Durham E-Theses

Design of a reliability methodology: Modelling the influence of temperature on gate Oxide reliability

Owens, Gethin Lloyd

How to cite:

Owens, Gethin Lloyd (2007) *Design of a reliability methodology: Modelling the influence of temperature on gate Oxide reliability*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/2695/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, Durham University, University Office, Old Elvet, Durham DH1 3HP
e-mail: e-theses.admin@dur.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

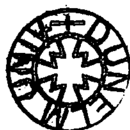
Design of a Reliability Methodology : Modelling the Influence of Temperature on Gate Oxide Reliability

Gethin Lloyd Owens

The copyright of this thesis rests with the author or the university to which it was submitted. No quotation from it, or information derived from it may be published without the prior written consent of the author or university, and any information derived from it should be acknowledged.

A Thesis presented for the degree of
Doctor of Philosophy

School of Engineering
University of Durham
Durham
UK



- 5 FEB 2007

Design of a Reliability Methodology : Modelling the Influence of Temperature on Gate Oxide Reliability

Gethin Lloyd Owens

Abstract

In any product design process there is a fundamental endeavour to design reliable products. This is to ensure that the product will have the capability to perform as designed for a required period of time. However in a world that demands faster and greater functionality of products, ensuring long-term reliability is ever becoming an increased concern. Advances in technology have provided huge performance benefits, but have also accelerated the onset of reliability problems. Compounding the problem is a growing demand for devices applicable to a wide range of operating environments such as elevated temperatures.

Traditionally reliability assurance has relied on failure analysis at the latter stages of the product development process. This involves a time consuming and costly cycle of re-design and testing that is simply no longer practical for today's time-to-market requirements. What is required is a quick and effective way of evaluating product reliability at the earliest stages of conceptual design and to allow the designer to repeatably examine and improve the product. In this thesis we address this problem with the design of a new reliability tool. This tool uses a novel methodology to integrate for the first time several failure mechanism models, thus providing a method of reliability analysis that can be used throughout the design process.

An Integrated Reliability Methodology (IRM) is presented that encompasses the changes that technology growth has brought with it and includes several new device degradation models. Each model is based on a physics of failure approach and includes on the effects of temperature. At all stages the models are verified experimentally on modern deep sub-micron devices. The research provides the foundations of a tool which gives the user the opportunity to make appropriate trade-offs between performance and reliability, and that can be implemented in the early stages of product development.

Declaration

The work in this thesis is based on research carried out in the School of Engineering, University of Durham, UK. No part of this thesis has been submitted elsewhere for any other degree or qualification and it all my own work unless referenced to the contrary in the text.

Copyright © 2006 by Gethin Lloyd Owens.
"The copyright of this thesis rests with the author. No quotations from it should be published without the author's prior written consent and information derived from it should be acknowledged".

Acknowledgements

Firstly I would like to thank my supervisor, Dr Simon Johnson, for his academic supervision and encouragement throughout my Ph.D. I remain grateful to Simon for giving me the opportunity to undertake a Ph.D. Discussions, guidance and feedback from Simon have provided invaluable for the development of the thesis and it has been a pleasure knowing and working with him. Additionally I would like to thank those at Goodrich, Bob Newman, Peter Shrimpling and Robert Edward's for their support and helpful suggestions throughout.

Thanks go to Andrew Maiden, a friend, colleague and housemate. Thanks not only for all his help and knowledge but also for sharing the experience. The long walks and life at Aden cottage (especially during the cold winter nights) will always be remembered. For her continued belief in me and the joy she has brought to me, I would like to thank my girlfriend, Joelle Davies. Your smile, and food parcels kept me going in those long days.

Finally a special thanks go to my family for their patience and for standing by me during the more difficult times. I would particularly like to thank my Dad for his love, and always being there for me, and my Mother who has given me the confidence to follow my ambitions in life. My love and gratitude is forever with you.

Contents

Abstract ii

Acknowledgements v

List of Figures xii

List of Tables xvii

Glossary 1

1 Introduction 3

1.1 Objectives and Focus of the Research 5

1.2 Organisation of Thesis 6

2 Reliability in CMOS Technology : A Review 7

2.1 Introduction 7

2.2 CMOS Reliability Review 8

2.3 Physics of Failure 16

2.4 Failure Methodologies 19

2.5 Failure Mechanisms 21

2.5.1 Electromigration 21

2.5.2 Gate Oxide Failures 24

2.5.2.1 Gate Induced Drain Leakage 25

2.5.2.2 Negative Bias Temperature Instability 27

2.5.3 Hot-Carrier Effects 31

2.6 High Temperature Effects 33

2.7 Alternative Materials 34

2.8 Silicon-On-Insulator Technology 36

2.9 Conclusion 38

3 High Temperature Device Model **40**

3.1 Introduction 40

3.2 Devices and Measurement 41

3.3 I-V Model 43

3.3.1 Threshold Voltage 43

3.3.2 Linear Region 45

3.3.3 Saturation Region 49

3.3.4 Mobility Model 50

3.3.5 Parasitic Drain/Source Resistance . . . 52
3.3.6 Subthreshold Region . . . 52
3.3.7 Transition Region . . . 55
3.3.8 I-V Model Test Results . . . 56
3.4 Temperature Effects . . . 63
3.4.1 Carrier Mobility . . . 63
3.4.2 Velocity Saturation . . . 64
3.4.3 Threshold Voltage . . . 66
3.4.4 Subthreshold Swing Factor . . . 68
3.4.5 Drain Source Resistance . . . 69
3.4.6 Temperature Results . . . 70
3.5 Conclusion . . . 72
4 NBTI **74**
4.1 Introduction . . . 74
4.2 Mechanism Overview . . . 75
4.3 Experimental Method . . . 78
4.4 NBTI Model . . . 79
4.5 Recovery Effect . . . 82
4.6 Lock-in . . . 87

4.7 Dynamic Degradation 90

4.8 Conclusion 93

5 GIDL **94**

5.1 Introduction 94

5.2 Mechanism Overview 95

5.3 Experimental Method 97

5.4 GIDL Model 98

5.5 Integration into I-V Model 109

5.6 Conclusion 113

6 Hot-Carrier Effects **114**

6.1 Introduction 114

6.2 Mechanism Overview 115

6.3 Experimental Method 118

6.4 Hot-Carrier Model 119

6.5 Influence of Temperature 124

6.6 Lifetime Prediction 130

6.7 Integration into the I-V Model 132

6.8 Conclusion 134

7 **An Integrated Reliability Methodology** **135**

7.1 Introduction 135

7.2 Methodology 136

7.3 Tool Design 138

7.4 Simulation Results 141

7.4.1 Scaling 141

7.4.2 Failure Mechanisms 146

7.4.2.1 NBTI 147

7.4.2.2 GIDL 151

7.4.2.3 Hot-Carrier Effects 153

7.5 Conclusion 157

8 **Future Work and Conclusions** **160**

8.1 Future Work 160

8.1.1 Layout Diagnostics 161

8.1.2 Competing Failures 162

8.1.3 Tool Implementation 163

8.2 Conclusions 164

Bibliography **167**

A Experimental Setup 181

B Parameter Optimization 183

List of Figures

2.1 Bathtub curve 9

2.2 Product development cycle 10

2.3 Reliability distribution 11

2.4 Physics of failure process 17

3.1 An image of 0.18μm n-channel device 41

3.2 Schematic diagram of experimental setup 42

3.3 Threshold voltage as a function of drain voltage 44

3.4 MOSFET gate construction 46

3.5 Mobility as a function of gate voltage 51

3.6 Subthreshold current as a function of gate voltage 54

3.7 Illustration of the transition region 56

3.8 V_{DS} - I_{DS} characteristic 58

3.9 Saturation voltage as a function of gate voltage 59

3.10 V_{GS} - I_{DS} charakteristic 60

3.11 V_{GS} - I_{DS} characteristic as a function of substrate bias 61

3.12 Logarithmic V_{GS} - I_{DS} characteristic 62

3.13 Carrier mobility as function of temperature 64

3.14 Velocity saturation as function of temperature 65

3.15 Fermi potential as a function of temperature 66

3.16 Threshold voltage as a function of temperature 67

3.17 Temperature dependence of the subthreshold region (p-channel) . . . 68

3.18 Temperature dependence of parasitic resistance 69

3.19 V_{GS} - I_{DS} characteristic at various temperatures 70

3.20 V_{DS} - I_{DS} characteristic at various temperatures 71

4.1 Stress time window 75

4.2 NBTI n-channel and p-channel degradation comparison 76

4.3 Interface trap creation 77

4.4 Threshold voltage shift under DC stress 80

4.5 NBTI stress as a function of temperature 81

4.6 Arrhenius plot showing temperature dependence of $\triangle V_{th}$ 82

4.7 Threshold voltage recovery after NBTI stress 83

4.8 Threshold voltage recovery as a function of gate voltage 84

4.9 Comparison between NBTI model and experimental results 85

4.10 Comparison between stress and recovery activation energies 86

4.11 Temperature dependence of NBTI stress and recovery 87

4.12 NBTI lock-in illustration 88

4.13 NBTI lock-in temperature dependence 89

4.14 Dynamic stress configuration 90

4.15 NBTI frequency dependence 91

4.16 NBTI lifetime comparison 92

5.1 Illustration of gate-to-drain overlap region 95

5.2 GIDL mechanism (n-channel) 96

5.3 GIDL characteristic (p-channel) 98

5.4 Comparison of first GIDL model with measured data 99

5.5 Illustration of GIDL transition function 101

5.6 Comparison of GIDL model with 1st design 102

5.7 Measured GIDL current at various temperatures 103

5.8 Change in GIDL in different operating regions (p-channel) 104

5.9 Change transition points with temperature (p-channel) 106

5.10 Model predictions at various temperatures 107

5.11 Illustration of device operating regions (n-channel) 109

5.12 Subthreshold region at various drain voltages 111

5.13 Complete I-V characteristic 112

6.1 Illustration of hot-carrier mechanism (n-channel) 115

6.2 Illustration of silicon-oxide interface energy bands 116

6.3 Transconductance degradation after hot-carrier stress 117

6.4 Hot-carrier stress configuration 118

6.5 Inaccuracy in substrate current prediction (n-channel) 120

6.6 Substrate current versus gate voltage (n-channel) 121

6.7 Substrate current versus gate voltage (p-channel) 122

6.8 Comparison between n- and p-channel substrate currents 123

6.9 Substrate current at various temperatures (high V_{DS}) 124

6.10 Substrate current at various temperatures (low V_{DS}) 125

6.11 Substrate current crossover point 126

6.12 Model predictions for substrate current transition 128

6.13 Substrate current characteristic above and below transition 129

6.14 Hot-carrier lifetime (n-channel) 131

6.15 Complete I-V characteristic 133

7.1 Integrated reliability methodology 137

7.2 Layers of MathCAD development 139

7.3 Embedded MathCAD file 140

7.4 Leakage current as a function of channel length (n-channel) 142

7.5 Saturation current as a function of channel length 143

7.6 Influence of temperature on scaling channel length (n-channel) 144

7.7 Delay time as a function of channel length 145

7.8 Snapshot of NBTI model in simulation tool 147

7.9 NBTI stress simulation 148

7.10 NBTI stress simulation at varied temperatures 149

7.11 NBTI stress simulation as a function of oxide thickness 150

7.12 Voltage dependency of GIDL for varied channel lengths 151

7.13 Temperature dependency of GIDL at varied channel lengths 152

7.14 Snapshot of HCE model in simulation tool 153

7.15 Hot-carrier stress simulation 154

7.16 Substrate current as a function of channel length (p-channel) 155

7.17 Maximum electric field as a function of channel length 156

7.18 Device lifetime as a function of channel length (n-channel) 156

8.1 Development of methodology for layout diagnostics 161

A.1 Schematic diagram of test rig 182

List of Tables

2.1 MOSFET road map 25

2.2 Comparison table of different semiconductor materials 34

3.1 Extracted model parameters 57

5.1 Extracted GIDL model parameters 108

7.1 Scaling parameters 146

Glossary

AGREE	Advisory Group on the Reliability of Electronic Equipment
AlN	Aluminium Nitride
BSIM	Berkeley Short-channel IGFET Model
CAD	Computer Aided Design
CALCE	Computer Aided Life Cycle Engineering Centre
CLM	Channel Length Modulation
CMOS	Complimentary Metal-Oxide Semiconductor
COTS	Commercial Off The Shelf
DIBL	Drain Induced Barrier Lowering
DRAM	Dynamic Random Access Memory
EEPROM	Erasable Programmable Read Only Memory
GaAs	Gallium Arsenide
GaN	Gallium Nitride
GIDL	Gate Induced Leakage Currents
GPIB	General Purpose Interface Bus
HCE	Hot-Carrier Effects
IC	Integrated Circuit
IRM	Integrated Reliability Methodology
I-V	Current-Voltage

JEDEC	Joint Electron Device Engineering Council
LDD	Lightly Doped Drain
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MTTF	Mean Time To Failure
NBTI	Negative Bias Temperature Instability
PBTI	Positive Bias Temperature Instability
PoF	Physics of Failure
R-D	Reaction-Diffusion
SiC	Silicon Carbide
SOI	Silicon-on-insulator
SOS	Silicon on Sapphire
TSMC™	Taiwan Semiconductor Manufacturing Company
ULSI	Ultra Large Scale Integration
ZTC	Zero-Temperature Coefficient

Chapter 1

Introduction

THE DESIGN OF accurate reliability methodologies must be seen as a vital element for the future evolution of Complimentary Metal-Oxide Semiconductor (CMOS) technologies. The ever continued scaling of CMOS technologies, has made it increasingly difficult to ensure device reliability. Added to this the influence of an external factor such as temperature, the process of predicting the lifetime of a product is made even more complex. Hence, development of a novel reliability methodology is needed to encompass external conditions such as temperature but to also prevent device modelling falling behind the fast changes in technology.

As we enter the age of the nanometer scale, it is essential any reliability or performance issues brought about by reduced dimensions are addressed. Deep sub-micron dimensions have introduced a set of new reliability challenges which are forcing a dramatic change in the approaches used to assure integrated circuit reliability in the product design flow. High speed and low power are not the only targets that designers have to aim for. The designer has to able to simulate how several factors may change the behaviour of known failure mechanisms or even bring about the introduction of a new failure mechanism.

Determining the lifetime of a particular device is dependent on numerous considerations. Device geometry, architecture, materials, and environmental and operational stresses are all factors which play a key role in reliability assessment. The environmental stresses include temperature, relative humidity, pressure, static charge, and their cycles, gradients, and transients. Accelerating operational stresses include voltage and current. While high levels of these stresses will overstress the device to catastrophic failure, smaller values of the same stresses may result in a performance degradation and not necessarily hard breakdown. Hence device modelling is not strictly confined to predicting when a device will fail, but also analysing the effects of any performance degradation over time.

