at one end of the structure is used to evaluate the resistivity of the conducting layer. Current is then passed between C and D and the voltage measured between A and B. The resistance R_{AB} can then be calculated from which the linewidth may be evaluated.

$$W = \frac{R_{AB}}{R_s L} \quad (7)$$

This technique can be used to measure the etch uniformity of polysilicon, metal, as well as sideways diffusion. It is sensitive to changes in dimension of $\pm 0.1 \mu\text{m}$ provided that the width of the voltage taps are not greater than the trackwidth. This, of course, assumes that etching is uniform and that variation of sheet resistivity with position is negligible.

3.4 ALIGNMENT ACCURACY

The masks which are used in integrated circuit fabrication are aligned to one another using special patterns which are present on each mask. Alignment errors can result from incorrect stepping on the mask, faulty alignment machines, operator error or wafer distortion. A close analysis of the spatial distribution of these errors can lead to the isolation of their source. With the advent of VLSI technology where improved alignment is required this type of measurement is becoming more important.

A structure [8] which can be used for measuring the superposition error between a conductor and contacts is shown in figure 6. A current is passed between the contact pads I_1 and I_2 . The voltage taps are spaced equal distances apart with V_1 , V_2 and V_4 defined by the conductor mask and V_3 by the contact mask. By measuring these voltages the superposition error can be calculated in the direction of the long axis of the conducting bar. For no misalignment

$$(V_4 - V_3) = (V_3 - V_2) = (V_2 - V_1) \quad (8)$$

The difference in voltage ΔV is given by

$$\Delta V = (V_1 - V_2) - (V_2 - V_3) \quad (9)$$

$$\Delta V = (V_1 - V_2) - (V_3 - V_4) \quad (10)$$

The sign of ΔV resulting from equations (9) and (10) will depend upon the direction of the superposition error. If the spacing between the taps is S the error in alignment is given by

$$e = \frac{\Delta V S}{(V_4 - V_3)} \quad (11)$$

A similar structure at 90° will give the other component of the superposition error.

Figure 7 shows a structure for measuring the misalignment between polysilicon and diffusion for a self aligned process [9]. This consists of two diffused resistors whose widths are dependent upon the degree of polysilicon misalignment. The ratio of these two resistances gives the degree of misalignment with once again two structures at 90° being required to give both components.

3.4 STEP COVERAGE

The step coverage of metals is especially important for VLSI circuits as is the integrity of interlayer dielectrics. Problems in this area are a major yield reducing factor and hence structures which identify these types of fault are of major interest. Figure 8 gives an example of a structure for measuring the effectiveness of aluminium step coverage. They are typically tapped at

increasingly long intervals to account for differing levels of yield. In the case of multilevel circuits there are also step coverage difficulties with metal over metal and figure 9 gives an example of a structure which highlights this problem. It also gives a measure of interlayer dielectric integrity. Interdigitated structures can also be designed to check the quality of etching over stepped surfaces [10].

3.5 CONTACT CHAINS

Contacts chains can be used to evaluate the number of contacts which can be used in a circuit design while still providing an acceptable yield. Figure 10 gives an example of a metal to polysilicon contact chain. These are tapped in a similar manner to the step coverage structures and by their nature take up a large area of the PCC.

3.6 CONTACT RESISTANCE

A knowledge of contact resistance is important and can be measured using the structure of figure 11. The sheet resistance is then given by [11]

$$R_s = (R_1 - R_2) \frac{W}{L_1 - L_2} \quad (12)$$

and the contact resistance (R_c) is

$$R_c = \frac{R_2 L_1 - R_1 L_2}{2(L_1 - L_2)} \quad (13)$$

These equations are valid if

$$R_c \approx R_f = \frac{1}{W} R_s \rho_c \quad (14)$$

where ρ_c is the specific contact resistivity and R_f is the front contact resistance. This is usually the case for $W > 5$ microns. One of the disadvantages of the above structure is that the measurement of R_c depends upon the subtraction of two large numbers. It also includes the effect of parasitic resistances and cannot take into account the change of resistivity in the diffused layer underneath the contact. The four terminal structure of figure 12 overcomes some of these problems [12].

A current is forced between pads 1 and 3 and the voltage measured between pads 2 and 4. The current is then reversed and the measurement performed again. The voltage and current pads are then interchanged and the above measurements repeated. These four values of resistance are then averaged. This Kelvin type measurement gives the interfacial contact resistance (R_c) from which the specific contact resistivity can be derived.

$$\rho_c = \frac{R_c}{A} \quad (15)$$

where A is the area of the contact. This assumes that the contact resistance is uniform across the whole contact area which may not always be the case [12].

3.7 OTHER DEVICES

No process control chip would be complete without capacitors and transistors of various dimensions. They can be used to give information such as doping profiles, SPICE parameters, oxide thickness as well as the basic transistor characteristics. References [13-15] give an indication of some of the measurements which can be made.

4. MEASUREMENT OF THRESHOLD VOLTAGE

The measurement of threshold voltage is complicated by the fact that MOS transistors do not abruptly turn on [15] as illustrated in figure 13. The exact measurement technique which should be used is dependent upon whether the measured value is required for just process monitoring or for use in a circuit simulator such as SPICE [16]. The methods which are to be described will all yield values within 100 - 200 mV of each other.

Figure 14 shows a configuration for measuring threshold voltage which is ideally suited to process monitoring. The measurement is very quick because it just requires a current to be set and a single voltage measurement performed provided that V_T is defined as the gate voltage required for a given source-drain current. This is an arbitrary value typically set to 1 μ A but it should be recognized that V_T will vary for transistors with different dimensions and therein lies its limitations.

Another approach is to measure the gate voltage at two set values of I_{DS} . A line is then drawn through the two measured points and its intersection with $I_{DS} = 0$ gives the threshold voltage as illustrated in figure 15. This method assumes that the transistor is in saturation in which case

$$I_{DS} = \frac{\beta}{2} (V_{GS} - V_T)^2 \quad (16)$$

$$\sqrt{I_{DS}} = \sqrt{\frac{\beta}{2}} (V_{GS} - V_T) \quad (17)$$

The slope of I_{DS} vs V_{GS} shown in figure 15 is then given by

$$\frac{d\sqrt{I_{DS}}}{dV_{GS}} = \sqrt{\frac{\beta}{2}} \quad (18)$$

This justifies measuring just two values provided that β is a constant. The assumption is reasonable provided the points used are in the linear portion of the characteristic shown in figure 15. The advantage of this method over the previous one is that the same threshold voltage results from measurements made on transistors with different dimensions. It is also possible to make a number of measurements in the linear region and then fit a line using linear regression in order that single measurements are not so heavily relied upon.

A similar alternative to this approach is to use what is known as the 10-40 method. The measurement circuit is that of figure 16. V_{DS} is set to 5 volts and V_{GS} is adjusted twice to obtain drain source currents of 10 and 40 μ A. The choice of these values is not critical provided both are in the linear portion of the characteristic of figure 13 and their ratio is 4:1. These requirements are due to the calculation which is used to derive the threshold voltage and is justified below. If the transistor is in saturation and equation (16) applies then substituting the values of I_{DS} together with their associated gate voltages and then dividing the two equations gives

$$\frac{10}{40} = \left(\frac{V_{10} - V_T}{V_{40} - V_T} \right)^2 \quad (19)$$

$$\frac{1}{2} = \frac{V_{10} - V_T}{V_{40} - V_T} \quad (20)$$

$$V_T = 2V_{10} - V_{40} \quad (21)$$

The above 10-40 method is not ideally suited to process monitoring because a search routine must be used to find V_{10} and V_{40} which is a time consuming procedure for automatic measurement systems.

If the measurement circuit of figure 12 is used then just two measurements are required and this significantly reduces testing times. Unfortunately the two techniques do not necessarily result in the same threshold voltage. The reason for this can best be described using figure 17. Provided that there is no channel length modulation so that the slope in the saturation region is zero then

the same current will result for $V_{DS} = V_{GS}$ and $V_{DS} = 5$ volts. If however, the saturation region has a slope then differences in the threshold voltage of hundreds of millivolts can result.

These types of technique are ideally suited for process monitoring where the exact definition of V_T is not so critical because the measurement can be used in a differential mode. What is more important is the consistency of the measurement technique.

Measurements for use in circuit simulators are different in that they require the measurement to be consistent with the equations which describe the transistor operation. The next example uses the SPICE definition of V_T and the measurement circuit is once again that of figure 16. The equation used in SPICE which relates I_{DS} to V_{GS} is

$$I_{DS} = \beta \left(V_{GS} - V_T - \frac{V_{DS}}{2} \right) V_{DS} \quad (22)$$

Equation (22) is only valid for small values of V_{DS} which is set to 0.1V and the gate voltage is then swept from zero volts to a value well above V_T to obtain the characteristic shown in figure 18. The threshold voltage is calculated by fitting a tangent to the curve where the slope is a maximum. This is the position where the transconductance is a maximum as the device turns on and the mobility will not be degraded. Rearranging equation (22) gives

$$\frac{I_{DS}}{\beta V_{DS}} = V_{GS} - V_T - \frac{V_{DS}}{2} \quad (23)$$

when $I_{DS} = 0$

$$V_T = V_{GS} - \frac{V_{DS}}{2} \quad (24)$$

The value of V_{GS} in equation (24) is given by the tangent when $I_{DS} = 0$ from which the threshold voltage can be calculated.

5. EXTRACTION OF PARAMETERS FOR CIRCUIT MODELLING

The simulation of circuit operation using programs such as SPICE is of crucial importance for the successful design of integrated circuits. The input files for these programs use data which has been extracted from PCCs. This requires that the PCC must contain transistors with different dimensions. The following description gives one method of extracting some of the more major parameters which have previously been discussed. It should be noted that while the equations used throughout this paper are very similar to those used in SPICE there are differences which alter the exact sequence of extraction.

The first requirement for this procedure is the oxide capacitance (C_{ox}) which can be obtained either from process control measurements or from an MOS capacitor on the PCC. The next step uses of a large transistor (30 x 30 μm) so that any edge effects do not significantly affect the measurement. The threshold voltage is measured for a number of different substrate biases (V_{BS}) as shown in figure 19 to enable ϕ_B , N_{sub} , V_{FB} and γ to be extracted. The following equations are required.

$$V_T = V_{FB} + 2\phi_B + \gamma\sqrt{2\phi_B - V_{BS}} \quad (25)$$

where

$$\gamma = \frac{2q\epsilon_0\epsilon_{si}N_{sub}}{C_{ox}} \quad (26)$$

$$\phi_B = \frac{kT}{q} \ln \frac{N_{sub}}{n_i} \quad (27)$$

The threshold characteristics are used to derive V_T for different substrate biases as described in

section 4. Then ϕ_B is calculated by setting N_{mb} to 10^{20} atoms cm^{-3} and substituting this value into equation (27). The threshold voltage is then plotted against $\sqrt{2\phi_B - V_{BS}}$ as shown in figure 20 and from equation (12.25) it can be observed that the slope is given by γ . This value of γ is then used to calculate N_{mb} using equation (26). The above procedure is repeated by substituting this new value of N_{mb} into equation (27) and repeating the iteration until N_{mb} and ϕ_B cease to change significantly.

The next step is to find the effective length and width of the transistors using the following two equations.

$$I_{DS} = \beta \left(V_{GS} - V_T - \frac{V_{DS}}{2} \right) V_{DS} \quad (28)$$

$$\beta = \mu_{eff} \frac{L_{eff}}{W_{eff}} C_{ox} \quad (29)$$

This extraction requires the threshold characteristics for a number of very wide transistors (30 μm) with a number of different lengths to be measured. This is performed with V_{DS} is set to a low value with no substrate bias so equation (28) is valid. From these measurements the value of β can be calculated for each transistor. If $1/\beta$ is plotted against L_m (the mask length) as shown in figure 21 then the effective length (L_{eff}) can be deduced.

$$L_{eff} = L_m - \Delta L \quad (30)$$

The effective width (W_{eff}) can be extracted in a similar manner using a number of very long transistors (30 μm) with various widths. The mask dimension (W_m) is then plotted against β as shown in figure 22 from which the effective width can be calculated.

$$W_{eff} = W_m - \Delta W \quad (31)$$

The next stage is to evaluate the mobility at zero field (μ_0). The threshold voltage characteristics of a transistor with typical dimensions are measured. A graph of μ_{eff} against V_{GS} is plotted as shown in figure 23 for values of $V_{GS} \gg V_T$. The value of μ_0 is given by extrapolating the line to V_{T0} when the mobility is at its maximum.

The mobility reduction with increased transverse electric field can be observed in figure 23 and is modelled by an empirical factor θ where

$$\mu_{eff} = \frac{\mu_0}{1 + \theta(V_{GS} - V_T)} \quad (32)$$

If μ_0/μ_{eff} is plotted against V_{GS} then from equation (32) θ is given by the slope which is illustrated by figure 24.

The following technique illustrates the type of procedure by which SPICE parameters are extracted. The accuracy of each parameter is dependent upon the accuracy of previous parameters and the assumptions made (eg. equation (22) is only valid for low values of V_{DS}). Extraction techniques for other parameters are outlined in reference [16].

SPICE parameters are measured using a number of assumptions and accuracy can often be improved by using optimisation techniques which curve fit the model parameters to the transistor characteristics. This procedure is readily implemented using programs such as SUXES [17] or SIMPAR [18]. If this approach is taken then the parameters obviously lose some of their physical meaning. With geometries being reduced many of the models are no longer adequate and much effort is now being devoted to their improvement.

6. PRESENTATION OF RESULTS

The results from parametric testing can be greatly enhanced by the measurements being presented in the correct manner. The spread of parameters such as V_T can be represented by bar charts but this contains no spatial information. With the emphasis for VLSI being for greater uniformity wafer mapping has become more important [19]. This trend has also been evident with much recent process control equipment [20]. Figure 25 shows the type of wafer map which can be used to represent the misalignment across a wafer. Other parameters can be displayed as contours or surface plots which together with the judicious use of colour can be used to help isolate the source of non-uniformity from the spatial distribution.

7. EQUIPMENT USED FOR MEASURING PCC'S

Ideally the system used to measure the PCC should be fast, accurate, flexible and simple to use. Obviously any system is a compromise between these parameters and the cost, with the importance of each one depending upon the function the equipment is required to perform. References [20-22] give a comparison of some of the systems which are now available. The three basic options as to the type of system are given below.

7.1 SINGLE USER

This is potentially the cheapest option and an example of such a system is given in figure 26. The desktop computer is used to control a set of instruments via the IEEE bus and a number of relays are used to connect them to the appropriate probes. Unless a full switching matrix is used, and that is very expensive, each set of measurements will require some manual intervention to link the instruments via relays to the probe card. Once these connections have been wired the relays can then be switched under program control to connect the instruments to the desired pins. The advantages of this system is that it is not too expensive and that instruments may be removed or new ones added making the configuration very flexible. One of the disadvantages is that the manual interconnect is both time consuming and error prone. As a result small changes are awkward. Another disadvantage is that while the system is being used for testing no software development can take place.

7.2 MULTIUSER SYSTEM USING A SWITCHING MATRIX

Figure 27 shows the configuration of a multi-user system which uses a computer and a single set of instruments which may be multiplexed to the appropriate probe station. To be able to do this some sophisticated software is used to control the now mandatory switching matrix. This type of system can support a number of users either developing programs or testing devices because all instrumentation is software switchable to any probe. If a large number of test stations are used then the measurement time per wafer increases due to the time taken switching relays and waiting for instruments to become available. Sources of error are due to voltage drops in both the matrix and the leads to the probe card, and noise from the matrix. Shielding and guarding along with Kelvin type connections keeps these problems to a minimum.

7.3 SYSTEM WITH INSTRUMENTATION IN CLOSE PROXIMITY TO THE PROBES

Some of the problems of noise in low level measurements may be eliminated by placing the sensitive measuring instruments near to the probes as shown in figure 28. This is more expensive with each test station having its own measurement instrumentation. Another approach shown in figure 29 eliminates the switching matrix by giving each pin its own stimulus and measurement device [23]. This is even more expensive but allows many tests to be made simultaneously which reduces measurement time.

8. FUTURE

Process control chips provide information on the process after all the steps have been completed. The information which can be obtained from them can be quite diverse depending upon both the process and the environment. The design of a PCC is not only a function of the process and the information required but also the equipment which is to be used in the measurement. This means that design for testability is as important for PCCs as it is for integrated circuits. This factor will in the future, perhaps lead to the incorporation of some of the testing system on the chip [24], by using digital test structures [25] and on chip switching [26].

With automated testing the amount of data which may be gathered is virtually unlimited. The storage, access, and presentation of this data has now opened up a whole new field which must be developed to allow us to use the measured data to its full extent. The problem of how to correlate data from parametric test, in-process measurement and functional test is at present being addressed by the software packages being used for computer aided manufacture.

REFERENCES

1. M.G. Buehler, "Comprehensive Test Patterns with Modular Test Structures: The 2 by N Probe Pad Array Approach", Solid State Technology, pp 74-89, Oct 1979.
2. C. Alcorn, D. Dworak, N. Haddad, W. Henley, P. Nixon, "Kerf Test Structure Designs for Process and Device Characterization", Solid State Technology, May 1985, pp 229-235.
3. L.J. Van der Pauw, "A Method of Measuring Specific Resistivity and Hall Effects of Discs with Arbitrary Shape", Phillips Res. Rep., Vol 13, Jan 1958, pp 1-9.
4. W. Vernsel, "Analysis of the Greek Cross, A Van der Pauw Structure with Finite Contacts", Solid State Electronics, Vol 22, pp 911-914
5. M.G. Buehler, "An Experimental Study of Various Cross Sheet Resistor Test Structures", Journ. Electrochem. Soc., Vol 145, 1978, pp 645-650.
6. M.G. Buehler, "A Planar Four-Probe Test Structure for Measuring Bulk Resistivity", IEEE. Trans. on Electron Devices, Vol ED-23, 1978, pp 968-974.
7. M.G. Buehler, "Bridge and Van der Pauw Sheet Resistors for Characterizing the Line Width of Conducting Layers", Journ. Electrochem. Soc., Vol 145, 1978, pp 650-654.
8. T.J. Russel, T.F. Leedy, R.L. Mattis, "A Comparison of Electrical and Visual Alignment Test Structures for Evaluating Alignment in Integrated Circuit Manufacture", IEDM Technical Digest, Dec 1977, pp 7A-7F.
9. I.J. Stemp, K.H. Nicholas, H.E. Brockman, "Automatic Testing and Analysis of Misregistrations Found in Semiconductor Processing", IEEE Trans. Electron Devices, Vol ED-26, no 4, April 1979, pp 729-732
10. C.N. Alcorn, "VLSI Multilevel Wiring Monitor", IEEE VLSI Multilevel Interconnect Conference, New Orleans, May 1984, pp 252-258.

11. H.H. Berger, "Contact Resistance on Diffused Resistors", IEEE International Solid State Circuits Conference, Penns. USA, 1969, pp 160-161.
12. S.J. Procter, L.W. Lindholm, J.A. Mazer, "Direct Measurement of Interfacial Contact Resistance, and Interfacial Contact Layer Uniformity", IEEE Trans ED, Vol ED-30, no 11, Nov 1983, pp 1535-1542.
13. C.G. Shirley, "A Computer-Controlled CV Characterization System", Semiconductor International, July 1982, pp 81-97.
14. M.G. Buehler, "Dopant Profiles Determined from Enhancement- Mode MOSFET dc Measurements", Appl. Phys. Lett., Vol 31, No 14, 15th Dec, 1977, pp 848-850.
15. H. Wallinga, "A Method for the Measurement of the Turn-On Condition in MOS Transistors", Solid State Electronics, Vol 14, No 11, pp 1093-1098.
16. A. Vladimirescu, S. Liu, "The Simulation of MOS Integrated Circuits Using SPICE2", Memorandum no. UCB/ERL M80/7, Berkley, Feb 1980.
17. K. Doganis, D.L. Scharfetter, "General Optimization and Extraction of IC Device Model Parameters", IEEE Trans. Electron Devices, Vol ED-30, no. 9, Sept 1983, pp 1419-1228.
18. W. Maes, K. De Meyer, L. Dupas, "SIMPARG: A Parameter Extraction Program to be used for any User-Defined Analytical Expression in the Field of Process and Device Modelling", 15th European Solid State Research Conference, ESSDERC 85, Aachen, Sept 1985, pp 153-154.
19. D.S. Perloff, F.E. Wahl, J.D. Reimer, "Contour Maps Reveal Non-Uniformity in Semiconductor Processing State State Technology, 1977, pp 30-42.
20. P.S. Burgraaf, "Instruments with Wafer Mapping Capability", Semiconductor International, March 1984, pp 52-57.
21. C. Chrones, "Parametric Test Systems for Wafer Processing", Semiconductor International, Oct 1980, pp 113-140.
22. G.C. Evans, "Semiconductor Parametric Testing Yesterday, Today and Tomorrow", Semiconductor Production, April/May 1982, pp 8-17.
23. U. Kaemph, "Automated Parametric Testers to Monitor the Integrated Circuit Process", Solid State Technology, Sept 1981, pp 81-87
24. M.G. Buehler, L.W. Lindholm, "Role of Test Chips in Coordinating Logic and Circuit Design and Layout Aids for VLSI", Solid State Technology, Sept 1981, pp 68-74.
25. B.M.M. Henderson, A.M. Gundlach, A.J. Walton, "Integrated Circuit Test Structure which Uses a Vernier to Electrically Measure Mask Misalignment", Electronics Letters, Vol 19, no, 21, 13th Oct 1983, pp 868-869.
26. A.J. Walton, J.M. Robertson, R. Holwill, B. Moore, "On Chip Switching for dc Parametric Testing", Electronics Lett., no 10, Vol 21, May 1985, pp 422-423.