In this thesis we consider how deep sub-micron devices degrade for a varied set of operating and environmental conditions and then develop new models to simulate such behaviour. Using a physics of failure approach, novel models are designed for well known failure mechanisms and also for mechanisms only very recently emerging. This provides the foundation for the development of a new integrated reliability tool. Combining each failure model into a single tool will allow analysis of how any technology changes and/or elevated temperatures affect the performance and lifetime of devices. The author believes the provision of this information is vital to recognise trends in the failure mode distributions for both present day and future technologies. At present we are entering a time when physical limitations may prevent any further scaling. Variations in channel dimensions and oxide thickness to take two examples are undoubtedly going to cause variance to the physical operation of associated failure mechanisms. Hence, the development of a integrated reliability tool would provide invaluable information to how certain failure mechanisms vary after such changes. Furthermore, the tool will provide the capability to predict if the present rate of scaling can continue.

1.1 Objectives and Focus of the Research

The aim of the research is to address the actual impact of temperature and device technology on the failure mechanisms utilising a physics of failure approach. This is to be achieved through the development of an integrated reliability tool and failure models. To accomplish this goal, the following tasks were set.

- To understand how temperature affects the operation of a CMOS device and the development of a new temperature dependent MOSFET model. This will provide current-voltage characteristics and individual parameter data.
- Determine the prominent failure mechanisms that occur in deep sub-micron devices. In particular knowledge about each physical mechanism and its temperature dependence.
- Concentrated research on gate oxide failure mechanisms: how has the reduction in gate oxide thickness affected device performance and reliability.
- Identify any new failure mechanisms such as Negative Bias Temperature Instability (NBTI): Investigate the physical mechanism behind this new failure and develop novel models.
- Studies into how leakage degradation such as Gate Induced Leakage Currents (GIDL) are effected by temperature in modern devices.
- Hot-Carrier Stress: In particular investigate the temperature dependence of hot-carrier damage in deep sub-micron devices.
- Development and design of failure/degradation models using a Physics of Failure (PoF) approach.
- Integrated Reliability Methodology (IRM): how best to integrate the PoF models in a novel way to provide accurate and efficient device analysis.

1.2 Organisation of Thesis

A critical review of reliability in CMOS technology given in chapter 2. This covers the relevant background to the subject from reliability fundamentals such as the bathtub curve through to a review of alternative technologies. In addition the prominent failure mechanisms and the main effects of temperature are discussed in this chapter. Following this, chapter 3 describes the development of a new temperature dependent MOSFET model. The mathematical foundation behind the model is presented along with a set of benchmark simulation and experimental test results. This chapter serves to update present MOSFET current-voltage modelling but also as a foundation for later chapters.

Chapters 4-6 describe the research into the three failure mechanisms covered in this thesis. Chapter 4 investigates a recent degradation phenomenon, Negative Bias Temperature Instability. In chapter 5 a new Gate Induced Drain Leakage model is presented. Finally chapter 6 covers the present understanding of hot-carrier effects including some new results on its temperature dependence. In each of these chapters a new failure model is developed based on experimental data.

Chapter 7 combines the research and model developments in the previous chapters. The current-voltage model and each failure model are integrated into a novel reliability tool and simulation findings presented.

Future work and possible tool developments such as layout diagnostics are included in chapter 8 along with conclusions of what has been accomplished. The appendices include further details of the experimental setup in Appendix A, and an example of parameter optimisation in Appendix B.

Chapter 2

Reliability in CMOS Technology : A Review

2.1 Introduction

A CRITICAL REVIEW of reliability in CMOS technology and the impact of higher temperatures on devices is given in this chapter. The reliability of a product is defined as its ability to fulfill its intended function, under stated conditions for a stated period of time. Reliability associated failures in this case are caused by a failure mechanism, that is induced generally by internal and external stresses. How these mechanisms are monitored, understood and modeled has seen considerable research in the past. The following provides a review of this work looking at different reliability techniques that have previously aimed to predict device lifetimes. A critical review of past and present failure mechanisms is presented. Furthermore the influence of temperature and the growing demand for devices applicable for higher temperatures is reviewed.

2.2 CMOS Reliability Review

As early as the 1940s product and system reliability was identified as a major engineering discipline. However engineers have always worked towards reliable design. For example, in 1860, A. Wohler [1] presented some of the earliest fatigue failure information, which occurred on stagecoach and railroad axles. The applied load versus cycles to failure diagrams, which resulted from Wohler's work, were used to identify the load condition (called a fatigue limit) below which 'no failures' should be expected.

Although both technology and reliability have moved on significantly, the fundamental endeavour to design products to not fail still remains. The establishment of the Ad Hoc group on Reliability of Electronic Equipment in 1950 and subsequent Advisory Group on the Reliability of Electronic Equipment (AGREE) began reliability engineering in the electronics field. It was not until 1956 though that the first reliability methodology was conceived with the publication of the RCA release TR-1100, titled "Reliability Stress Analysis for Electronic Equipment," which presented models for computing rates of component failures [2]. It was this document that first applied the term reliability to integrated circuit manufacturing. It defined reliability as the lifetime during which a component is expected to perform its desired function. This definition can be represented graphically with the now well known mortality (bathtub) curve shown on the next page in figure 2.1. The first stage shows a high failure rate. This period is called the "early period," "burn-in period," "break-in period," or "infant mortality period." During this period, failures occur which are due to design or manufacturing weaknesses [3]. In other words, failure is due to weak or substandard components in which the probability of failure depends on how long the component has been operating. The next region is known as its useful life since the components can be used to the greatest advantage. Failures in this stage are mainly

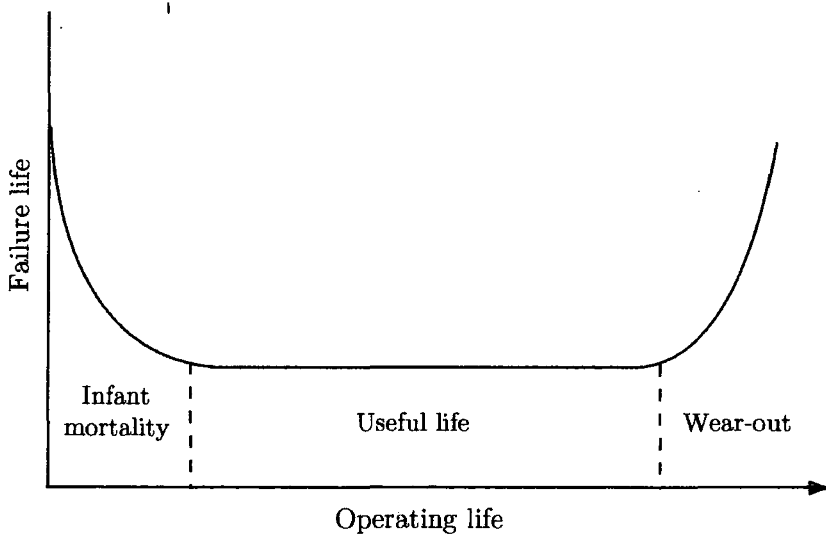


Figure 2.1: Bathtub curve

due to a low level of residual defects or electrical overstress/electrostatic discharge events. The failures are known as "chance," "random", "catastrophic failures" since they occur randomly and unpredictably. Failures are inevitably going to rise though as time goes on and this is what we see from figure 2.1. The final stage is the wear-out period where degradation failures begin to appear as a consequence of ageing or wear when the components are nearing their "rated life." In general, reliability is concerned with all three periods. In a lot of cases though, early life can be made as short as desired (even eliminated) by proper design, fabrication, and assembly, or by deliberate burn-in periods. Also if design and application are correct, the wear-out period should never be reached within operating life. This is because components are replaced as they fail during useful life, and each component is replaced (even if it has not failed) no later than at the end of its useful life.

The bathtub curve is ideal in illustrating reliability over the operating lifetime of a component. It is important also to consider how reliability changes throughout the whole product development cycle. Figure 2.2 illustrates this process from initial concept to production. Low reliability of the initial prototype could be a flaw in de-

sign, an unknown manufacturing process problem, or the cumulative effect of several environmental stresses. As these factors are determined and overcome throughout the development process, the reliability starts to improve. With the transition to the manufacturing production line though there is usually a fall in reliability. In many cases this is because the manufacturing production line environment can be very different from the environments of the research and development line. Additionally, at this point in the product life cycle the human involvement factor is usually at its most drastic transition, causing variances in the fabrication process.

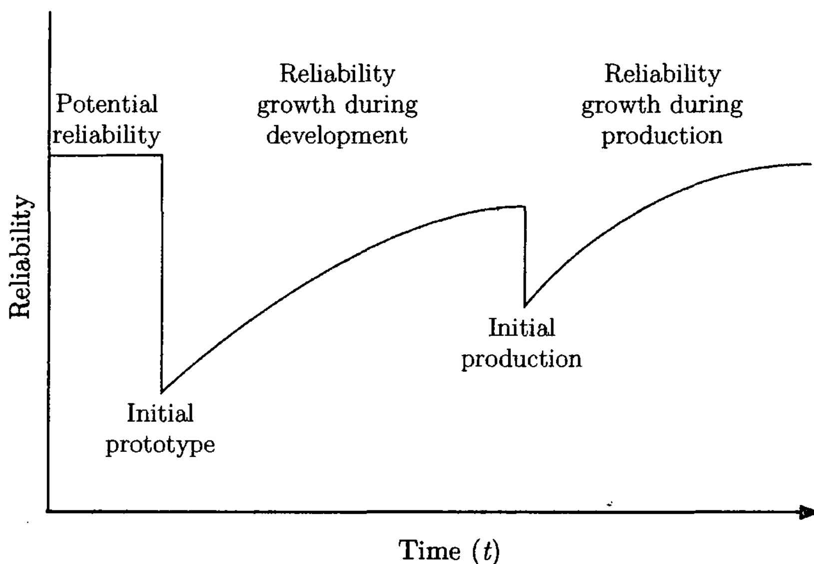


Figure 2.2: Product development cycle

From the manufacturing perspective it is very important to shorten this development time as much as possible. Traditionally, reliability assurance has relied mainly on failure detection at the end of the product development process. Due to the reliability analysis taking place towards the end of the development cycle any re-design work adds significantly to design time and cost. Ideally a manufacturer can use methods to allow reliability growth to occur in conjunction with product development. This would allow the manufacturer to gain a quicker time-to-market time of a reliable product that will not require costly warranty repair or replacement.

To quantify reliability we must take the concept of reliability as a probability distribution. One of the most common and simplest ways is to use the cumulative failure function *F(t)*. It is defined as the cumulative probability that a component fails at time *t*, or a fraction of the total number of components that have failed. It originates from the probability density function *f(t)*. This is a plot of the frequency at which components fail as a function of time divided by the whole population.

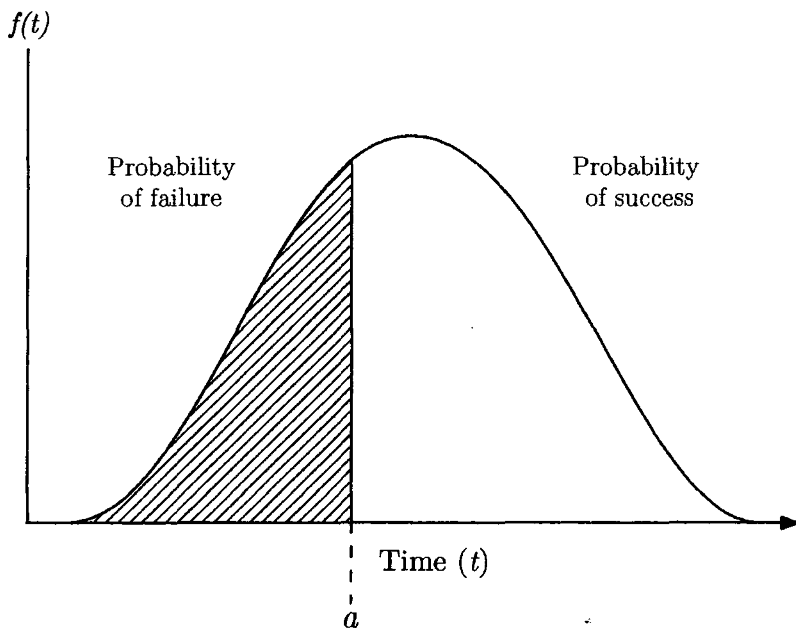


Figure 2.3: Reliability distribution

The probability density function can take many forms but one of the most common and simplest is the normal (Gaussian) curve. Figure 2.3 shows the normal distribution probability density function *f(t)*. This describes the probability of a failure occurring between two different points in time. However, in terms of quantifying component reliability, determining the probability of failure occurring before or after a certain time would be of greater use. This is where the cumulative failure function comes in. It measures the area under the failure probability curve, up to a given time, in this case, point *a*. This returns the probability of a failure occurring before

point a. Given that the area under the probability density function is always equal to 1, subtracting the cumulative density function from 1 would result in the probability of a failure occurring after a given time. This is the widely used reliability function *R(t)* and is defined as the fraction of the surviving good parts at any time.

$$R(t) = 1 - F(t) \tag{2.1}$$

The cumulative failure function *F(t)* is then found by taking the integral of the probability density function *f(t)* between zero and time a.

$$F(t) = \int_0^a f(t)dt \tag{2.2}$$

Therefore the reliability function *R(t)* can be expressed as:

$$R(t) = 1 - \int_0^a f(t)dt \tag{2.3}$$

$$R(t) = \int_a^\infty f(t)dt \tag{2.4}$$

From this follows the failure rate which is also known as the hazard rate function. The failure rate is the rate of change of the cumulative failure probability divided by the probability that the unit will not already be failed at time t. Thus it gives us the instantaneous failure frequency based on accumulated age and is defined as:

$$\lambda(t) = \frac{f(t)}{R(t)} \tag{2.5}$$

It is simply the probability density failure function divided by the reliability function, and has the units of failure per unit time among surviving parts, e.g. two failures per month. There is a clear and important distinction though to be made between this and the probability failure function. The probability failure function *f(t)* is the *unconditional probability* that the component will fail between to time a and time b.

The instantaneous failure rate $\lambda(t)$ is the *conditional probability* that the component will fail in the same time interval, given it has reached age T without failure. By analogy, the probability that a brand new component will fail in its fifteenth and sixteenth year of use is very small. However the probability of the same component failing in that same period, provided the component has lasted fifteen years is much greater. Another function that is very useful when trying to quantify component reliability is the the Mean Time To Failure (MTTF). This can again be derived from the probability density function *f(t)* and is the weighted average of all time values from zero to infinity, weighted according to the density. Hence the mean time to first failure μ is expressed as:

$$MTTF = \mu = \frac{\int_0^\infty t\ f(t)dt}{\int_0^\infty f(t)dt} \tag{2.6}$$

Given that the denominator will normally be 1, because the component has a cumulative probability of 1 of failing some time from zero to infinity, the mean time to fail is usually be expressed as

$$MTTF = \mu = \int_0^\infty t\ f(t)dt \tag{2.7}$$

For simplicity up to now I have only dealt with the normal probability distribution. The probability distribution curve can take many forms though that can resolve the above functions such as the MTTF. The common distributions including the normal distribution are the lognormal distribution, the Weibull distribution, and the exponential distribution. From figure 2.3 we can see how the normal distributions are appropriate when there is a strong tendency for the variable to take a central value, and positive or negative deviations from this central value are equally likely. Alternatively where components have a constant failure rate, it has been very common to use the exponential distribution.