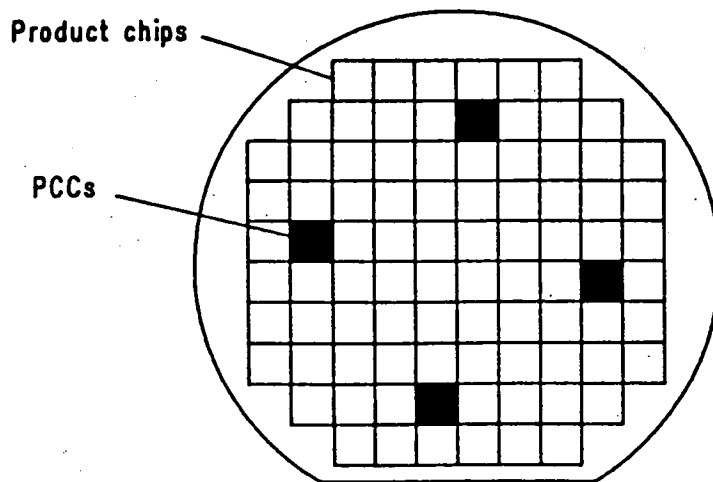


Figure 1. A Wafer showing the position of PCC drop-ins.

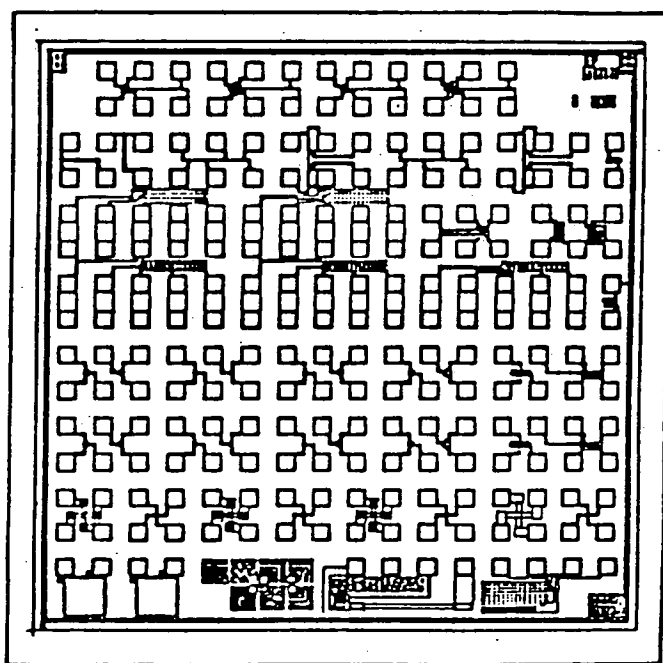


Figure 2. A PCC which uses the 2xN probe arrangement.

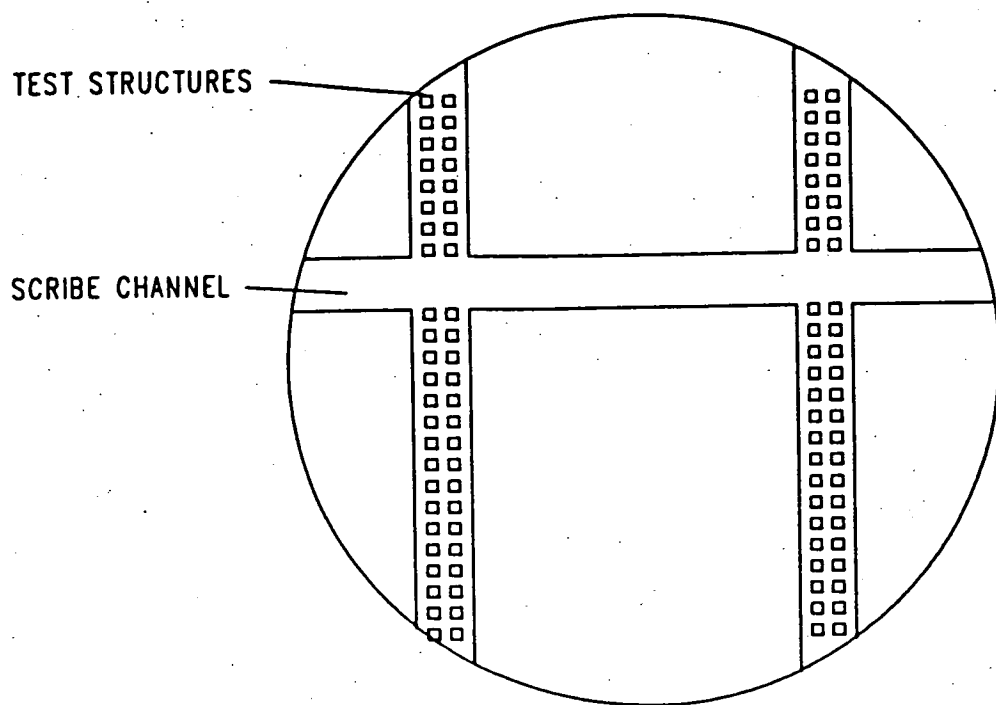


Figure 3. Test structures located in the scribe channel.

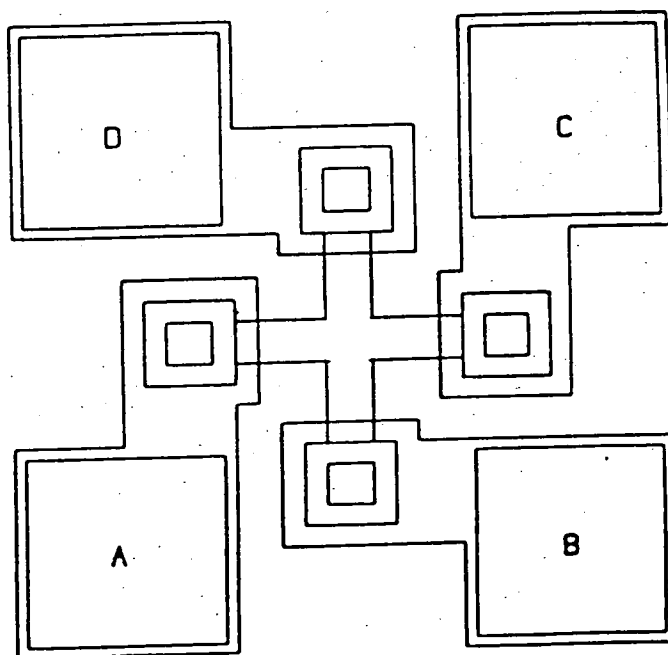


Figure 4. The Greek cross.

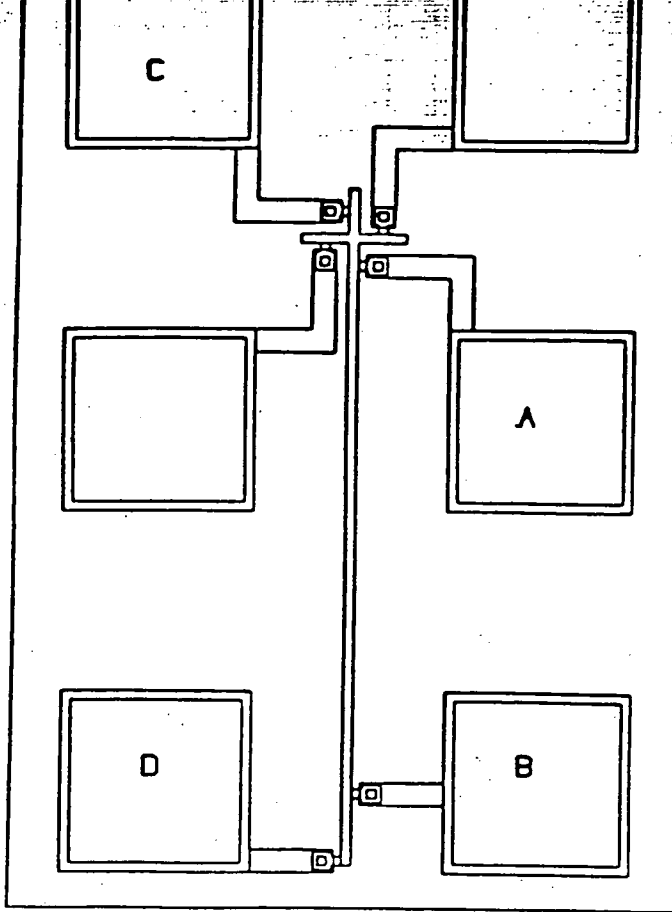


Figure 5. A structure for measuring the linewidth of conductors.

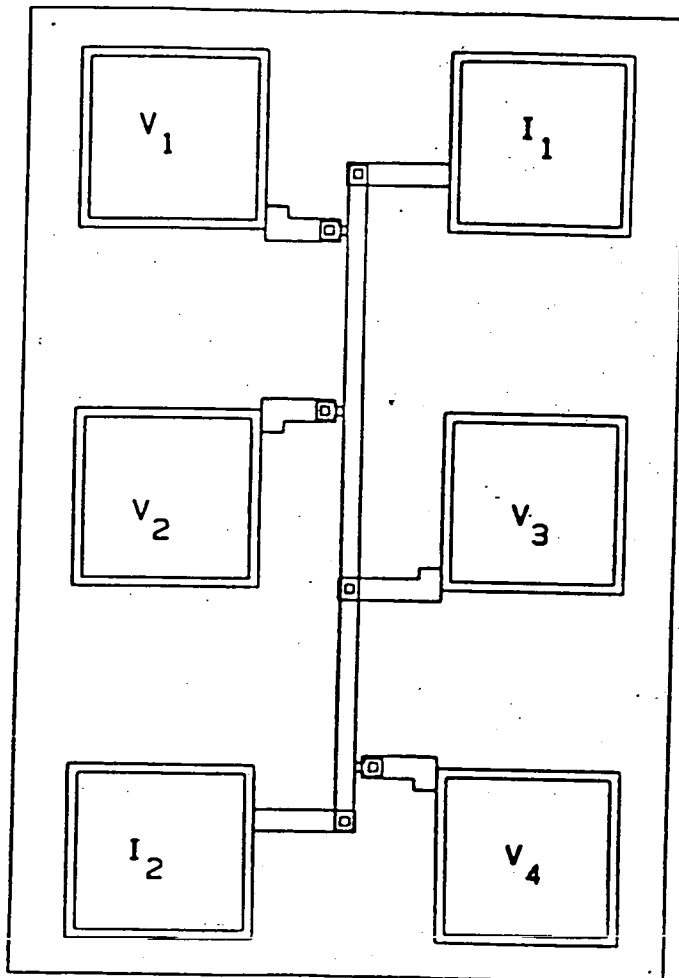


Figure 6. A structure for measuring the superposition error between a conductor and contacts.

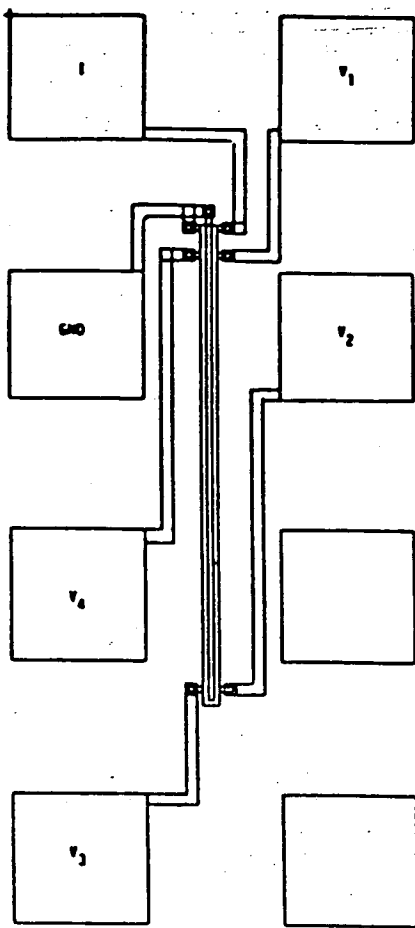


Figure 7. A structure for measuring the misalignment error between polysilicon and diffusion.

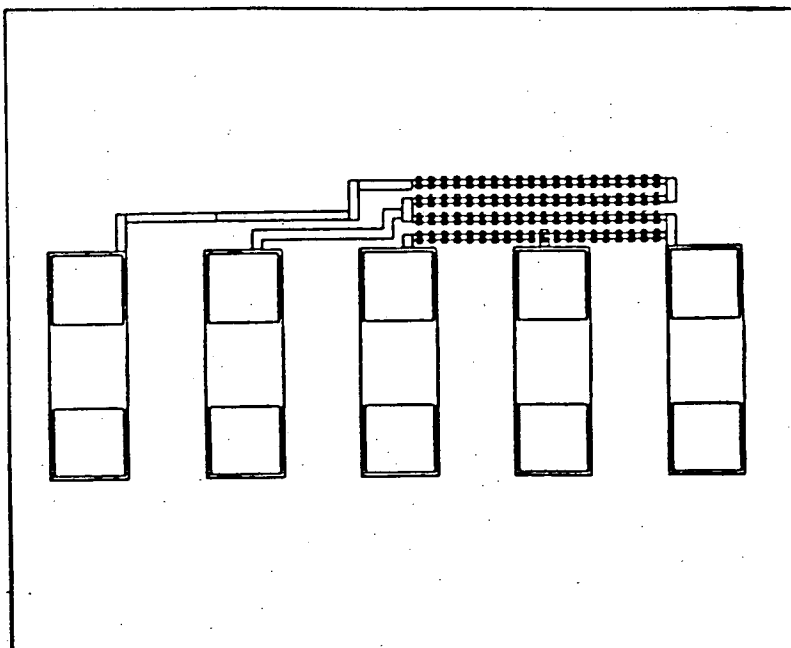


Figure 8. A structure for evaluating the effectiveness of aluminium step coverage.

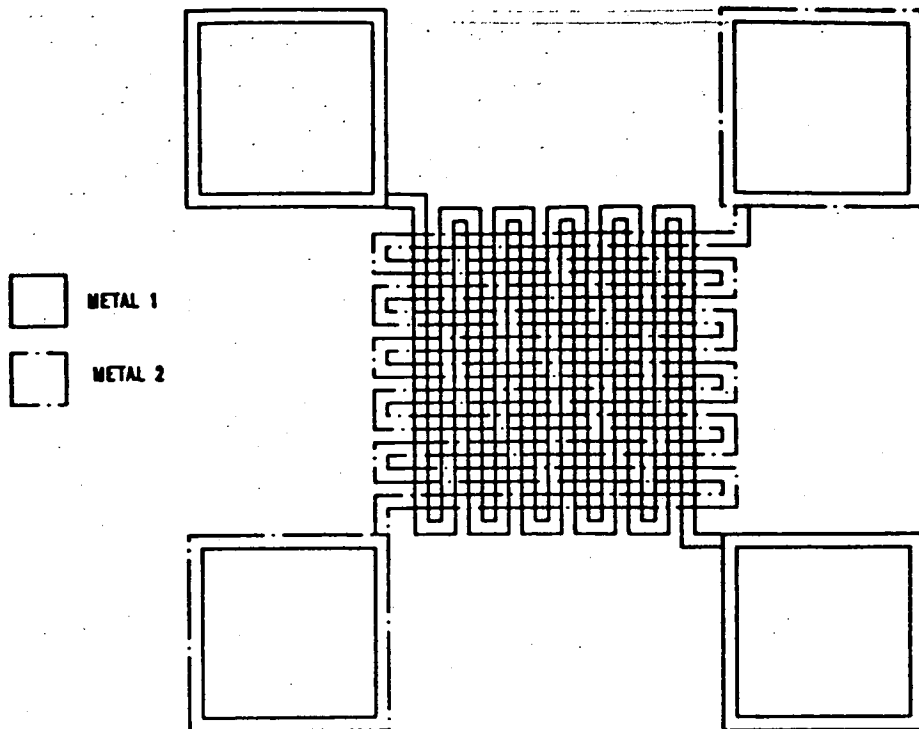


Figure 9. A structure for examining interconnect problems for two level metal systems.

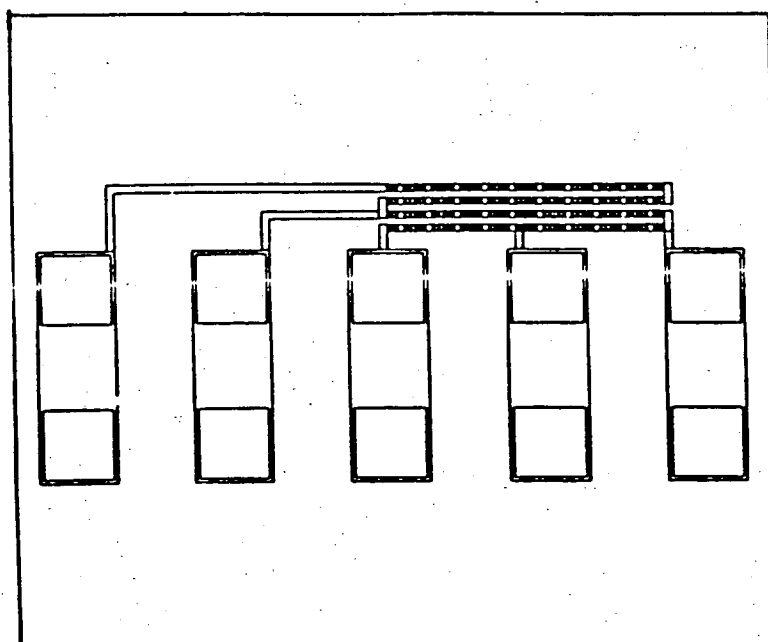


Figure 10. A metal - polysilicon contact chain.

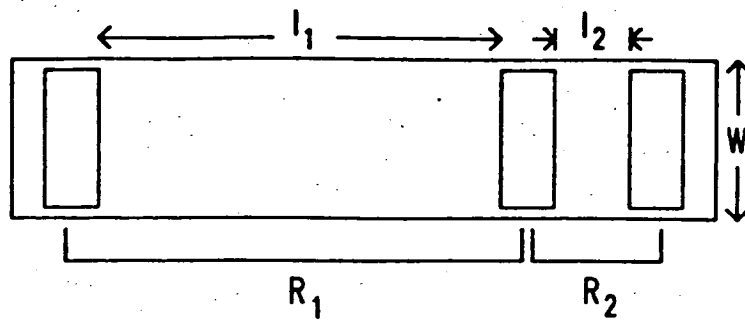


Figure 11. Structure for measuring contact resistance.

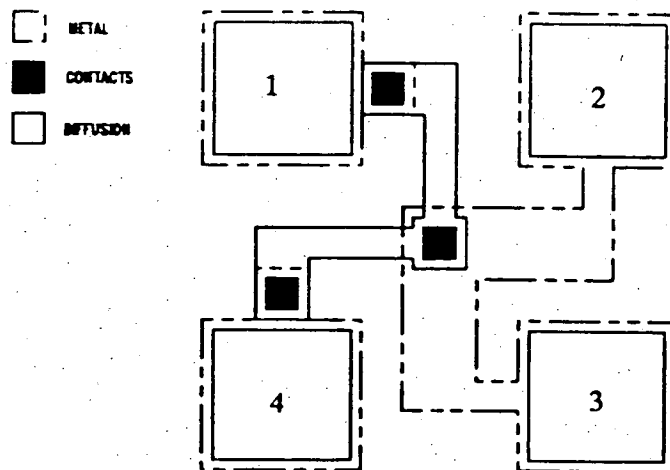


Figure 12. A Kelvin structure for measuring contact resistance.

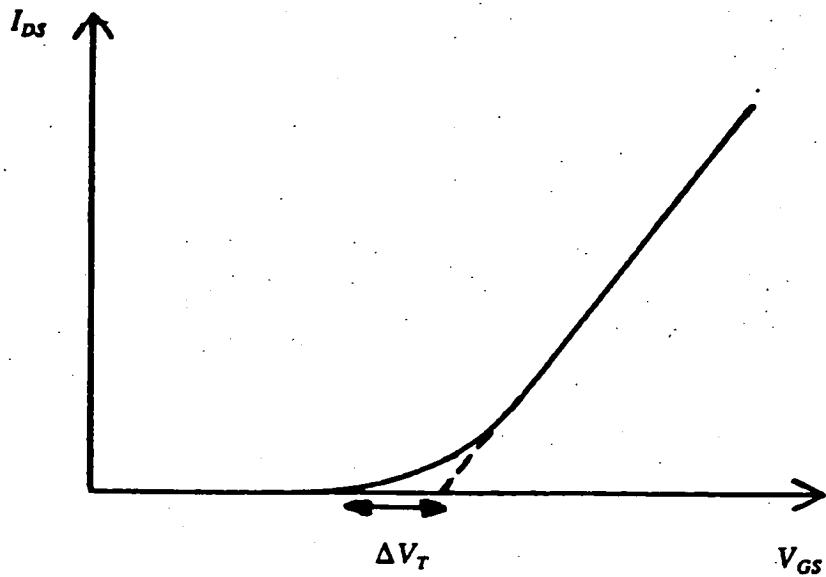


Figure 13. Illustration of the gradual turn on of MOS transistors which makes the definition of V_T difficult.