The exponential distribution is described mathematically as:

$$f(t) = \lambda \, e^{-\lambda t} \tag{2.8}$$

where λ is the failure rate. The mean time to failure μ, and the reliability *R(t)* of this distribution is expressed as:

$$MTTF = \mu = \frac{1}{\lambda} \tag{2.9}$$

$$R(t) = e^{-\lambda t} \tag{2.10}$$

As an example consider a batch of 1000 devices, with each functioning device having a probability of 0.1 of failing on any given day, regardless of how many days it has already been functioning (constant failure rate). We would expect 100 devices to fail on the first day leaving 900 functioning devices. On the second day we would expect to lose 0.1 of our remaining devices, thus leaving us with 810 functioning devices. Given a constant failure rate the decrease in functioning components with respect to time is exponential. It is however important to evaluate whether the devices really do have a constant failure rate. This is because this form of distribution assumes the probability of failure for a functioning device at any given time is independent of how long it has already been functioning. Therefore it does not consider the 'wear-out' stage. However in general, components are replaced as they fail during useful life, and each component is replaced no later than at the end of its useful life.

In this section I have only briefly covered what I feel are two of the most frequently used distributions in reliability engineering. There is scope for much more detailed analysis and comparisons. Kapur and Lamberson [4] have written a book titled "Reliability in Engineering Design" which has an excellent review and detailed comparisons of all the distributions.

A vital and common element to all failure distributions and failure analysis is the collation of reliability data from real-life measurements. The most direct way to obtain the data would involve testing a large number of samples in normal use conditions and monitor their performance against the failure criteria over time. Given that most applications have lifetimes of several years this approach is clearly not practical. One method which has been widely used in the electronic industry has been accelerated life testing. This employs a variety of high stress test methods that shorten the life of a product or accelerate the degradation of the products performance. Thereafter the accelerated lifetime values for different stress levels can be used to extrapolate the lifetime to normal operating conditions. Using this method efficiently obtains performance data that yields reasonable estimates of the life of a product or performance under normal conditions. Common stresses used to accelerate the failure include temperature, voltage, humidity and vibration.

For a long time, mathematically determining how different stresses affect the lifetime has been fulfilled by the use of the Arrhenius equation. This relates how increased temperature accelerates the age of a product as compared to its normal operating temperature and is given by:

$$A_f = A\ e^{E_a/k_b(1/T_u - 1/T_s)} \tag{2.11}$$

where A_f is the acceleration factor, A a proportional multiplier, E_a is the activation energy of the chemical reaction or physical process (eV), k_b is Boltzmann's constant[1], T_u is the steady state use temperature (Kelvin) and T_s is the stress temperature. The activation energy is derived from empirical data gathered during accelerated testing and it represents the effect that the applied stress has on the product under test.

[1]Boltzmann's constant = 8.617×10^{-5} eV/K

The concept of using the century old Arrhenius relationship to model electronic component failure rates was first formally introduced in the RADC [2] release TR-1100, titled "Reliability Stress Analysis for Electronic Equipment,"[2]. From this the first reliability prediction handbook for electronics, MIL-HDBK-217A presented data to confirm that the Arrhenius relationship can be applied to the thermal ageing processes of semiconductor components. An important point to highlight is that the use of Arrhenius is only valid under the assumption that the dominant component failure mechanisms depend on steady state temperature. This assumption eliminates any failure mechanism that may have a temperature threshold below which the mechanism is not active, plus those which have a negative temperature dependence. Despite this, reliability requirements based on constant failure rates lead developers to continue to use this approach as a basis. It is the authors belief that reliability could be quantified much more accurately with not only the use of mathematical approaches such as Arrhenius, but also with a realisation of the physical cause of a failure. This is why this work follows an approach to fundamentally understand the physical mechanism behind the failure before any models and predictions are made.

2.3 Physics of Failure

Even with the MIL standard being updated by a number of companies and Universities such as Boeing and the Computer Aided Life Cycle Engineering Centre (CALCE) at the University of Maryland, rapid improvements and increased complexity of microelectronic devices pushed the application of MIL-HDBK-217B beyond reason. The last version of the document was published in 1991. At this time many companies had discarded traditional methods of reliability prediction that were included in this publication. In its place, they started to use a new reliability assessment technique, based on the root cause of a failure, called physics of failure.

[2]Rome Air Development Centre

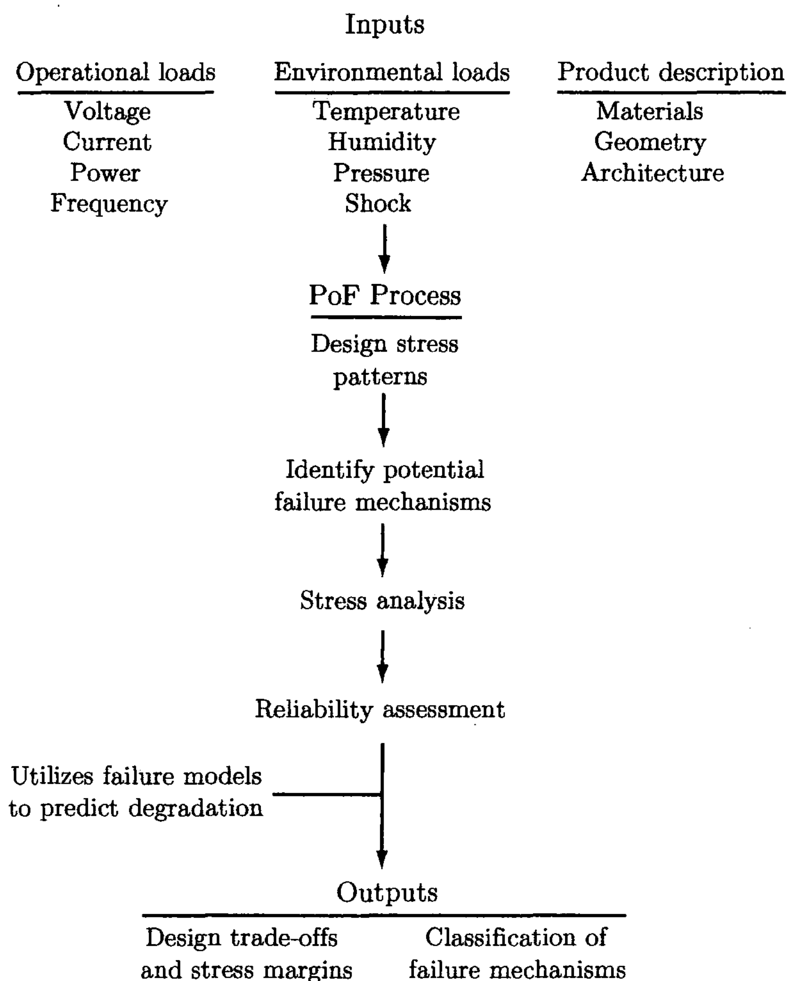


Figure 2.4: Physics of failure process

The underlying process behind the physics of failure approach is shown in figure 2.4. Physics of failure provides a method to understand the interaction, or physics, between a product's materials within specific use environments by identifying the underlying physical causes. This allows predictions to be made on how the failure mode will be affected by specific change to a device or technology. It is based on the fact that failure mechanisms are governed by fundamental mechanical, electrical, thermal, and chemical processes. Hence to exploit its potential, a firm understanding of the physical failure mechanisms present in the device is necessary.

The process begins with the preliminary design, looking at the device geometry, architecture and materials. Failure mainly results from the application of an over-stress, or by a lower stress level over a period of time. The next step is to define the applied life-cycle loads. Once the inputs have been established the process then moves to the main element of the physics of failure process: to determine the sources of reliability risks under life-cycle loads by identifying the potential weak-links and dominant failure mechanisms. This step is best described in two stages.

- Stress Analysis - to decide how the stress factors affect the device. This is to identify which failures may occur, where they occur, their cause (failure mode) and the consequences of such a failure.
- Lifetime Prediction - after identifying the likely failure mechanisms, the analysis involves the reliability prediction. Implementation of device models allows device degradation and lifetimes to be predicted. This allows not only the most dominant failure mechanism to be identified, but also to plot trends of the less likely failures. This will allow prediction of changes in failure distributions that would not be noticed in an approach that only identifies the most dominant failure mechanism.

In addition to establishing the failure modes of a device and the dominant mechanisms, a physics of failure approach can be used to ensure the device has adequate stress margins and that it meets the reliability targets. By understanding the root causes of the failure it also allows potential problems in technology changes, i.e. technology change to be identified and solved before they occur. However, the use of physics of failure is only one part of the complete process for predicting reliability. For complete reliability analysis the physics of failure or any other approach must fit into a reliability methodology or simulation tool.

2.4 Failure Methodologies

The traditional reliability prediction methods such as the MIL standards were based on analysis almost at the end of the design process. This meant that at this point in the product development cycle only minor design changes could be made. Also the tests often did not address the actual failure mechanisms occurring in the application environment. The stuck at fault test is one such example. Classically designers assumed they could model a circuit dependent on each node being stuck at one or zero. Banerjee *et al* [5] proposed that in a real life environment an MOS circuit can produce logic levels that cannot be classified as either a zero or one. Today this hypothesis has clearly been proven with the emergence of soft failures such as stress induced leakage currents that in some cases lead to a degradation but not necessarily complete circuit failure. Furthermore, some manufactures have previously used the same models for both temperature and voltage acceleration. It is unlikely that we can accept that in case of failure one activation energy can be used for all failures in acceleration factor calculations, knowing that different failure modes have different dependencies on temperature and voltage.

Shoucair [6] previously called for the development of new CAD tools to include a number of features that would be desirable for purposes of rendering high-temperature IC design more accurate, flexible, and realistic. It is the authors belief that the same call is now needed to advance the understanding again and provide a novel tool to include the most recently discovered failure mechanisms. This is something Fergerson and Shen [7] highlighted in their analysis of reliability methodologies during the last decade. Some good developments have been made, such as that of Li *et al* [8]. Li and his co-workers proposed the Failure Rate-Based SPICE (FaRBS) reliability simulation methodology. It combined modules of SPICE simulation with wear-out models and acceleration factor models to identify critical or failure prone

circuit block in a system. Both McClusky *et al* [9] in the form of Computer Aided Design of Microelectronic Packages II (CADMP-II) and Charpenal *et al* [10] in the form of FIDES developed similar tools. FIDES arose from a consortium of industry groups and was particularly interesting from the aspect it was designed to incorporate Commercial Off The Shelf (COTS) components into the reliability assessment. A number of simulation tools have been developed that concentrate on hot-carrier effects. Texas instruments developed HOTRON [11] as a proprietary simulator of hot-electron effects. The University of Southern California developed RELY [12] to simulate hot-electron effects and electromigration. Also hot-carrier simulation tools BERT[3] [13] and CAS[4] [14] were developed by the University of California, Berkeley.

Many of these tools have been developed for relatively large devices in comparison to today's technologies. Additionally at present there is not an accurate simulation tool that can take into account a range of temperatures and includes the most recently discovered failures mechanisms for deep sub-micron devices. In 2006 Li *et al* [15] stated that the advancement of device failure modelling had fallen behind the development of CMOS technology which has raised many new issues related to both circuit performance and reliability. Success in today's globally competitive environment requires the ability to repeatably examine and constantly improve manufacturing processing. Thus a time, and hence cost effective reliability test tool is required. For the future generation of devices, it is no longer practical to design the product and then measure and improve the reliability at the latter stages of development. A physical understanding of the failures to allow identification and prioritisation in the early stages of development is the key. The development of this idea continues in chapter 7, but it is clear to see that something new is needed to cope with the combination of elevated temperatures and any new failure mechanisms the dramatic changes in technology growth may have brought with it.

[3] Berkeley Reliability Tool
[4] Circuit Ageing Simulator

2.5 Failure Mechanisms

As well as providing the methodology for predicting reliability, if a physics of failure approach is to be harnessed it is critical that there is a fundamental understanding of the failure mechanisms themselves. The history of CMOS development has been dogged by three major causes of IC failures. These have been electromigration, gate oxide wear-out and hot-carrier effects.

2.5.1 Electromigration

Electromigration is the dominant mechanism in terms of interconnect failure and has been a major failure mechanism in discrete solid state devices since 1970. The first ICs were constructed with pure aluminium (Al) metal lines that were 10μm in width or more. At the same time they were very thin, in the order of 3000A. This meant a high current density was being passed along a material with a low melting temperature, which implies fast diffusion at low temperature. This combination was a recipe for disaster and failed parts were soon returned from the field. Under inspection many failures were caused by very fine cracks in the metal resulting in open circuits. This structural damage was due to ion transport in metal lines. If the current density is high enough, the momentum exchange between conducting electrons and diffusing metal atoms can be significant, resulting in noticeable mass transport generating electromigration damage. The damage manifests itself in the movement of vacancies and interstitials. The vacancies coalesce into voids and interstitial become hillocks. The voids in turn, decrease the cross sectional area of the circuit metallisation and increase local resistance and current density at that point in the track. Conversely, hillocks can cause a short to the adjacent or overhead metallisation.

Modeling of electromigtation was first significantly contributed to by a landmark paper by Black [16]. He proposed an empirical model for prediction of electromigration lifetime as a function of current density J, and metal line temperature in the following form:

$$t_{50} = \frac{A}{J^n}\, e^{\frac{E_a}{k_b T}} \tag{2.12}$$

Where t_{50} is the median time to failure, A is a material and process dependent constant, k_b Boltzmann's constant, T the absolute temperature, and E_a is the activation energy for failure. Traditionally, according to Blacks original work it was observed that electromigration failure followed a $1/J^2$ inverse square law. After 1967 most subsequent experiments indicated a current exponent ranging from 2 to 3, usually 2 and although the mechanism was not clear, most experiments were consistent with Black's relationship [17]. There is an assumption though in Black's model, that the mechanism leading to electromigration induced failures under the accelerated lifetime test conditions are the same as those under device operating conditions.

At the time that Black developed his model, the physical solution to the electromigration problem was simple: make the tracks thicker. However as device miniaturisation demanded smaller and smaller interconnects, this became no longer a solution. Today tracks are often in the order of 0.5μm thick, carrying a current of 1mA which can result in a current density of 10^6A/cm^2. The continued increase in current density over the past 40 years has meant new methods to counter the electromigration problems have needed to be devised. The first idea to make the metallisation more resistant to electromigration was to add a small amount (up to 4%) of copper to the aluminium. This solution bought some time but in the past five years there has again been the need to develop a better interconnect for two main reasons. Firstly, there is a resistance-capacitance delay in fine tracks. The use of narrower lines, not only increases the resistance, increasing the likelihood of electro-

migration but also the capacitance between the lines increases, bringing down signal propagation. The second problem is the high cost of building a multilayer interconnect structure. The solution adopted by industry has been to turn to pure copper as the on-chip conductor material. Since copper has a much higher melting point (1083°C) than aluminium (660°C), diffusion should be much lower in copper than aluminium at the same device operation temperature [18]. Experiments by Tu [18], Lloyd [19], and Tao *et al* [20] on copper tracks found the electromigration lifetime increased. Tao *et al* [20] found that in both DC and pulsed-DC tests, the lifetimes of the pure copper interconnects were about one and two orders of magnitude longer than those of aluminium alloyed with copper. His work was also interesting in that it aimed to make the interlayer as thin as possible. Because there had been issues with adhesion of copper to the $Si0_2$ dielectric layer [18], previous test structures had employed a interlayer metal with significant thickness (compared to the copper thickness) to improve adhesion. Tao's and his colleagues work cleverly used a seeding layer of 15nm Cobalt. Firstly, this solved the adhesion problem as cobalt adheres well to $Si0_2$. Secondly with a reduction in interlayer thickness, its effect on the electromigration properties of copper interconnects was kept to a minimum.