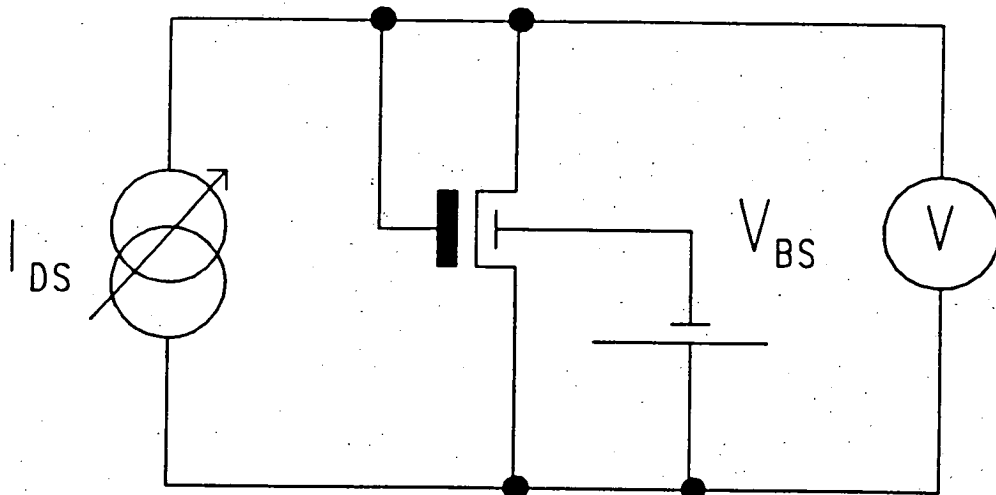


Figure 14. Circuit which can be used for measuring V_T by forcing current and measuring V_{GS} . Note that the voltmeter must have a high input impedance compared with the channel of the device under test.

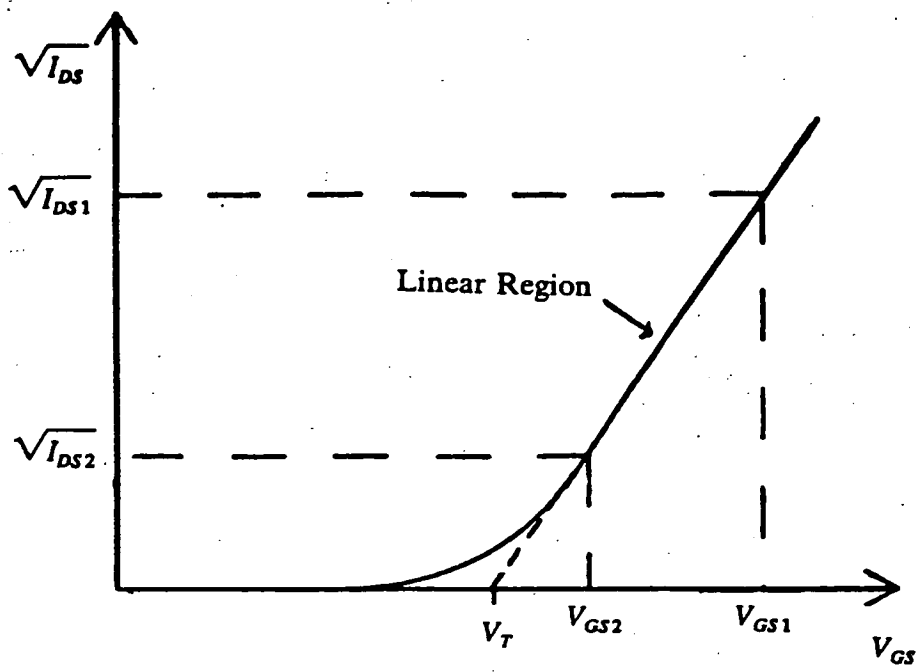


Figure 15. The $\sqrt{V_{DS}}$ vs V_{GS} method of measuring V_T .

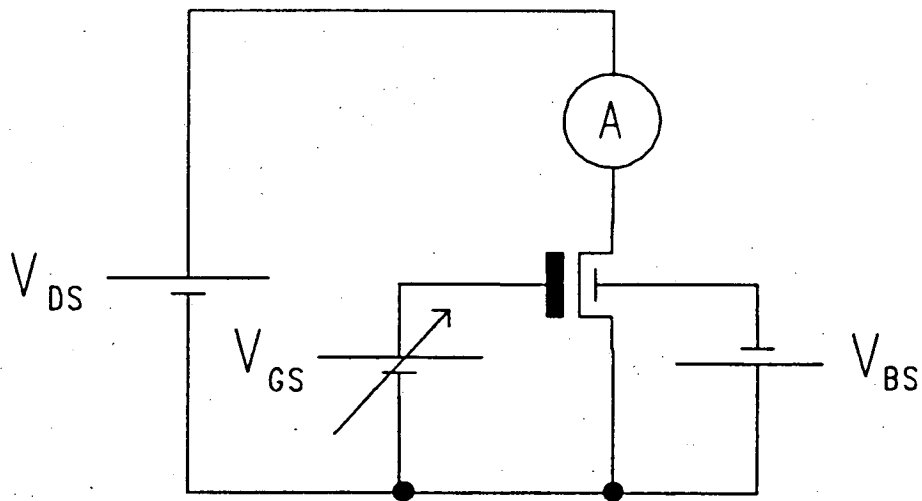


Figure 16. An alternative circuit for measuring V_T .

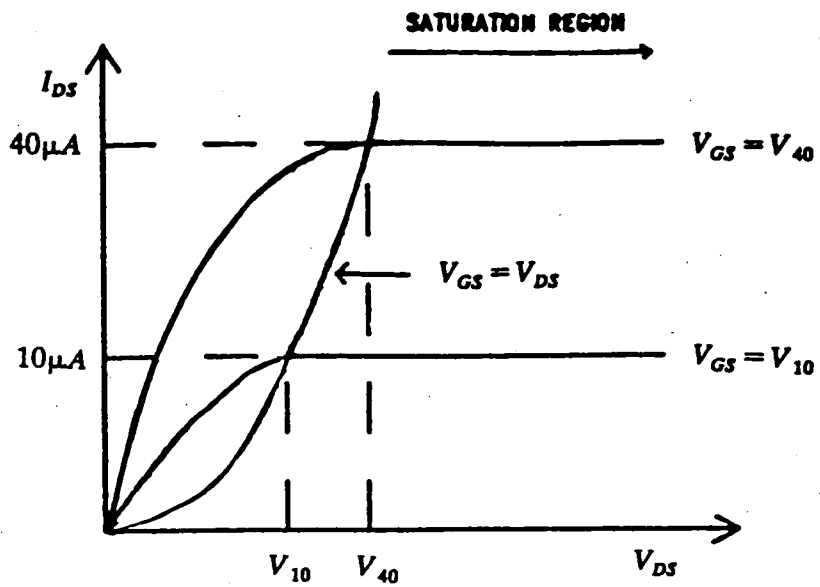


Figure 17. Illustration of V_{10} and V_{40} . It can be observed that if the transistor characteristic is flat in the saturated region the the 10-40 method will give the same result for $V_{DS} \geq V_{GS}$.

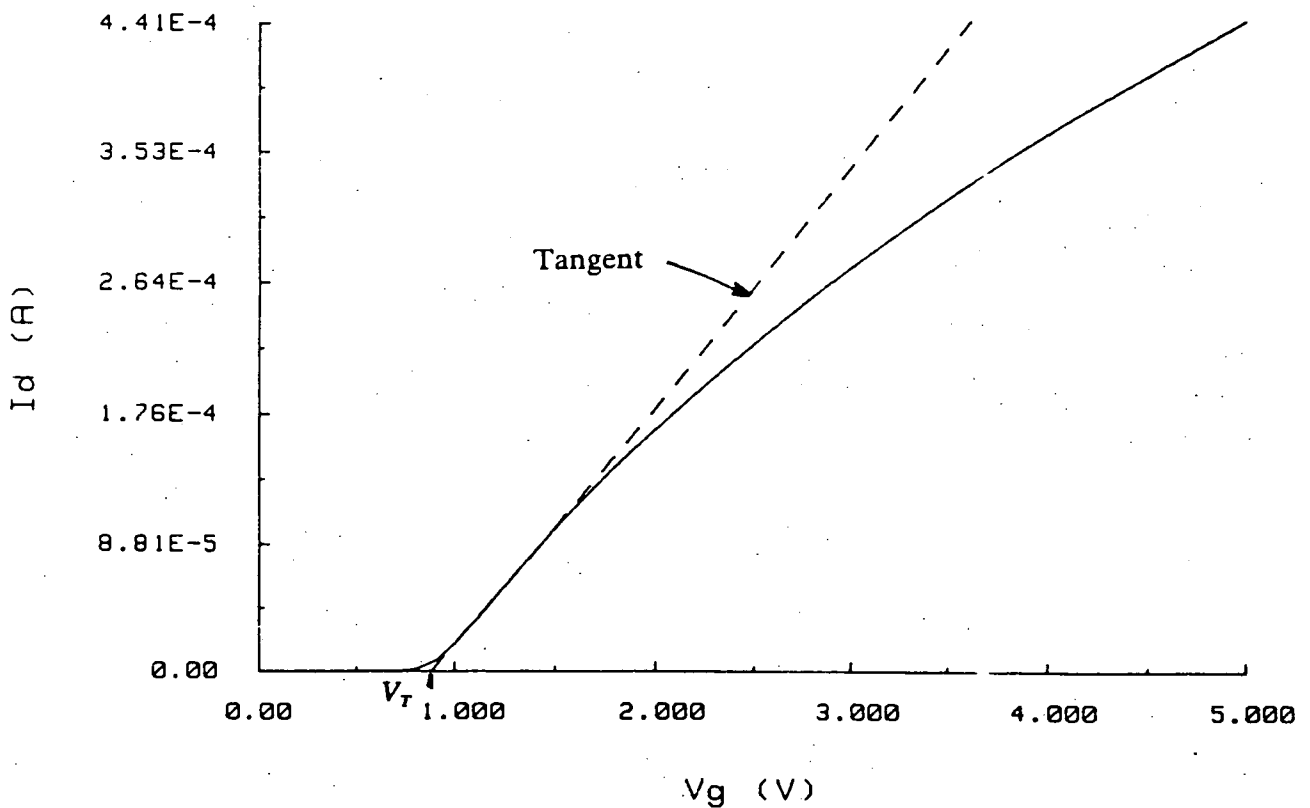


Figure 18. The measurement of V_T for SPICE parameter extraction.

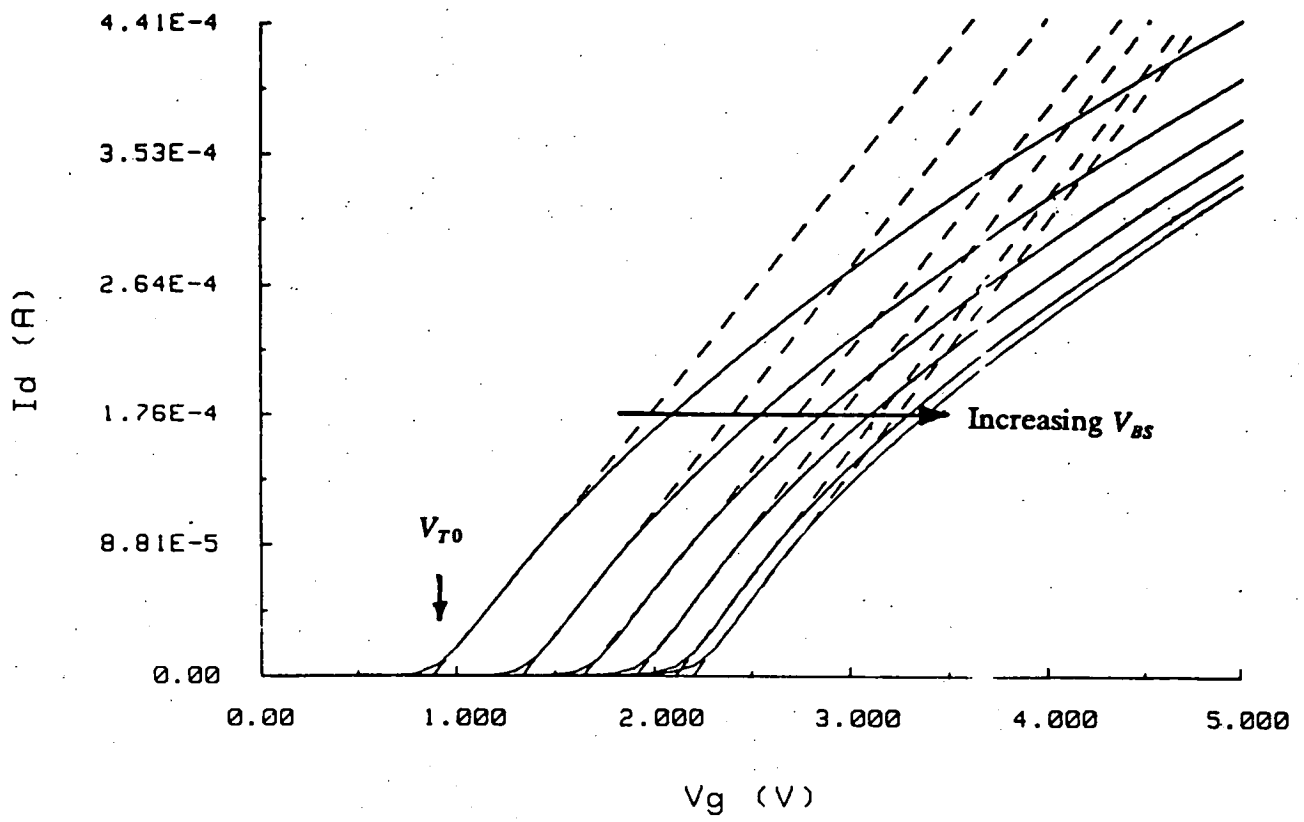


Figure 19. Gate Voltage characteristics for different values of V_{BS} .

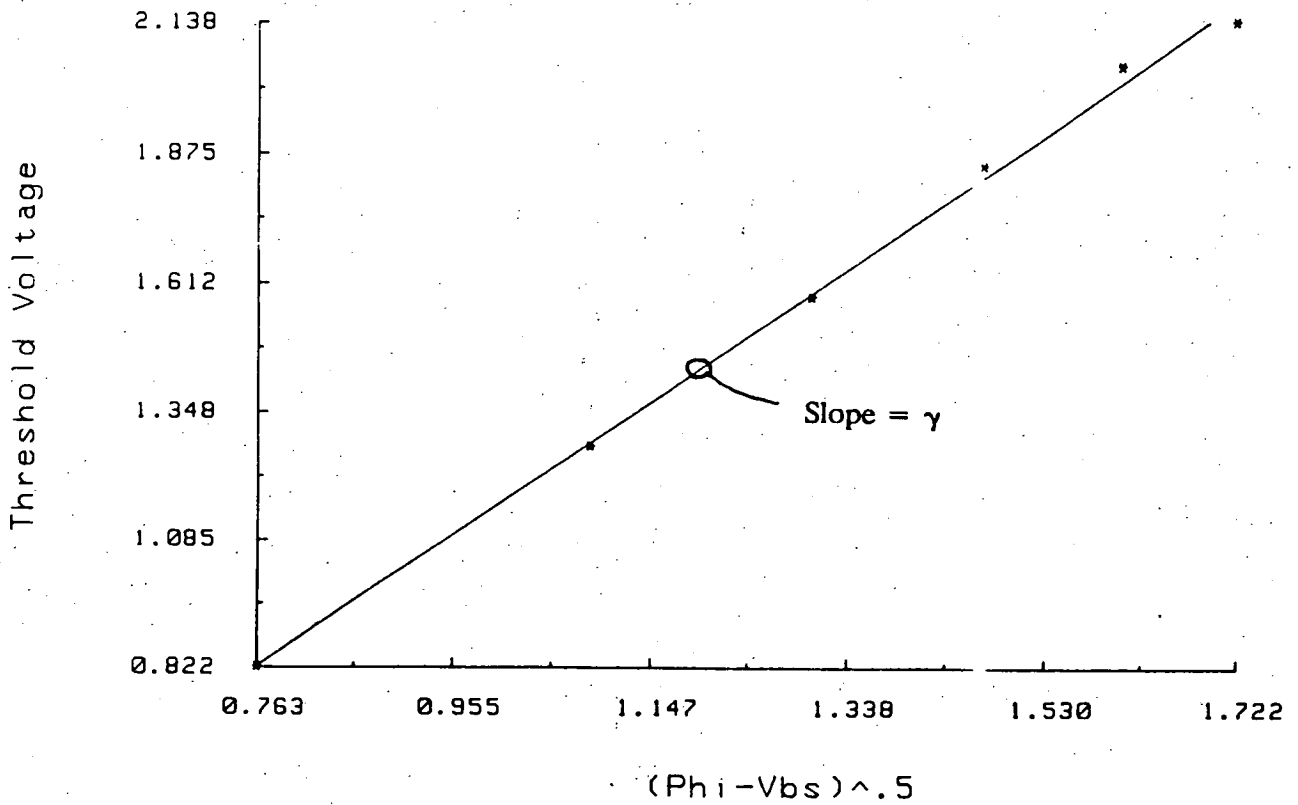


Figure 20. Threshold voltage against $\sqrt{2\phi_B - V_{BS}}$ from which γ can be derived from the slope.

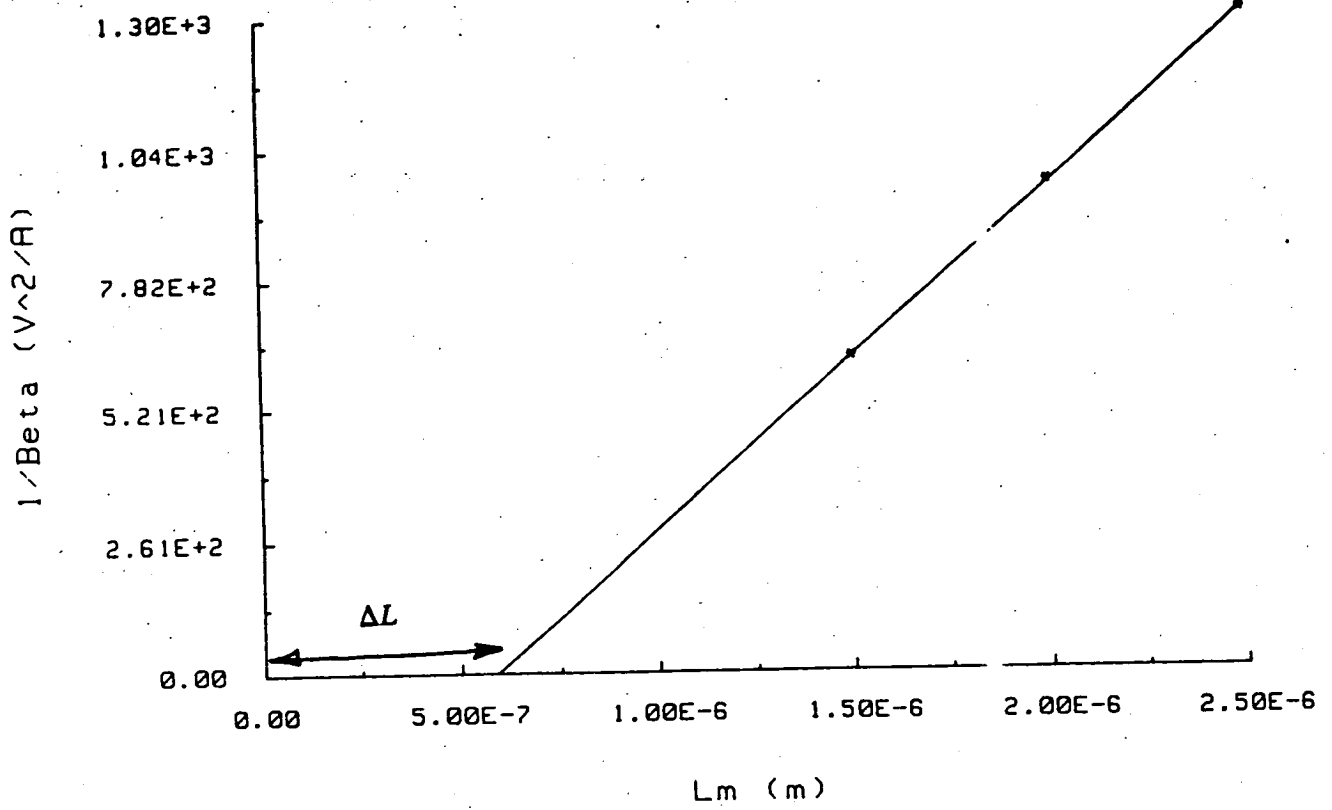


Figure 21. Derivation of the effective length.