Developments in the use of copper has helped to make its introduction easier. Some authors [19, 21] still see the processing problems of using copper to be an issue. Further downscaling and also technological advancements continue to require more reliable interconnects under conditions where metallisation is inherently less reliable. One example is the introduction of flip chip solder joints, which tend to become weak links in the system. Due to the unique geometry of the joints, current crowding at the contact interface causes increased current density and electromigration damage. This and a number of other reasons documented by Tu [18] mean electromigration in flip chip solder joints is now competing with on-chip metallisation as the major electromigration reliability problem in microelectronic devices.

Although this is a very brief history of its origins, problems and solutions, it is clear electromigration is always going to be an important reliability issue. The promise of developing future metallisation schemes that will eliminate the problem has so far not been solved [21]. Recent advances have shown that with careful design electromigration can at least be controlled such that advanced microcircuits can be designed with the required reliability. As a result it was felt it would be more beneficial for the time spent on this project to be aimed towards other more prominent failure mechanisms. Although electromigration for a long time has been seen as the major failure mechanism is CMOS devices, continued device scaling has brought about more pressing issues needing study including gate oxide failures and hot-carrier effects. This can also be seen as a reflection of the amount of current research and publications work in this area. The following section goes on to highlight such failures and why it is the authors belief they pose the greatest reliability problems for future technologies.

2.5.2 Gate Oxide Failures

In today's technology, transistor dimensions have been scaled dramatically and one of the most critical issues of this scaling has been the reduction of the gate oxide thickness [22]. Table 2.1 on the following page shows the road map of MOSFET minimum feature size as predicted by the semiconductor industry, showing the prospect of less than 2nm dielectrics by the year 2014 for standard CMOS fabrication. As a result of the reduction in oxide thickness, oxide related failures have increased and many see these to be the limiting failure modes for future technologies [22, 23, 24, 25, 26]. With oxide thickness now being measured in atomic units, there is cause to be concerned and subsequently a major element of this research aims to investigate the failure modes that are currently a problem and which pose a limiting factor for further scaling.

Parameters		1991	1994	1997	2001	2005	2007	2009
Channel length	(μm)	0.5	0.35	0.25	0.18	0.13	0.09	0.065
t_{ox}	(nm)	13.5	9.0	8.0	7.0	4.5	3.5	2.7

Table 2.1: MOSFET road map

For many years gate-oxide was more of a production problem than an intrinsic reliability problem. Most of the research conducted was concerned with extrinsic failures as the intrinsic quality of the oxide never posed a serious reliability threat under typical operating conditions [22]. However, over the past ten years as oxide thickness (t_{ox}) has decreased, the intrinsic reliability limits have been approached and the study of gate-oxide breakdown has become no longer limited by extrinsic failures. The dominating factor to account for this was the change in feature size without proportional scaling of the power supply voltage. This resulted in a significant increase of the horizontal and vertical electrical fields in the channel of the device [27]. As a result new problems that posed an additional constraint on the acceptable supply voltage and or oxide thickness became evident for deep sub-micron devices [28, 29, 30].

2.5.2.1 Gate Induced Drain Leakage

Gate-induced leakage current is the name given to a tunnelling effect that occurs in the gate-to-drain overlap region of MOS devices in the OFF-state. It is caused by high transverse and lateral electric fields close to the drain region, generated by gate-source (V_{GS}) and drain-source (V_{DS}) voltages [31]. Huang *et al* [32] presented a paper in 1998 that recognised GIDL current as the major drain leakage current phenomenon in OFF-state MOSFET's. Before this Chan *et al* [28] reported GIDL as a new mechanism in 1987, showing the presence of leakage in the sub-half micron regime. A significant drain leakage was detected in thin gate oxide (15.5nm) MOSFET's at drain voltages much lower than the junction breakdown voltage.

Chan *et al* discussed device design considerations for minimising the GIDL current. In order to limit the leakage current to 0.1pA/μm of the channel width, it was proposed that the oxide field in the gate-to-drain overlap region should be limited to 1.9MV/cm. An important proposal by Chan *et al* was also to attribute the gate-induced leakage to the band-to-band tunnelling process. The belief that band-to-band tunnelling is the root cause of GIDL has continued to the present time. Endoh *et al* in 1990 [33], Guo *et al* in 1998 [34], Chen *et al* in 2001[35], Touhami *et al* in 2002 [31], and Lopez *et al* in 2004 [36], all published notable papers stating that they saw band-to-band tunnelling as the major leakage mechanism responsible for GIDL. Endoh *et al* [33] was one of the first to propose a model to describe the band-to-band tunnelling effect and the leakage current assessment. It built upon Chan's *et al* [28] model that neglected the dependence of the transverse electric field on the drain doping profile and used a fixed value of band bending. Both Chen *et al* [35] and Touhami *et al* [31] developed models that considered doping concentration, vertical field, and lateral field with results showing good agreement with measurement data over a range of gate and drain biases. None of the above authors mentioned above have considered the influence of temperature on GIDL. It has not been until very recently that researchers have started to investigate this [37].

Limited data is available on how temperature affects GIDL current. It was proposed in a report by Slisher *et al* [38] that GIDL is independent of temperature whereas both Lopez *et al* [36] and Bouhdada *et al* [39] presented results showing an increase in GIDL at elevated temperatures. Lopez *et al* [36] hypothesised that the variation in leakage was due to the variation of the energy band gap of silicon with temperature. Bouhdada *et al* [39] had an alternative outlook. Again, results showed an increase in GIDL current with temperature but it was believed to be due to increased carrier generation. Several authors have studied how changes in device dimensions and process parameters have effected GIDL. Chung *et al* [40] found that GIDL is

enhanced by either decreasing oxide thickness, or by increasing drain voltage. In the same study GIDL was seen to be virtually independent of channel length because the tunnelling depends only on conditions in the immediate gate-to-drain overlap region. Bouhdada *et al* [41] found that at higher doping densities, the electric field increases in the overlap region, encouraging carrier generation, and hence causing an increase in GIDL current.

The review of published data has provided a valuable background to GIDL. I believe there is still a requirement for the development of a computationally efficient expression for GIDL ideal for use within a complete reliability tool. Currently there is only one GIDL model for use in a SPICE like simulator, which is a complicated expression in the BSIM4 model [42]. Our aim is to develop a simple, easy to use model, that would ideally have the capability to predict the effect of high temperatures on deep sub-micron devices. Chapter 5 details the development of one such model along with further detailed analysis of the physical mechanism behind GIDL.

2.5.2.2 Negative Bias Temperature Instability

The degradation phenomenon of Negative Bias Temperature Instability (NBTI) is characterised by an increase of threshold voltage and decrease of drive current, as a result of elevated temperatures and a high negative gate bias stress. Typical stress temperatures lie in the 100-250°C range with oxide electric fields typically below 6 MV/cm i.e fields below those that lead to hot-carrier degredation. Potentially parameter variations caused by NBTI can lead to transistor mismatch, timing issues and reduced switching speed. Given that the number of digital transistors approach millions of devices in modern products, such NBTI degradation characteristics have today become a critical issue for circuit designers.

NBTI has been known since the very early days of MOS device development, having been observed as early as 1967 [43]. Deal [44] named it Instability Number VI. Goetzberger [45] and his co-workers at Bell laboratories were one of the first groups to investigate NBTI degradation. They stressed 100 nm oxides at 300°C and found an increased interface trap creation for negative gate voltage stress compared to positive stress. The generation of interface traps under negative gate voltage stress was discovered to have a time dependence of $t^{0.25}$. Their investigations also found a higher trap density for p-channel devices compared to n-channel devices. Nevertheless, NBTI effects remained marginal for many years, especially when compared to mechanisms like hot-carrier effects [46].

NBTI effects have become a critical reliability issue for modern deep sub-micron devices in recent times for three main reasons. Firstly slower scaling of operating voltages for both digital and analog circuits compared to more aggressive oxide thickness scaling has gradually increased the effective field across the oxide. Secondly thinner oxides have brought the poly-silicon gate closer to the $Si/Si0_2$ interface increasing the chances of hydrogen diffusion, trap creation and NBTI susceptibility. Thirdly the introduction of nitrogen atoms into the oxide has enhanced NBTI degradation. To increase the dielectric constant nitrogen was added, intended to improve hot-carrier resistance and gate current leakage. However, nitrogen tends to increased NBTI problems. Tan *et al* [47] presented some excellent results showing the enhancement of NBTI with incorporation of nitrogen into the oxide. Comparisons between pure $Si0_2$ and various interfacial nitrogen concentrations showed an increase in NBTI degradation, sometimes by a factor of 10 as in the case when comparing pure $Si0_2$ to 15.5% nitrogen concentrate. These results were complimented by Huard and his co-workers [46] who further investigated the effect of nitrogen, showing an accelerated NBTI degradation for increased nitrogen concentrations.

As well as highlighting the influence of nitrogen on NBTI, Huard *et al* [46] also presented some results on oxide field and temperature dependence of NBTI. A range of gate oxide thickness (1.6 to 6.5nm), were stressed for ten thousand seconds at varying gate voltages and temperatures. The investigation found that both trap creation and charge trapping were increased for higher gate voltage stress levels. Results also showed an increase in threshold voltage shift at higher temperatures, which was proposed to be due to increased interface trap creation. Similar characteristics were observed by Aono *et al* [48] and Mahapatra *et al* [49] who pointed out that ΔV_{th} has a tendency to saturate with increasing stress time. This is a characteristic noted by Goetzberger *et al* [45] over thirty years ago. Since then many groups have reported the time-dependent shift to follow a power law dependence with an exponent around 0.25. This value was what early studies by Jeppson and Svensson [50] proposed but there after the value has been widely debated. As Ershov *et al* [51] stated, one of the reasons for this debate is that the exponent is not a universal parameter, and may depend on stress and measurement conditions. This was proven by Aono *et al* [48] who showed that the coefficient decreases with increasing stress time and can vary from 0.3 down to 0.16. Further deviations from the single exponent power law were presented by Alam *et al* [52].

A key characteristic of NBTI first noted by Schlünder *et al* [53] in 1999 was a recovery effect. He noticed that the combination of high temperature and high positive gate voltage resulted in strong relaxation of the NBTI effects. Ershov *et al* [51] performed experiments with measurements taken during and after sequences of stresses revealing the the recovery contains a fast initial transient followed by a very slow non-exponential transient, which appears to saturate with time. This would suggest that NBTI degradation contains a level of permanent damage, possibly related to fixed positive charge in the gate oxide. The obvious question that springs to mind when considering the recovery effect, is how NBTI degradation would be

effected by dynamic stressing. The majority of published research has focused on static stressing. Tan *et al* [54] presented some insight into the effects of pulsed stress finding an increased lifetime enhancement for dynamic signals compared to constant voltage stressing. Chen *et al* [55] presented results showing dynamic NBTI (DNBTI) degradation was less than that of equivalent static NBTI (SNBTI) stress and DNBTI effects were frequency independent. In contrast Mahapatra *et al* [49] proposed that interface trap creation and thus $\triangle V_{th}$ decreased at higher frequencies.

Exact modelling of NBTI has been made difficult because the physics and root cause of the degradation is still not fully understood [46]. It is believed that NBTI is a result of a build up of positive charge either at the $Si/Si0_2$ interface or in the oxide layer, which in-turn leads to an increase in threshold voltage. A number of authors [43, 49, 51, 56] have attributed the formation of this positive charge to the diffusion of hydrogen away from the $Si/Si0_2$ interface. Mahapatra *et al* [49] explained this process using the Reaction-Diffusion (R-D) model. The model states that under NBTI stress the interface trap is due to the dissociation of Si-H bonds at the $Si/Si0_2$ interface (reaction) and the subsequent movement of released H species away from the interface (diffusion) leaving behind a positively charged interface state. According to the R-D model, the movement of hydrogen species back to the interface when the stress is removed would also give an explanation to the the recovery effect. This is what Mahapatra *et al* [49] proposed but Huard *et al* [46] argued against this. His results compared the reduction of the threshold voltage with interface trap density after stress removal. The results showed a strong decrease (more than 50%) of the threshold voltage whereas the interface trap density remained almost unchanged. He therefore proposed that the recovery effect should not be accounted for by the re-passivation of hydrogen, but failed to detail any alternative reason.

It is acknowledged that'at present exact determination of the the process that leads to NBTI degradation and its recovery is not fully understood. It is clear also just from this review that there are a number of issues still be resolved, including the effects of dynamic stressing. The use of models such as the reaction-diffusion type has helped put together a more complete understanding of NBTI physics, but it could be said that the NBTI signatures noticed by Goetzberger [45] and his co-workers over thirty years ago are still valid today.

2.5.3 Hot-Carrier Effects

Hot-carrier damage leads to degradation of the threshold voltage, transconductance (on-resistance), and drain current driving capabilities. The driving force behind the degradation is the channel electric field. This has always posed an important consideration to the development of CMOS transistor technology and there has been substantial research into hot-carriers effects and the cause of the degradation [57, 58, 59]. However the repeating problems brought about by aggressive scaling has resulted in its effects being enhanced in recent years. Ning *et al* [60] was one of the first to recognise the problem hot-carriers posed to device scaling. Chen *et al* [61] continued Ning's early work to apply the lucky electron model in the study of substrate hot electron injection. The lucky electron model developed by Shockley hinges on there being a supply of 'lucky' hot electrons. Two factors that constitute the electron being 'lucky' are firstly if it has enough kinetic energy from the channel to become 'hot'. Within the literature several analytical models for substrate current behaviour have been reported [57, 62, 63]. Hu *et al* [57] used the lucky electron model as a basis for a landmark paper that explored the use of the substrate current as a measure of hot-carrier effects. Using the substrate current as a link to hot-carrier degradation he proposed a method for predicting device lifetime that has been used extensively since by numerous researchers.

The model presented by Hu *et al* [57] was based on the belief that both the substrate current and hot-electron effects had a common driving force - the channel electric field, or more specifically the maximum channel electric field E_m. Since Hu's work, several authors have attempted to refine his original substrate current model. Kolhatkar and Dutta [64] proposed a model but some of the fitting parameters used in their analytical approximation were difficult to extract. In recent years a number of lifetime models have been developed that characterise the hot-carrier effect for modern sub-micron technologies [57, 65, 66]. These semi-empirical models were shown to be valid only down to 0.25μm technology [66]. However for smaller technologies as used in this study it needs to be confirmed if they are still valid.