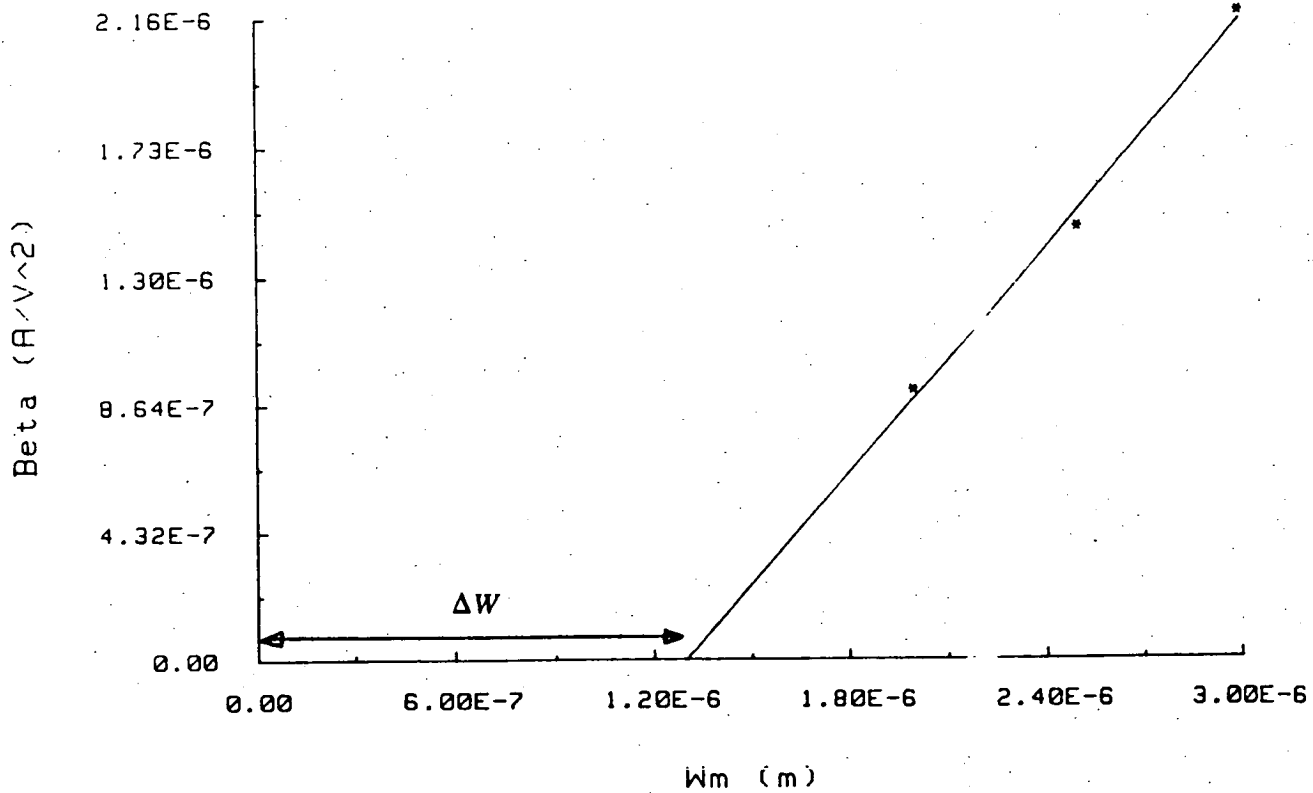


Figure 22. Derivation of the effective width.

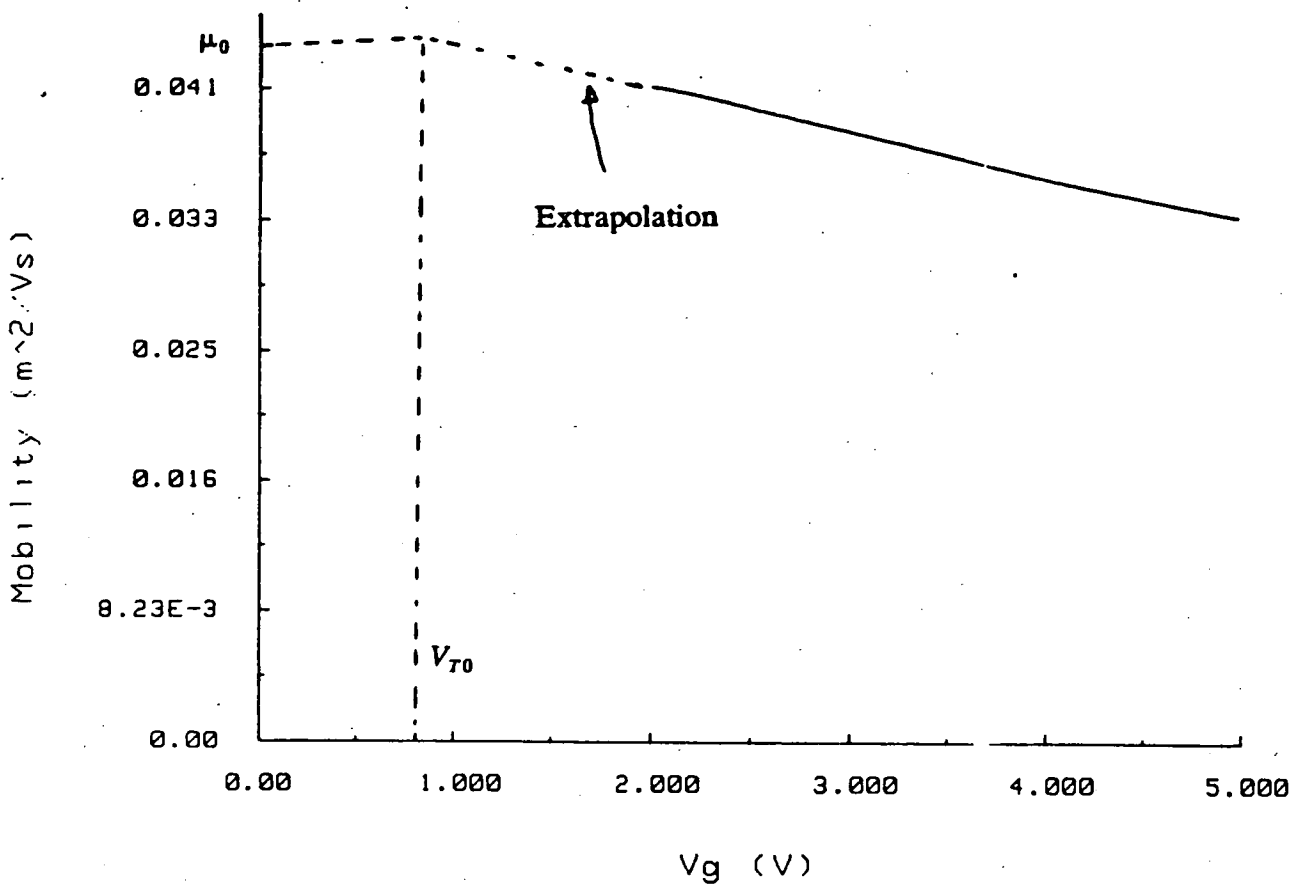


Figure 23. Plot of μ_{eff} against V_{GS} from which μ_o can be derived by extrapolation.

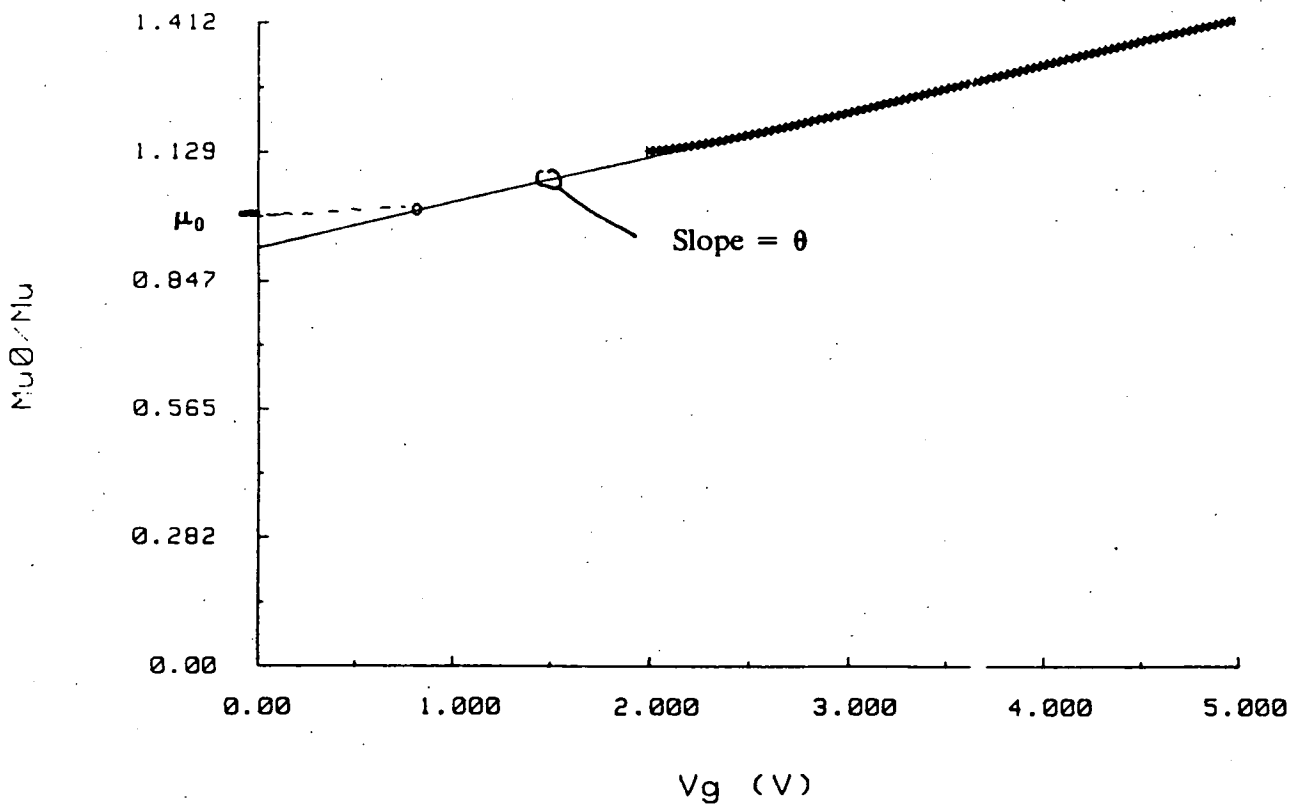
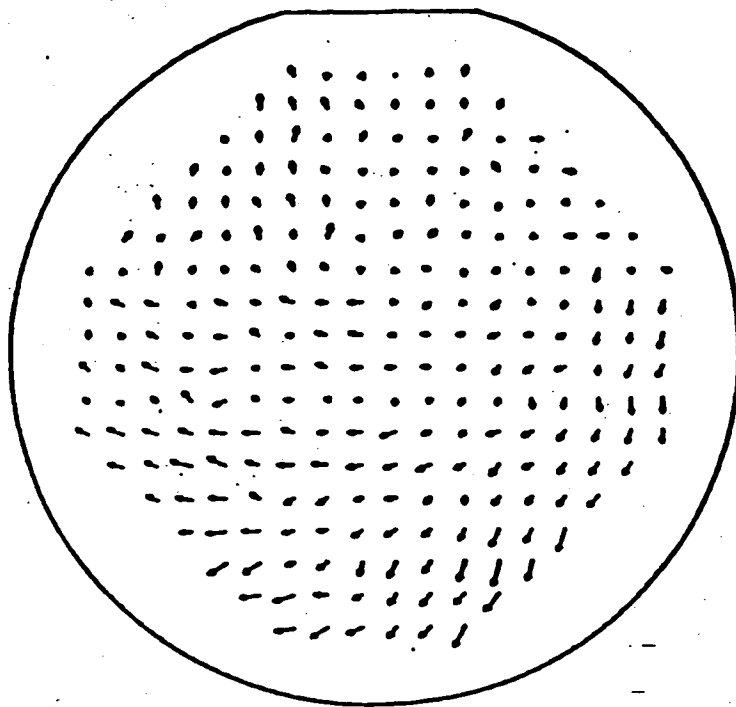


Figure 24. Plot of μ_o/μ_{eff} against V_{GS} from which θ is given by the slope.