There has been limited research in the temperature dependence of hot-carrier effects. One reason being that it is generally believed that the substrate current and hence hot-carrier effects decrease at higher temperatures. A comprehensive study by Bravaix *et al* [67] confirmed this finding that on one hand high temperatures cause mobility degradation which reduces the current gain and device switching speed, but on the other hand, also contributes to a reduced sensitivity of the device characteristics to hot-carrier effects. However recently there has been some concern that the assumption that hot-carrier degradation is reduced at elevated temperatures may not be true for deep sub-micron devices [67, 68].

As we approach oxide thickness on an atomic level and higher electrical fields than ever before, it is the author's belief that further hot-carrier consideration is needed. This is particularly needed to further the understanding of the hot-carrier mechanism and to clarify its temperature dependence. Additionally, most studies concentrate on hot-carrier effects in n-channel devices only, due to increased degradation [69]. We aim in our studies to investigate both types of devices for complete analysis.

2.6 High Temperature Effects

Devices have found their way into many applications with varied conditions, and one aspect in particular is a growing demand for devices applicable for higher operating temperatures. Existing semiconductor products are typically released for maximum ambient temperatures of 125°C although many applications require devices to operate at much higher temperatures. Three of the largest potential application of high temperature devices are within aerospace, automotive, and well-logging [70, 71, 72]. An example where reliable high temperature devices were needed was seen in a Russian remote space exploration to Venus. The Venera 1 was prevented from probing the surface of Venus and ended up only gathering atmospheric data due to the lack of electronics capable of withstanding the higher temperatures [73, 74].

There are a number of physical effects that make it difficult to build devices that operate reliably at high temperatures. Among the prominent limitations of conventional CMOS technology are increasing intrinsic carrier density, and junction leakage currents increasing the chance of latch-up. Decreasing mobility and threshold voltage add to the degradation of device performance with implication to noise margins and switching speeds in digital circuits. Decreasing gain-bandwidth products, and increasing input-offset voltages are also likely performance problems faced in analog circuits at elevated temperatures. These effects are well understood with noteworthy contributions from a number of authors. Prince *et al* [75] observed a decrease in threshold voltage with increasing temperature, while Palkuti [76] *et al* studied integrated circuits up to 300°C reporting degradation's due to leakage currents, and reduced mobility. Shoucair has provided a great wealth of significant research, with analysis of digital and analog circuits at elevated temperatures [77], detailed experimental and analytical models of leakage currents [78], and future outlooks relating to issues of downscaling, and process evolution [79].

2.7 Alternative Materials

Concerning the use of alternative materials and technologies, this research has focused solely on bulk CMOS. A number of wider band-gap semiconductor materials such as Silicon Carbide (SiC), diamond and nitrides, offer promising alternatives for high temperature applications. The author does not dismiss the clear advantages of using wider band-gap materials and recent developments by a number of researchers [73, 80, 81, 82, 83] in this area show great promise. Table 2.2 [84] shows the key properties of a range of wide band-gap materials.

Property		Si	GaAs	SiC	GaN	AiN
Band-gap	(eV)	1.12	1.43	3.26	3.4	6.1
Breakdown field	(V/μm)	30	30	250	250	1200
Thermal conductivity	(W/cm K)	1.5	0.5	4.5	1.5	3.3
Saturated velocity	(cm/s)	1E7	1E7	2E7	2.4E7	1.8E7
Electron mobility	(cm^2/V s)	600	4000	400	1500	-
Hole mobility	(cm^2/V s)	150	-	-	30	-

Table 2.2: Comparison table of different semiconductor materials

With a bad-gap of 1.43eV, versus 1.12eV for silicon, Gallium Arsenide (GaAs) appears to have the potential for higher temperature operation. The onset of diffusion-dominated leakage currents generally occur at much higher temperatures than in Silicon, because of the relatively higher band-gap [83]. Fricke *et al* [85] presented a GaAs fabrication that provided reliable device operation up to 300°C, demonstrating the possibilities of high temperature operation. There is a major difference though between Si and GaAs in the fabrication process. Si forms a tough, adherent native oxide, while GaAs oxides are not adherent. This means that the conventional MOSFET cannot be made in GaAs, and MESFETS's are the most common transistor type in GaAs technology [80]. With no oxide in a MESFET, the channel is isolated from the gates by reverse biased Schottky junctions. In turn experiments

by Shoucair *et al* [83] have shown the MESFET's to display a significant high gate leakage current. Dreike *et al* [80] also presented data highlighting the problem of leaky gates in MESFET's plus problems faced particularly in the area of achieving reliable ohmic contacts to the devices. Both Dreike and Shoucair agree that further developments in GaAs technology are needed to exploit its full potential but also state these efforts would likely be better spent on a larger band-gap semiconductor or heterostructure technology for high temperature operation [80].

Silicon Carbide has been investigated for a number of years for its potential use as a wide band-gap semiconductor [86, 87]. During the late 1960s, SiC was considered to be the semiconductor material of the future [73]. Recent research in material growth, doping, oxidation, and metalisation has seen considerably more commercial interest and activity in the past decade. Cree Inc. in particular have carried out extensive R&D. They have now launched their second generation SiC microwave metal semiconductor field effect transistor (MESFET) process. Aimed towards the wireless base station business it was a 10W power device which operates up to 2.4 GHz and apparently showed no degradation at 175°C and only limited degradation at 240°C junction temperature. GE Global Research Centre have also been looking at SiC devices. In 1993 they developed a NMOS integrated circuit, including a 300°C operational amplifier. However it was reported they faced oxide reliability problems with the $SiC/Si0_2$ interface prone to interface state densities and poor dielectric reliability at elevated temperature [84]. Another focus of research is recent years has been the growth of nitrides epitaxially on dissimilar substrates such as GaAs, SiC and sapphire. The growth of Gallium Nitride (GaN) or Aluminium Nitride (AlN) to form a heterostructure with a SiC substrate has shown improved electrical and thermal performance. They offer many potential solutions to high temperature problems such as reduction in leakage but have suffered from high development costs.

Invented in the 1960's at Rockwell, Silicon on Sapphire (SOS) is an additional technology that has long held great promise. Early on it was recognized for its high speed and low power potential. Peregrine Semiconductor developed an ultra thin silicon process which left a high quality ultra thin layer of silicon on a insulated sapphire substrate. Elimination of the substrate capacitance allowed for higher speeds at lower power and avoided voltage dependent capacitance distortions. Kronberg [88] compared results of SOS devices with comparable bulk CMOS devices. He found SOS devices showed most of the effects seen in bulk silicon MOSFET's. The SOS devices though were found to have much smaller leakage levels and useful characteristics were seen at 300°C. The major trade off for SOS versus other technologies, bulk silicon in-particular, is cost. The higher material starting cost of SOS technology is a difficulty for development and as a result SOS is still to become a viable alternative.

The above alternatives clearly posses potential as materials suited for harsh environment electronics. However the most common low power and high speed device technology today is still silicon CMOS [6]. Along with the interests of the project sponsor, it is the authors belief that to maximise the capacity of this project conventional CMOS technology is the viable option.

2.8 Silicon-On-Insulator Technology

The use of Silicon-on-insulator (SOI) technology has come in response to the desire for increased performance and lower operating voltages as conventional CMOS technology approaches its fundamental limits. Silicon-on-insulator devices differ from bulk CMOS by placing the transistors silicon junction area on top of an electrical insulator. Advantages over bulk devices include speed, power dissipation and reduction of leakage currents [89].

For a long time SOI has been seen as a serious contender to substantially extend the temperature range of silicon integrated components [90]. The device structure reduces high leakage currents that can lead to latch-up occurring in bulk CMOS. Because each source and drain junction is surrounded by insulator, the junction capacitance is largely reduced compared to bulk silicon transistors. This means the circuits can operate at higher speeds or substantially lower power at the same speed. Results [90, 91, 92, 93] have demonstrated SOI technology to offer a real opportunity for integrating at reasonable cost, digital and analogue circuits which can withstand temperatures in excess of 300°C. One important feature is smaller threshold voltage shifts. Reichert *et al* [94] showed reduced threshold voltage shift for elevated temperatures up to 200°C.

One company that has developed silicon-on-insulator technology for high temperature use is Honeywell electronics. They have developed a reliable high temperature process designed to produce integrated circuits capable of operation up to 300°C. The process is a fully isolated 1.25μm analag/0.7μm digital process. Reduction in leakage currents is provided by all junctions bottomed out to the isolation, while maintaining transistor switching frequencies through reduced junction capacitance. Examples of SOI device fabrication by Honeywell, are a family of high-performance static random access memories designed for high temperature operation, that have demonstrated over 2500 hours of operation at 300°C under dynamic burn-in-conditions. Very recently the application of silicon-on-insulator devices for high performance products has also seen significant progress. Companies such as AMD and IBM have started to use SOI technology for their ultra deep sub-micron processes. Both seem to have harnessed the technology successfully without any reliability concerns and have shown a 20-25% improvement in device cycle time compared to that of similar CMOS devices.

2.9 Conclusion

The aim of this chapter has been to provide a critical review of reliability in CMOS technology with emphasis on those parts of the subject that are of particular interest to the research in later chapters. This has covered classical reliability functions such as the bathtub curve, the reliability probability distribution, the Arrhenius equation through to new reliability approaches such as physics of failure.

An overview of reliability methodologies is given, highlighting the drawbacks of traditional methods such as the stuck at fault model, and new alternatives identified. Proposed is the development of a novel reliability simulation tool to allow fault identification in the early stages of development, something vital in today's market. Such a simulation tool requires a clear understanding of the root causes of failure, of which the most prominent mechanisms were reviewed in section 2.5.

Electromigration was once the ruling CMOS failure mechanism but aggressive scaling of device dimensions has revealed new prominent failures. With gate oxide thicknesses approaching atomic levels concerns over failure mechanisms such as Gate Induced Drain Leakage, and Negative Bias Temperature Instability have arisen. The review of present GIDL studies highlighted the need for accurate deep sub-micron models that consider the effect of elevated temperatures. NBTI is a new degradation phenomenon, the root cause of which is still not fully understood. Investigations are required into the effects of dynamic stressing, and the development of a new model capable of predicting the cumulative and recovery effect. The past and present significance of hot-carrier effects is also reviewed. This points towards the need for further consideration to whether the classical belief that hot-carrier effects are reduced at high temperatures is still valid for deep sub-micron devices.

In section 2.6, the additional influence and difficulties presented by elevated temperatures is given. The largest potential users are within the aerospace, automotive, well-logging and space applications. Additionally a review of past research into degradation factors related to enhanced temperatures, shows an expected decrease in threshold voltage and channel mobility, while an increase in junction leakage currents. These issues are addressed again in more detail in chapter 3.

Consideration of alternative materials and technologies concluded that despite having some unique advantages for high temperature use, there are still key challenges in process and manufacturing that make them not a viable option for this project. Cost, and process availability means any new ideas are much more open to viable development and implementation within bulk CMOS. The author believes for near-term research and development pay-off bulk CMOS and possibly commercial off the shelf components are the most practical options.

The author concludes from the literature review that there must be a clear understanding of how the product will fail to accurately quantify the reliability of electronic products. This means a critical understanding of the physical root causes behind the failure mechanisms. In particular failures related to the gate oxide and its interface are in need of further investigation for deep sub-micron devices. Bulk CMOS is by far the most widely used and viable technology for development and research. It provides the greatest potential gain for creating new methodologies to predict device degradations, failures, and ultimately allow problems to be spotted before they occur. This principle forms the basis for the following chapters and the design of an integrated reliability methodology.

Chapter 3

High Temperature Device Model

3.1 Introduction

DEVELOPMENT OF A new simple but accurate high temperature MOSFET model is demonstrated in this chapter. This model is designed to use fewer parameters than other models, whilst accurately predicting the effects of temperature on state-of-the art devices.

During the design stages of any model I believe there are two words which should always be kept in the back of the designer's mind: accuracy and efficiency. They are of course conflicting terms, but as designers we are responsible for defining the compromise. Accuracy is obviously an important consideration but must not be the overriding aim. We must consider efficiency not only in terms of computational time but also in the model's usability. How well the end user understands the model is a sign of good design. Therefore, providing the end user with a model that is both accurate but also useable, is a main aim in the development of this model.

At present, compact models are of growing complexity and the number of model parameters has tended to double about every 10 years [95]. The models presented here deviate from that trend, with the aim of developing a model with fewer parameters. Simple expressions taking into account temperature effects for each of the MOSFET's operation regions are given. This has been done by retaining the basic functional form of fully physical models whilst improving the accuracy of the temperature dependent components. A smooth transition between each region is provided and the model allows for easy extraction of parameters. This approach of simplifying the expressions is novel for present day modelling trends, but I believe usability is an important characteristic if the model is to be successful in the hands of the end user. One such end user is the project sponsor company, Goodrich Aerospace. Regular discussions with them allowed me to present this approach to them. The provision of models that they could understand and modify to their needs was seen as particular importance and met with positive feedback.

3.2 Devices and Measurement

Throughout this work conventional n-channel and p-channel MOSFETs were used. Fabricated using a 0.18μm (1.8V) process by Taiwan Semiconductor Manufacturing Company (TSMC™) with a channel width of 10μm and gate oxide thickness of 4.2nm. Figure 3.1 shows an image of one such n-channel device.

Figure 3.1: An image of 0.18μm n-channel device

For testing purposes, test structures each containing five n-channel and five p-channel MOSFET's were separately bonded into dual in-line packages. Measurements were carried out with the use of Keithley™ 2400 and 6487 source/measure units[1]. This provided a link directly from the devices under test to a PC via GPIB[2]. Below is a schematic diagram of the experimental setup.

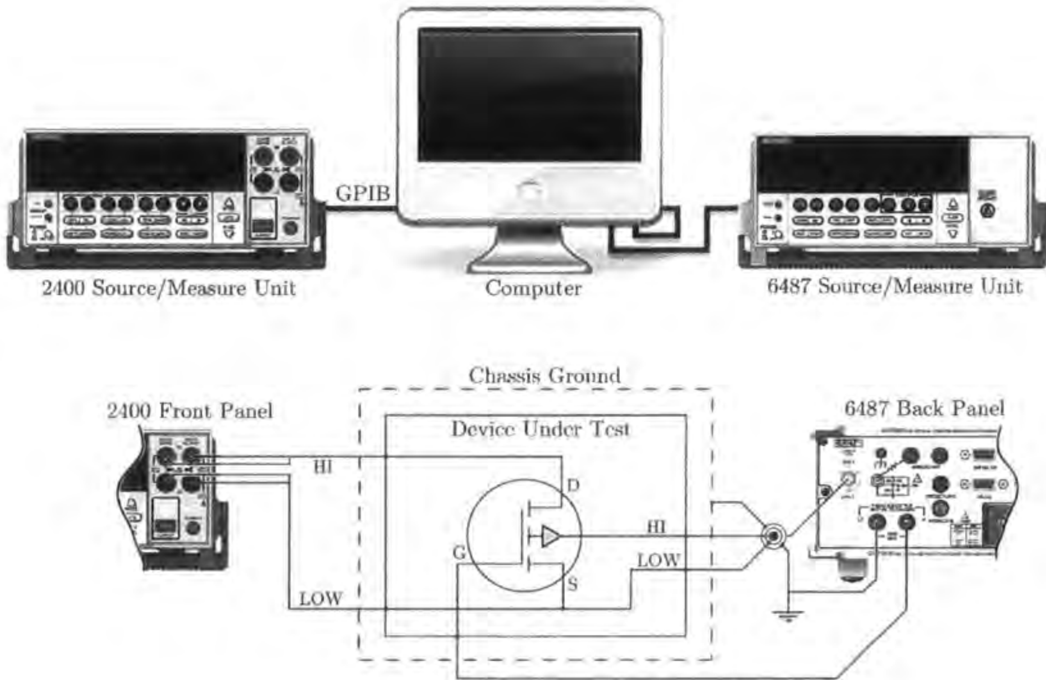


Figure 3.2: Schematic diagram of experimental setup

Using Labview™ to control the units, a set of custom test patterns were created to collate the large quantities of experimental data and provide for varying test conditions. A variable heating element (not shown) provided a temperature range of 27-155°C. Further information and a detailed diagram of the experimental test rig is contained within Appendix A.

[1]Measurement resolution ±10pA
[2]General Purpose Interface Bus

3.3 I-V Model

Current-voltage (I-V) characteristics are vital for any designer in his or her circuit design. This work concentrates on the influence of temperature on the I-V characteristic. Current-voltage equations are derived as a function of threshold voltage, and each operating region is considered separately.

3.3.1 Threshold Voltage

The threshold voltage V_{th} represents the gate bias V_{GS} at which a conducting channel is formed between the source and the drain junctions, and the device in deemed 'ON'. If V_{GS} is below V_{th}, then the device is in the weak inversion (subthreshold) region. For V_{GS} above V_{th}, the device is in the strong inversion region. V_{th} is the gate voltage which results in a surface potential ϕ_s at the SiO_2 interface that is equal to twice the potential difference between the bulk Fermi level E_F, and the intrinsic Fermi level E_i. Simplistically, V_{th} as a function of substrate bias V_{BS} is expressed as:

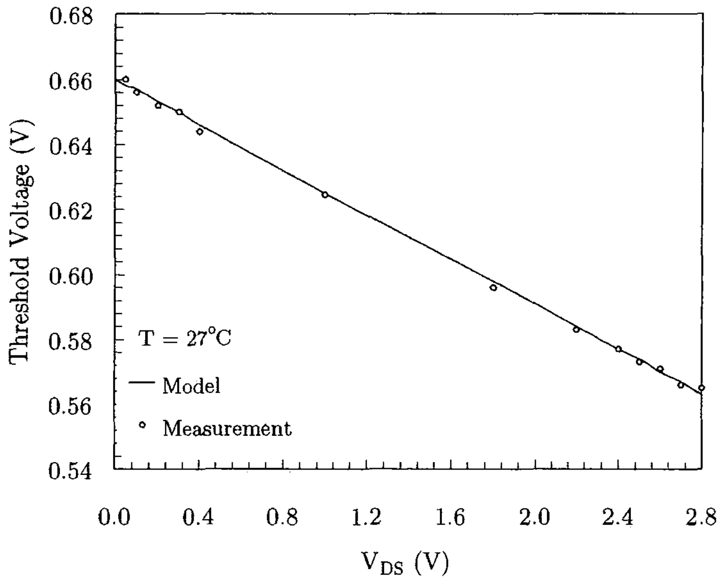
$$V_{th} = V_{th0} - (\gamma \cdot V_{BS}) \tag{3.1}$$

where V_{th0} is the ideal threshold value and γ the substrate bias effect coefficient. Equation 3.1 is only valid for large devices and pays no consideration to the short channel effects seen in today's technologies. For short channel devices, V_{th} is affected by the drain voltage V_{DS}. This is called Drain Induced Barrier Lowering (DIBL). Equation 3.1 can be modified to account for this affect as below, where σ is the DIBL factor.

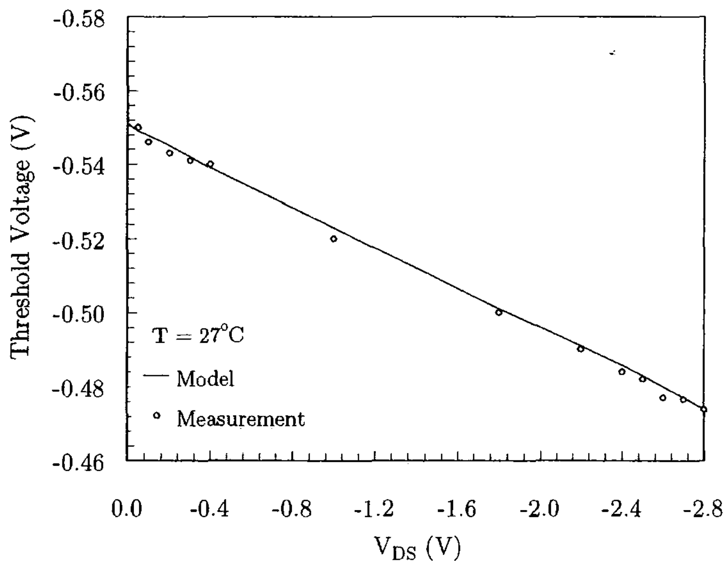
$$V_{th} = V_{th0} - (\sigma \cdot V_{DS}) - (\gamma \cdot V_{BS}) \tag{3.2}$$

As V_{DS} is increased, the potential barrier from the source to channel junction is lowered allowing electrons to be injected into the channel region more freely.

Figure 3.3 shows good agreement between experimental data and equation 3.1 for device types. It indicates a linear relationship between $\triangle V_{th}$ and V_{DS} where the gradient of the lines is σ. Solid lines represent the model calculation and symbols represent the measurement data. This way of illustrating model/measurement data is continued throughout the thesis unless indicated otherwise in the figures.



(a) n-channel



(b) p-channel

Figure 3.3: Threshold voltage as a function of drain voltage

Threshold voltage values were calculated using the constant current method as documented in the Joint Electron Device Engineering Council (JEDEC) Standard No.90 [96]. JEDEC states that the threshold voltage is the gate voltage at which the drain current I_{DS} is equal to a constant current, multiplied by the ratio of gate width to gate length. This is expressed as:

$$V_{th} = V_{GS}\left(@I_{DS} = I_{DS0} \cdot \frac{W}{L}\right) \tag{3.3}$$

where I_{DS0} is typically 0.1μA or -0.025μA for n-channel or p-channel respectively. This value can change but must be selected such that V_{th} is in the subthreshold region of the device. Throughout this thesis I_{DS0} values of 0.375μA and -0.06μA were used for all the threshold voltage calculations.

3.3.2 Linear Region

The linear region of a devices operating characteristic describes the current-voltage behaviour of a MOSFET when a small drain bias is applied. The region is also sometimes called the ohmic region, because the MOSFET acts as a resistor whose resistance can be modulated by changing the gate bias. When the drain voltage is zero, the drain current is zero regardless of the gate-to-source voltage. However, as the drain voltage is increased, a channel is created between the drain and the source that allows current to flow. The amount of current that flows in the channel is defined as the total charge in the inversion layer, divided by the time the carriers need to flow from the source to the drain.

$$I_{DS} = \frac{Q_{inv}}{t_r} \tag{3.4}$$

where Q_{inv} is the inversion layer charge per unit area and t_r is the transit time.

Assuming the velocity, v, of the carriers between the drain and source is constant, the transit time is equal to the channel length divided by the carrier velocity.

$$t_r = \frac{L}{v} \tag{3.5}$$

The carrier velocity is the product of the carrier mobility, μ, and channel electric field, E, where $E = V_{DS}/L$. Hence equation 3.5 becomes:

$$t_r = \frac{L^2}{\mu \cdot V_{DS}} \tag{3.6}$$

The inversion layer charge Q_{inv} can be approximated by the product of the gate oxide capacitance per unit area C_{ox}, and the effective gate voltage $(V_{GS} - V_{th})$.

$$Q_{inv} = C_{ox} \cdot (V_{GS} - V_{th}) \tag{3.7}$$

Referring to figure 3.4, the gate oxide capacitance is given by $C_{ox} = \varepsilon_{ox} A/t_{ox}$, where ε_{ox} is the permittivity of the oxide material, A the area of the gate over the channel and t_{ox} the oxide thickness.

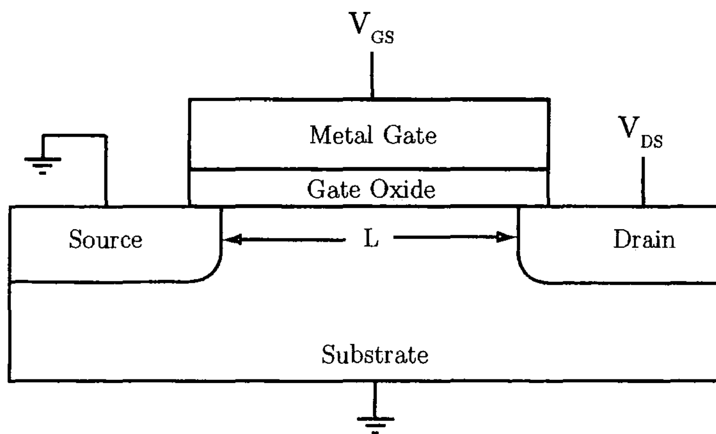


Figure 3.4: MOSFET gate construction

Expanding the gate oxide capacitance parameter in equation 3.7, the inversion layer charge can be expressed as:

$$Q_{inv} = \frac{L \cdot W \cdot \varepsilon_{ox}}{t_{ox}} \cdot (V_{GS} - V_{th}) \tag{3.8}$$

Hence in the form of equation 3.4 the current flowing in the channel of a device is the combination of equation 3.8 with that of 3.6. This leads to the following expression for the linear drain current:

$$I_{DS} = \mu \cdot \frac{W}{L} \cdot C_{OX} \cdot (V_{GS} - V_{th}) \cdot V_{DS} \tag{3.9}$$

This model is valid if it is assumed the carrier velocity, electric field and inversion layer charge are uniform between the drain and source. Using equation 3.9 the drain current flowing in a device will rise linearly with applied drain voltage. In reality though, the drain current reaches a peak and saturates. By considering a variable inversion layer charge this effect can be easily modelled. It is based on the fact that the current at each point in the channel is constant, and is related to a local channel voltage. Considering a small section within the device with length dy and a channel voltage V_C, equation 3.9 becomes:

$$I_{DS} = \mu \cdot \frac{W}{dy} \cdot C_{OX} \cdot (V_{GS} - V_{th} - V_C)\ dV_C \tag{3.10}$$

The drain voltage is replaced with the change in channel voltage over distance dy, namely dV_C. To model the carrier velocity more accurately in the channel, we split the channel length into segments of length dy and consider the change of channel voltage across each segment dV_C. Thus:

$$E_{(y)} = \frac{dV_{C(y)}}{dy} \tag{3.11}$$

Given that the carrier velocity, $v = \mu E$, where the carrier mobility μ is constant, a semi-empirical model can be used to describe the relationship between the channel field and carrier velocity, expressed as:

$$v_{(y)} = \frac{\mu \cdot E_{(y)}}{1 + E_{(y)}/E_{sat}} \tag{3.12}$$

where E_{sat} is the critical field at which the carrier velocity becomes saturated given by $E_{sat} = 2v_{sat}/\mu$. Values of v_{sat} are typically around 7 x 10^6cm/s. Equation 3.10 can then be modified to model a variable electric field and carrier drift velocity along the channel. This assumes the carrier mobility μ is constant.

$$I_{DS} = v_{(y)} \cdot C_{OX} \cdot W \cdot (V_{GS} - V_{th} - V_{C(y)})$$

$$I_{DS} = \frac{\mu \cdot E_{(y)}}{1 + E_{(y)}/E_{sat}} \cdot C_{OX} \cdot W \cdot (V_{GST} - V_{C(y)}) \tag{3.13}$$

V_{GST} is the effective gate bias given by the applied gate voltage minus the threshold voltage $V_{GST} = V_{GS} - V_{th}$. Rearranging equation 3.13 for the electric field we get:

$$E_{(y)} = \frac{I_{DS}}{\mu \cdot W \cdot C_{ox} \cdot (V_{GST} - V_{C(y)}) - I_{DS}/E_{sat}} \tag{3.14}$$

Integrating both sides from the source to the drain, so that y varies between zero and the gate length, and the channel voltage V_C varies between zero and the drain voltage we get:

$$I_{DS} = \beta \cdot \frac{1}{1 + V_{DS}/E_{sat}L} \cdot \left[(V_{GS} - V_{th}) - \frac{V_{DS}}{2} \right] \cdot V_{DS} \tag{3.15}$$

where β is the transconductance parameter.

$$\beta = \mu \cdot C_{ox} \cdot \frac{W}{L} \tag{3.16}$$

Using this quadratic function the drain current increases with applied drain bias, and then reaches a maximum value at V_{DSsat}. According to equation 3.15 the current would then decrease and eventually become negative. In reality though, the current voltage characteristic continues to rise with a weak dependence on the drain voltage. This section is called the saturation region.

3.3.3 Saturation Region

When the drain voltage reaches the saturation point, the channel pinches-off and the inversion layer no longer reaches the drain junction. On the onset of channel pinch off, the lateral electric field in the channel is equal to E_{sat} and the carrier velocity equal to v_{sat}. The saturation voltage V_{DSsat} is the corresponding channel voltage at this point. Therefore at saturation we can substitute the velocity v_{sat} for $v_{(y)}$ and channel voltage V_{DSsat} for $V_{C(y)}$ in equation 3.13. The resultant drain current at saturation becomes:

$$I_{DS} = v_{sat} \cdot C_{ox} \cdot W \cdot (V_{GST} - V_{DSsat}) \tag{3.17}$$

The saturation voltage can be derived by letting $v_{(y)} = v_{sat}$ and $V_{DS} = V_{DSsat}$ in equation 3.15 and setting it equal to equation 3.17. The resultant saturation voltage equation presented in [97] is expressed as:

$$V_{DSsat} = \frac{E_{sat} \cdot L_{eff} \cdot V_{GST}}{(E_{sat} \cdot L_{eff}) + V_{GST}} \tag{3.18}$$

This model is consistent for $V_{DS} = V_{DSsat}$, taking into account velocity saturation effects. We must also consider the condition when $V_{DS} > V_{DSsat}$. As the drain voltage is increased beyond the saturation voltage, the pinch-off point moves further away from the drain. This effectively reduces the channel length and adds more charge carriers to the channel region. For short channel devices this difference is

not negligible and the drain current in short channel devices increases for increasing drain bias after saturation. This is called Channel Length Modulation (CLM). In the development of this model a semi empirical expression is used to consider these effects so that equation 3.17 now becomes:

$$I_{DS} = v_{sat} \cdot C_{ox} \cdot W \cdot (V_{GST} - V_{DSsat}) \cdot (1 + \lambda V_{DS}) \tag{3.19}$$

where λ is the channel length modulation factor, typically < 0.1. To ensure continuity between the linear and saturation regions, the channel length modulation factor is also included in equation 3.15 for the linear region. Thus, the model for the drain current in the strong inversion regime of a MOSFET becomes:

$$I_{DS_0} = \begin{cases} \beta \cdot \frac{1}{1+V_{DS}/E_{sat}L} \cdot \left(V_{GST} - \frac{V_{DS}^2}{2}\right) \cdot (1 + \lambda V_{DS}) & \text{if } V_{DS} < V_{DSsat}, \\ v_{sat} \cdot C_{ox} \cdot W \cdot (V_{GST} - V_{DSsat}) \cdot (1 + \lambda V_{DS}) & \text{if } V_{DS} > V_{DSsat} \end{cases} \tag{3.20}$$

3.3.4 Mobility Model

Up to now it has been assumed that μ, the carrier mobility, is constant. It has been shown [98] though that μ has a dependency on V_{GS} that must be taken into account if we are to compare our I-V model to real data. A widely used expression for mobility field dependency is:

$$\mu = \frac{\mu_0}{1 + \mu_v(V_{GST})} \tag{3.21}$$

where μ_0 is the low field mobility and μ_v is the gate bias factor. I have found this model does not match my experimental results and therefore propose a different model. I have created a simple empirical model given by:

$$\mu = \mu_0 - \mu_v(V_{GST}) \tag{3.22}$$

Figure 3.5 shows a comparison between measured data and theoretical mobility values calculated using equation 3.22. It also highlights the inaccuracy of equation 3.21 for modelling the mobility field dependency in deep sub-micron devices. All measurements and calculations were determined at room temperature.