<u>Statistical Parameters:</u>	
Tx	-0.145 um
Ty	-0.031 um
Ex	0.027 um/cm
Ey	0.071 um/cm
Sx	-6.447 urad
Sy	-7.167 urad
<u>Correlation Coeffs.:</u>	
x-axis	0.762
y-axis	0.882
<u>Resid. Root Mean Sq.:</u>	
x-axis	0.115 um
y-axis	0.104 um
Vector	0.155 um

Figure 25. Wafer map of misalignments.

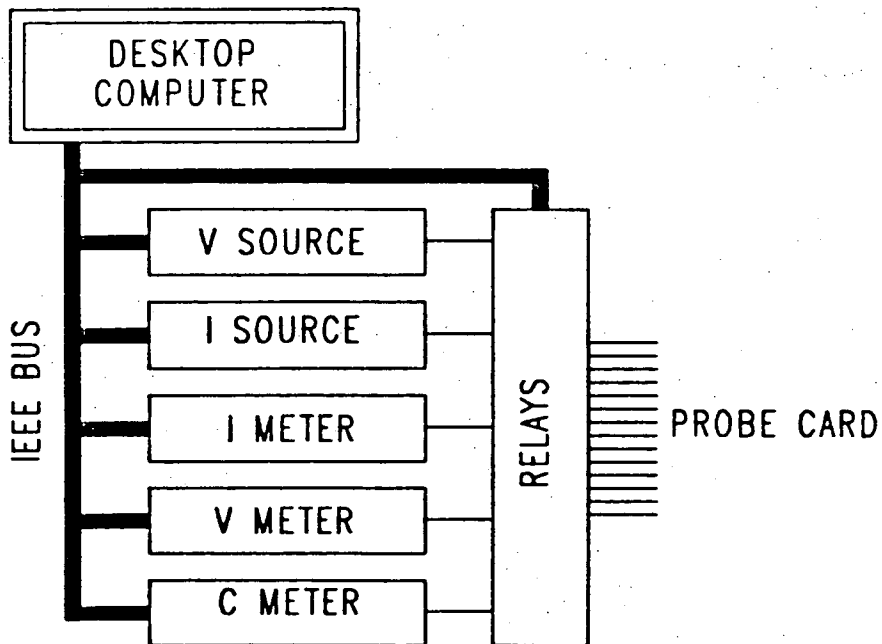


Figure 26. A single user parametric test system.

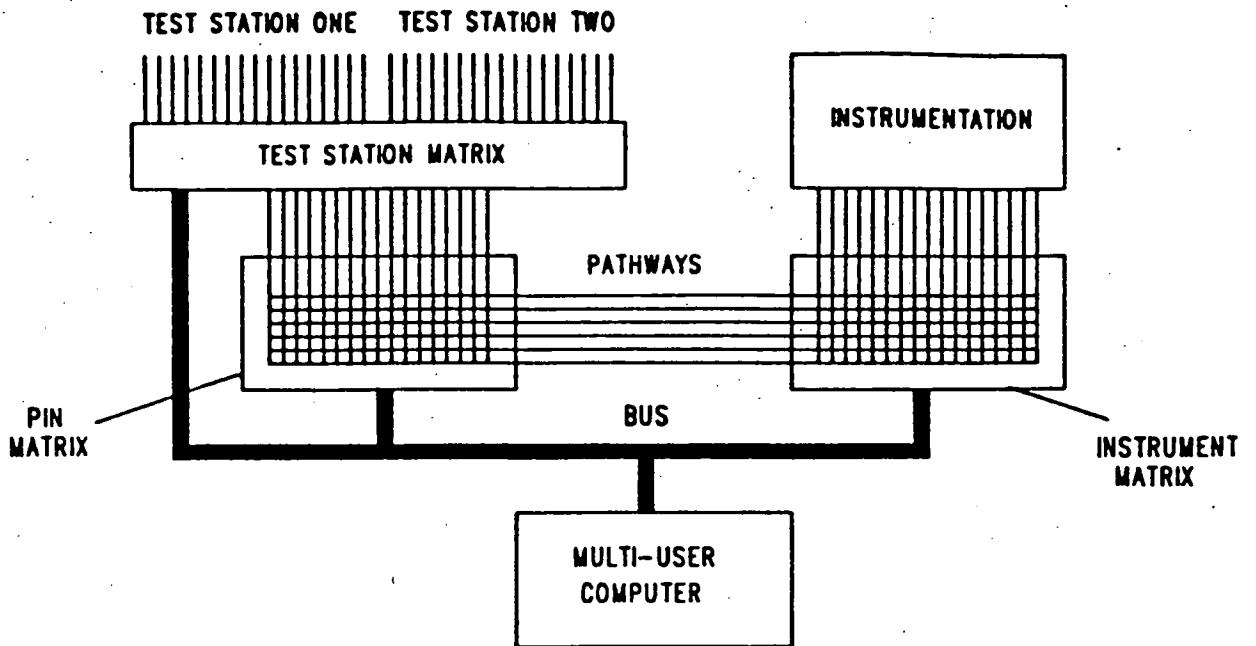


Figure 27. A multi-user parametric test system.

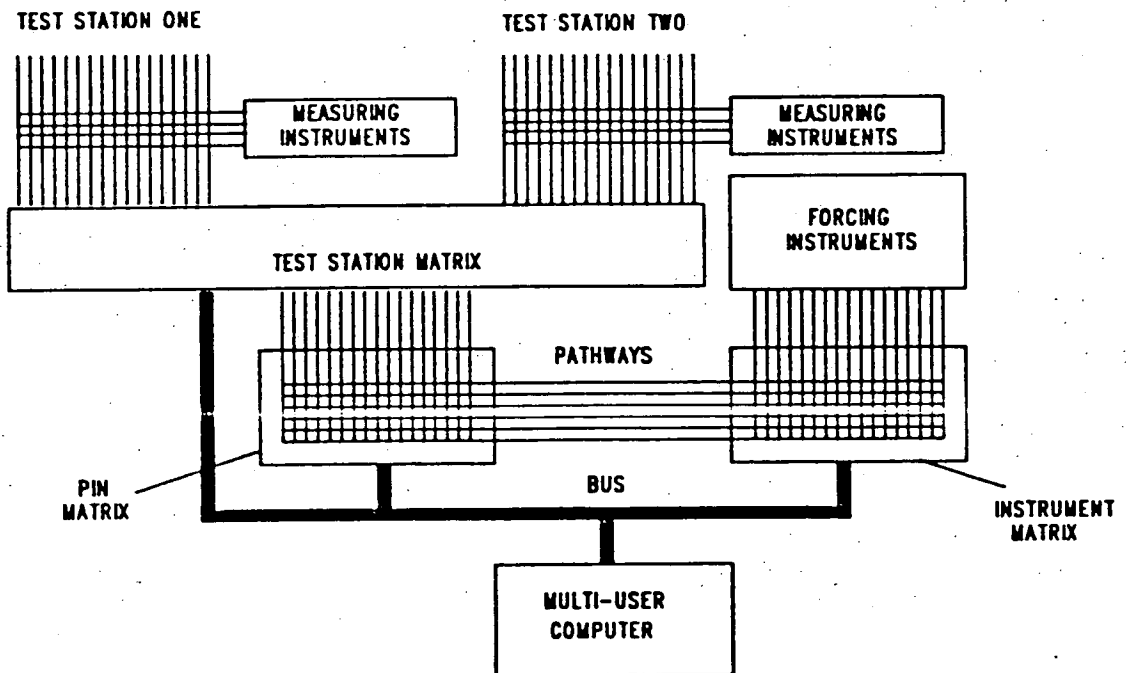


Figure 28. A multi-user parametric test system with instrumentation in close proximity to the probes.

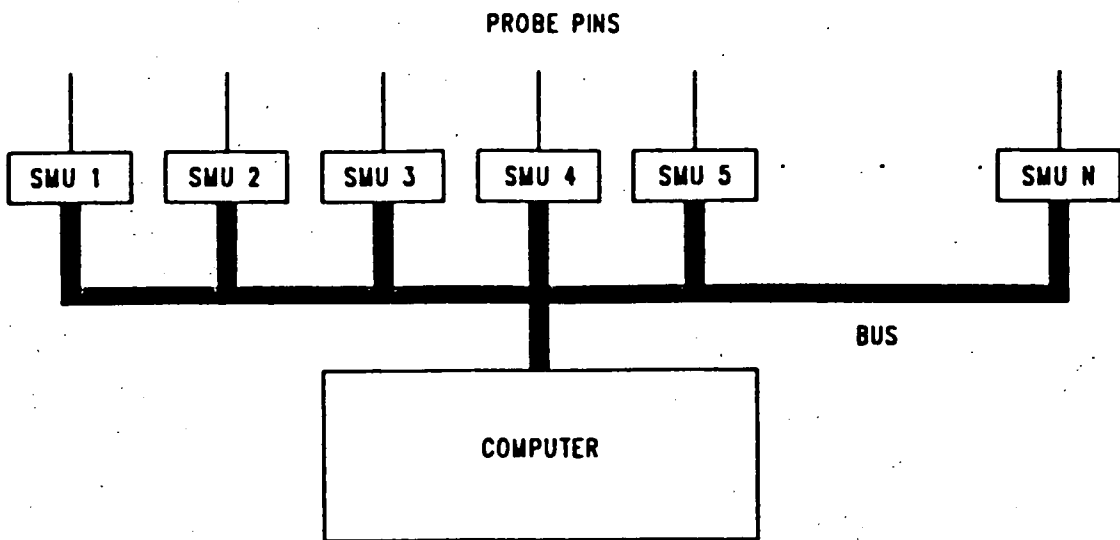


Figure 29. A parametric test system with each pin having its own source-measure unit.