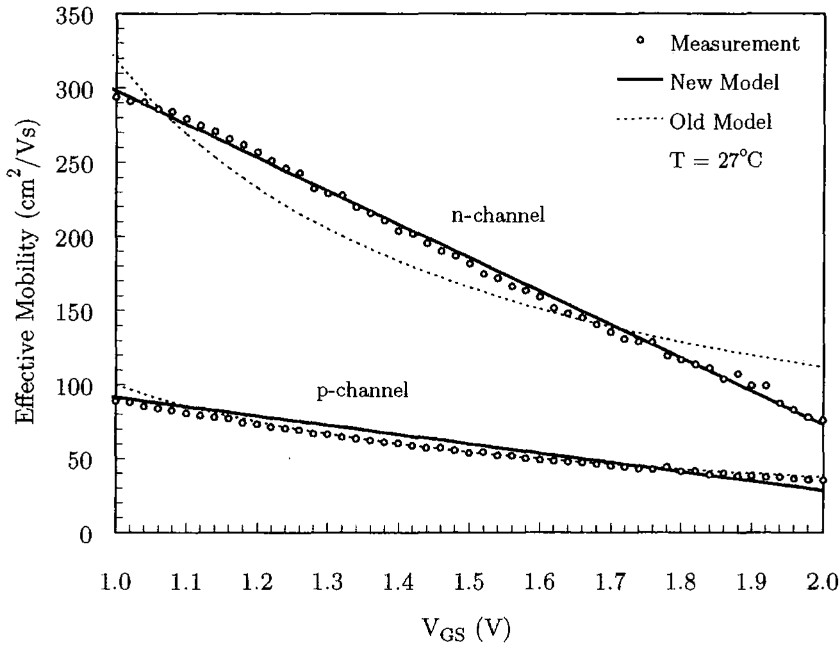


Figure 3.5: Mobility as a function of gate voltage

Good agreement can be seen between the measured and the new model data. A simple extraction of μ_0 can be determined experimentally from the point where the device transconductance (Gm) is at a maximum.

$$\mu_0 = \frac{Gm_{(max)} \cdot L}{C_{ox} \cdot W \cdot V_{DS}} \tag{3.23}$$

This yields μ_0 values of 523cm^2/Vs and 154cm^2/Vs for n-channel and p-channel devices respectively. This mobility model can be used for both electrons and holes but with different extracted values for the parameters μ_0 and μ_v. All instances of μ in the previous calculations are replaced with that of equation 3.22.

3.3.5 Parasitic Drain/Source Resistance

An additional short channel effect that has been incorporated into the model is the influence of parasitic resistance (R_{DS}). As a MOSFET's channel length is scaled down, the impact of R_{DS} on I_{DS} becomes more influential, especially at low drain bias. To model R_{DS} in a direct method would lead to a significantly more complicated drain current expression. The development of this model is based on R_{DS} comprising of R_D, the drain resistance and R_S the source resistance:

$$R_{DS} = R_D + R_S \tag{3.24}$$

For efficiency and simplicity it is assumed that $R_D = R_S$ even though the author acknowledges that asymmetric Lightly Doped Drain (LDD) structures result in a non-equal R_D and R_S. This is implemented into the model as below:

$$I_{DS} = \frac{I_{DS_0}}{1 + \frac{R_{DS} \cdot I_{DS_0}}{V_{DS}}} \tag{3.25}$$

where I_{DS_0} is the previously derived drain current expression (equation 3.20).

3.3.6 Subthreshold Region

In section 3.3.1 it was stated that current flows in the channel of a MOSFET when the surface potential is equal to, or greater than $2\phi_F$, and $V_{GS} > V_{th}$. This current is called drift current. However when $V_{GS} < V_{th}$, there also exists a concentration of electrons near the surface, which create leakage current between the drain and source even when the surface is not in strong inversion. This current is called diffusion current, and is what characterises the subthreshold region. For deep submicron devices the modelling of the subthreshold current is vital because as devices are scaled the relative magnitude of the OFF-state current becomes more significant.

This problem of increased OFF-state current is illustrated later in chapter 5. However in this chapter we adopt the following simple expression for the subthreshold current. This is an established model [99] that characterises the drain current as rising exponentially with the gate voltage.

$$I_S = I_{S_0} \cdot \exp\left(\frac{V_{GST} - v_{off}}{nV_{tm}}\right) \cdot \left[1 - \exp\left(\frac{-V_{DS}}{V_{tm}}\right)\right] \tag{3.26}$$

where the parameter v_{off} is the offset voltage, n the subthreshold swing factor, V_{tm} the thermal voltage, and I_{S_0} is the current at $V_{GS} = 0$ given by:

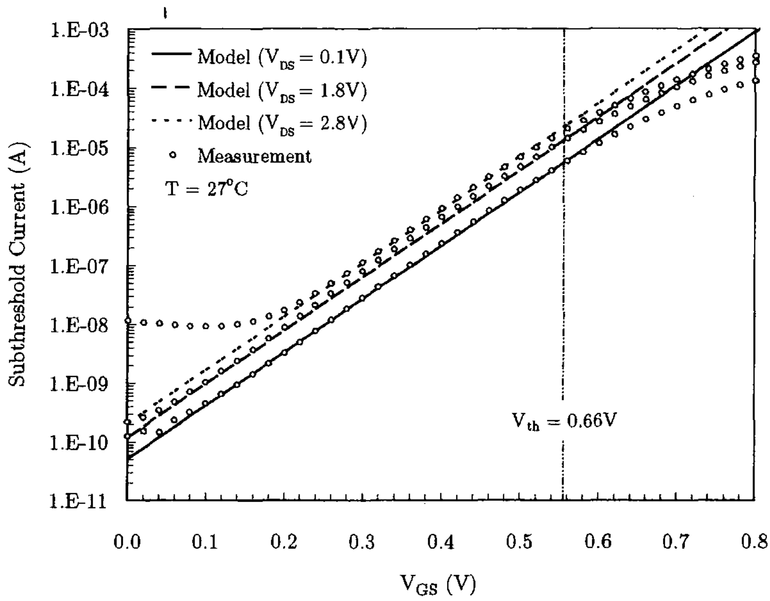
$$I_{S_0} = \mu \cdot \frac{W}{L} \cdot C_{ox} \cdot {V_{tm}}^2 \tag{3.27}$$

The thermal voltage is expressed as:

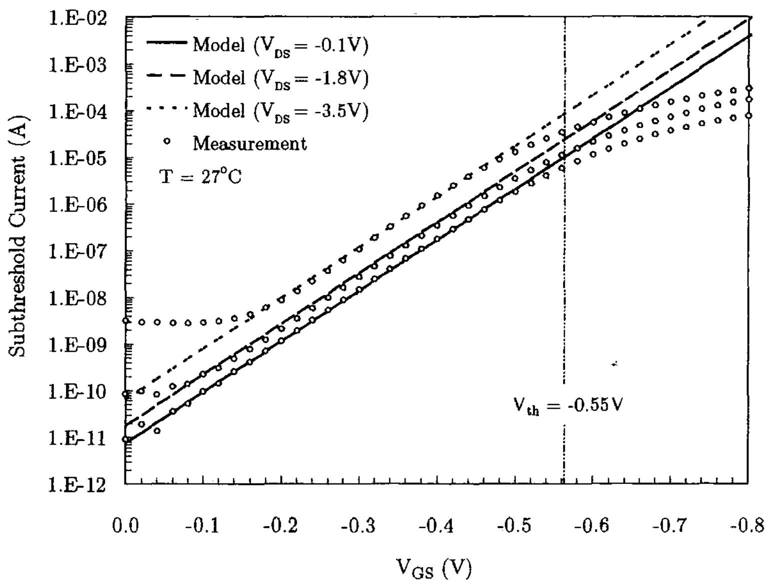
$$V_{tm} = k_b T / q \tag{3.28}$$

where q is electron charge[3]. All other terms have been previously quantified. Figure 3.6 illustrates a comparison between measured and simulated data using equation 3.26. Good agreement can be seen in the middle region of the curves. For long channel devices, the subthreshold current is independent of the drain bias. In short channel devices, I_S can be significantly larger for higher drain bias due to DIBL effects mentioned in section 3.3.1. The model allows for these DIBL effects. Parameter extraction for the subthreshold model is simple with only v_{off} and n to determine. For the results presented in figure 3.6(a) v_{off} and n were -0.062 and 1.85, and for figure 3.6(b), -0.030 and 1.54 respectively. There are clear inaccuracies though at higher drain voltages and also above V_{th}. The former is due to a tunnelling process called GIDL that was reviewed in chapter 2. It occurs at the gate-to-drain overlap region when there is a high lateral electric field close to the drain region. As a result a new OFF-state leakage component is seen. Further examination and modelling of

[3]Electron charge = 1.602177×10^{-19}C (Coulombs)



(a) n-channel



(b) p-channel

Figure 3.6: Subthreshold current as a function of gate voltage

GIDL is given in chapter 5 and the present subthreshold model is modified to characterise its effects. The later inaccuracy in figure 3.6, seen above V_{th} is one of the common difficulties in modelling a MOSFET I-V characteristic. As V_{GS} approaches V_{th}, the device is moving from weak to strong inversion. This means for a complete I-V model, we need a smooth transition between these two regions.

3.3.7 Transition Region

There is no simple physical model available for the drain current in this region. One approach is to define a unified expression across all regions. It is the author's belief though that this approach is inefficient for computing the whole I-V model and goes against the aims of this work. Providing separate expressions for each region is not only more computationally efficient but also allows the user greater freedom to modify and analyse individual components. In BSIM2[4] a spline function is used, but determining all the coefficients of the spline is time consuming. I therefore propose the use of a much simpler way to accurately model the transition region.

To provide a continuous I_{DS} characteristic, lower $V_{GS(low)}$ and upper $V_{GS(high)}$ transition points must be defined. A mid-point drain current $I_{DS(mp)}$ must also be determined. Then with the use of a weighting factor, a smooth curve between $V_{GS(low)}$ and $V_{GS(high)}$ through $I_{DS(mp)}$ can be provided by the following equation [99]:

$$I_{DS_{(Tran)}} = (1 - w^2) I_{S(low)} + \left[2\,(1 - w)\,(w I_{DS(mp)})\right] + (w^2 I_{DS(high)}) \tag{3.29}$$

where $I_{S(low)}$ and $I_{DS(high)}$ are the corresponding subthreshold and drain current values at $V_{GS(low)}$ and $V_{GS(high)}$, and w is the weighting factor that varies between zero and one. The drain current mid-point ($I_{DS(mp)}$) was found to be dependent on drain bias and I have defined an empirical expression to match the measured data.

$$w = \left(V_{GS} - V_{GS(low)}\right) \cdot \left(\frac{1}{V_{GS(high)} - V_{GS(low)}}\right) \tag{3.30}$$

$$I_{DS(mp)} = M_a + \left(V_{DS} \cdot M_{b)}\right) \tag{3.31}$$

where M_a and M_b are fitting parameters.

[4]Berkeley Short-channel IGFET Model

Figure 3.7 shows how effective this transition function is at avoiding any discontinuities at the boundaries between regions.

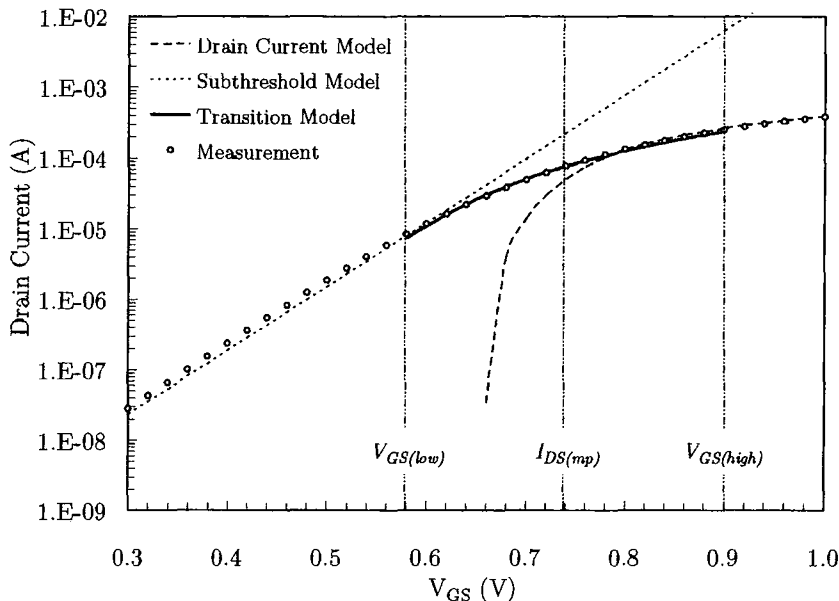


Figure 3.7: Illustration of the transition region

3.3.8 I-V Model Test Results

A set of benchmark tests were performed to check the model's general applicability, robustness (lack of discontinuities), and accuracy. Results for deep sub-micron n-channel and p-channel devices at room temperature are shown. The core operating voltage V_{DD} of the devices in the measurements is 1.8V.

At the introduction of this chapter, I promoted an approach to modelling that encouraged simplicity and the use of minimal model parameters. Table 3.1 shows the parameters extracted and used for model calculations in all results presented in this section. A total of 8 I_{DS} measurements were taken for each device type and employed to extract the 9 model parameters needed. The threshold voltage, DIBL

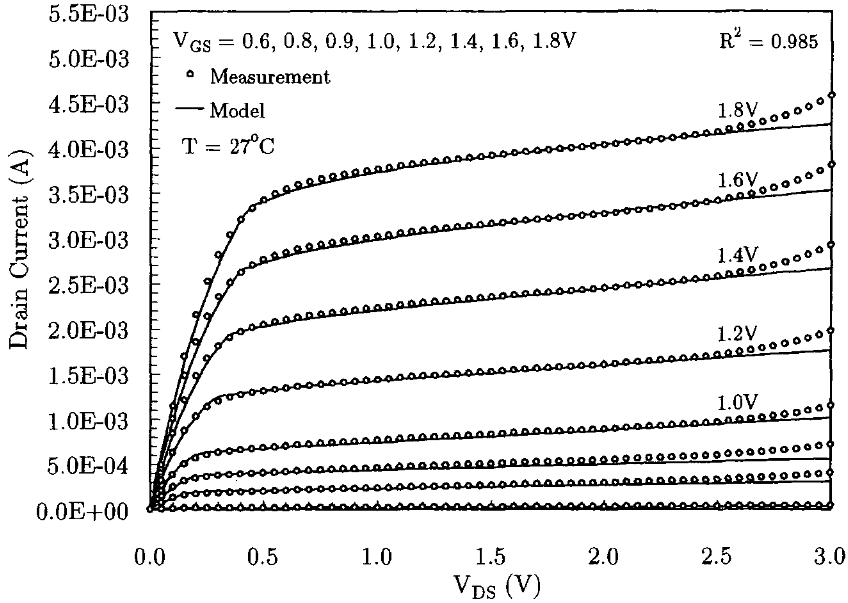
Model Parameters		n-channel	p-channel
Gate Length	L (μm)	0.18	0.18
Gate Width	W (μm)	10	10
Oxide Thickness	t_{ox} (μm)	4.2	4.2
Threshold Voltage	V_{th0} (V)	0.66	-0.55
V_{th} DIBL Factor	σ	0.035	0.02
Low Field Mobility	μ_0 (cm^2/Vs)	372	111
Mobility Field Factor	μ_v	35	10
Velocity Saturation	v_{sat} (cm/s)	6.8x10^6	6.6x10^6
CLM Factor	λ	0.01	0.03
Parasitic Resistance	R_{DS} (Ω)	22	65
Subthreshold Offset	V_{off} (V)	-0.062	-0.028
Subthreshold Swing	n	1.85	1.54

Table 3.1: Extracted model parameters

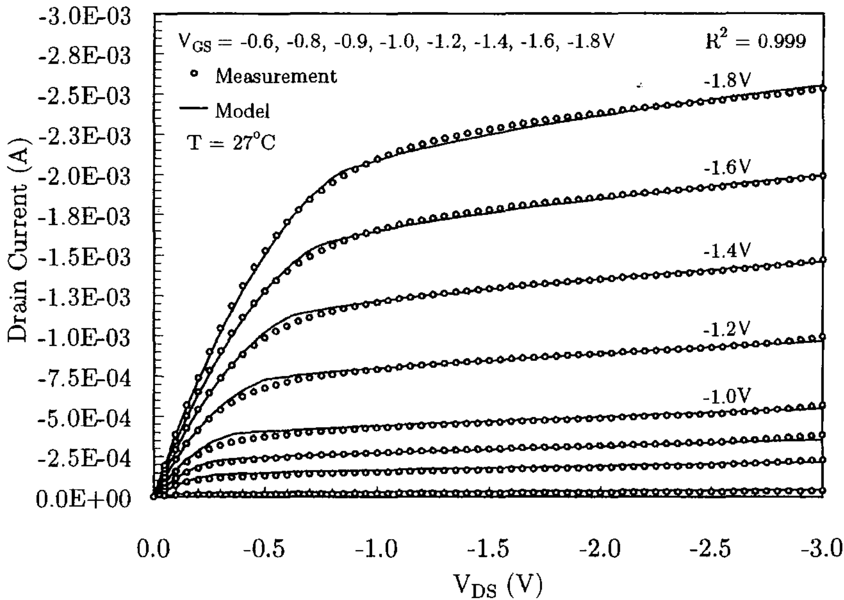
factor, mobility and subthreshold parameters can be extracted from the linear region characteristic V_{GS} - I_{DS}. The channel length modulation (CLM) parameter is extracted by fitting the model to I_{DS} - V_{DS} at V_{GS} = 1.8V (V_{BS} = 0). The field dependence of the mobility was also fitted to the I_{DS} - V_{DS} characteristic.

With only a small number of parameters needing to be found, the models were fitted manually for best fit. However the author acknowledges that in practice where the characterisation of a large number of devices may be needed, it would be more efficient to use an optimisation technique such as least squares. Least Squares is a mathematical optimisation method which attempts to find a "best fit" function that has the minimal sum of the deviations squared (least square error) from a given set of measured data. An example of using such a technique for the calculation of the subthreshold parameters n_0 and V_{off} is described in Appendix B. The resultant optimised parameters for a n-channel device were n_0 = 1.93 and V_{off} = 0.065, which are close to those manually extracted as shown in table 3.1. An alternative approach is to use a program such as MATLAB™ and its optimisation toolbox. This type of program is of great help and can be used to automate tasks such as the least squares optimisation shown in Appendix B.

Using the parameters defined in table 3.1, figure 3.8 shows the simulated $I_{DS} - V_{DS}$ characteristic and measured data. The gate voltage is varied between V_{th} and V_{DD}. Excellent agreement between model and measured data is seen with R^2 values of 0.985 and 0.999 for n-channel and p-channel devices respectively. The R-squared



(a) n-channel



(b) p-channel

Figure 3.8: V_{DS} - I_{DS} characteristic

value is the proportion of the variance in the y attributable to the variance in x, and hence a quantifiable value of fit between model and measurement. A smooth transition between the linear region and saturation region is illustrated. At high drain bias ($V_{DS} > 2.5$V) the n-channel model is seen to deviate from the experimental data. When the electrical field near the drain is very large (>0.1MV/cm), some electrons coming from the source will be energetic enough to cause an additional current component. This is the onset of impact ionisation and the creation of electron-hole pairs adding a substrate leakage to the drain current . Further analysis of this is given in chapter 6, but at this stage it gives us a nice indication as to why hot-carrier degradation is greater in n-channel devices than p-channel. Figure 3.8 also illustrates how the point at which velocity saturation occurs ($V_{DS(sat)}$) increases with higher gate bias. N-channel devices reach $V_{DS(sat)}$ earlier due to higher carrier mobility. Figure 3.9 compares the saturation voltages as a function of gate bias for both n-channel and p-channel devices.

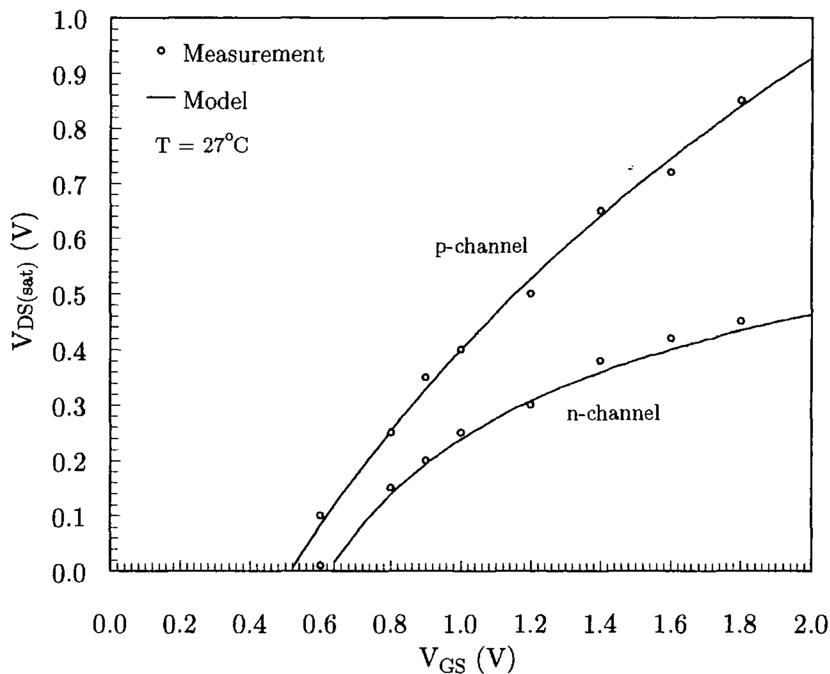
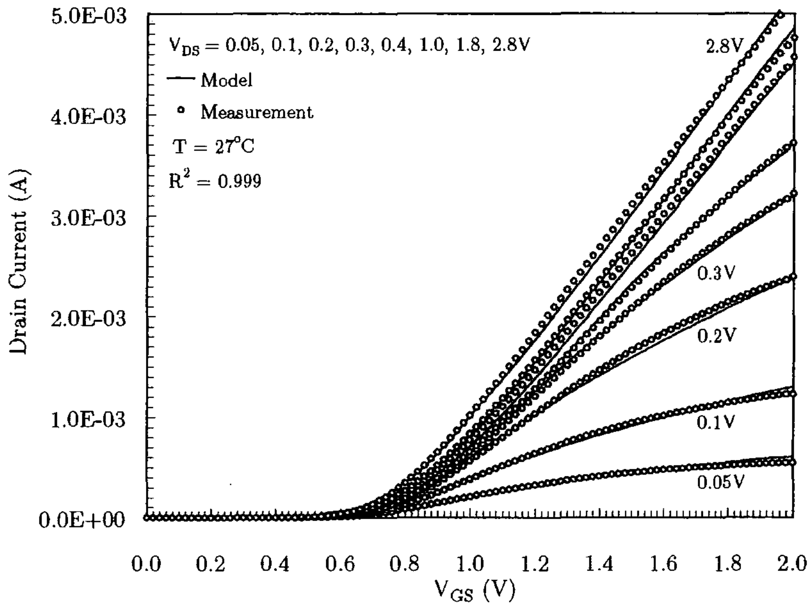
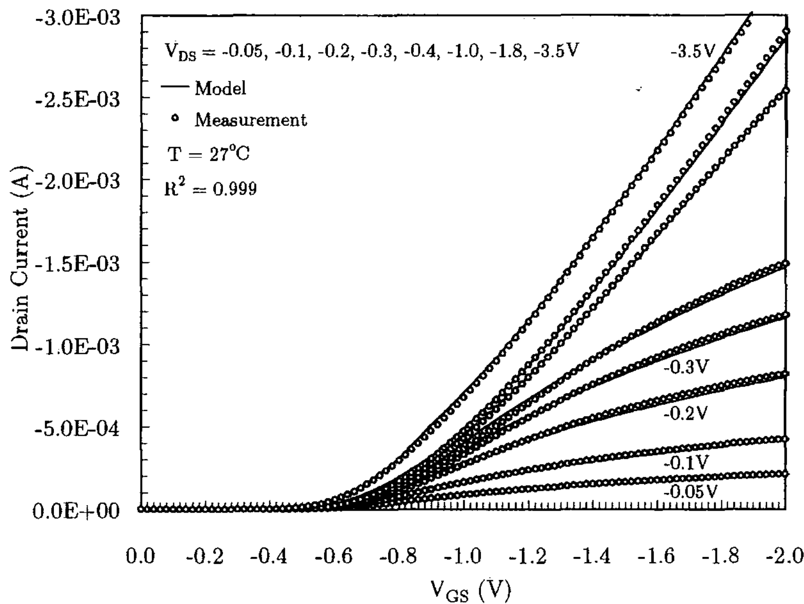


Figure 3.9: Saturation voltage as a function of gate voltage

The next piece of analysis looks at the $I_{DS} - V_{GS}$ characteristic. For constant drain bias, V_{GS} is swept from zero to two volts taking measurements every 20mV. In figure 3.10 the drain current increases with gate bias when $V_{GS} > V_{th}$. The relative drain current values are also higher, as one might expect for higher drain bias.



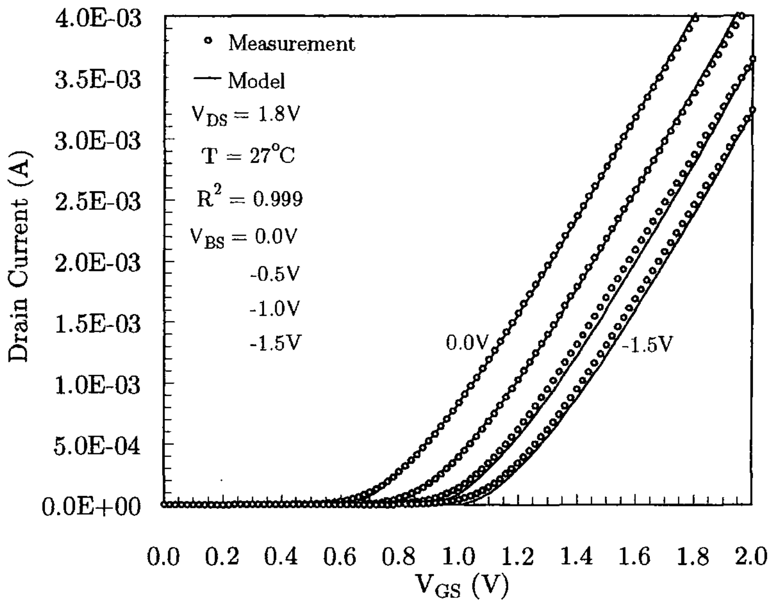
(a) n-channel



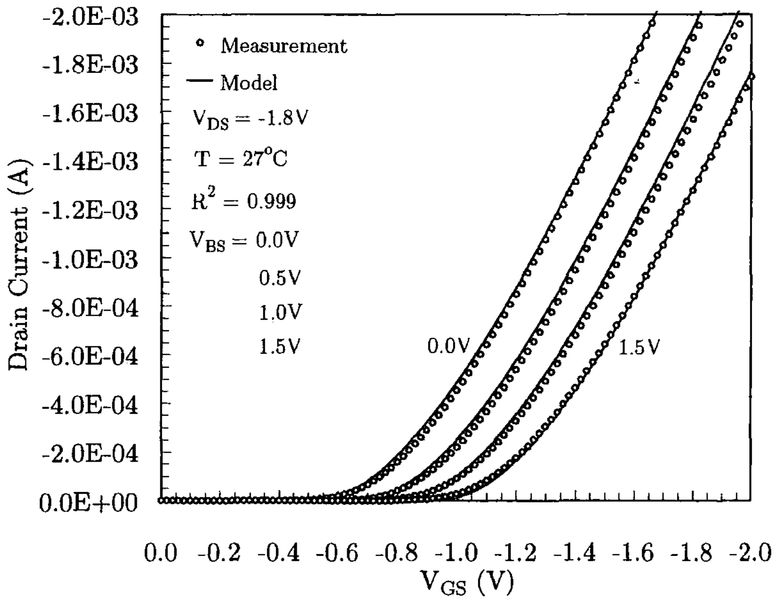
(b) p-channel

Figure 3.10: V_{GS} - I_{DS} characteristic

A reduction in threshold voltage at high drain bias due to DIBL effects, introduced by equation 3.1, can be seen for both n and p-channel devices. Additionally the importance of considering R_{DS} is seen. The effect of the parasitic resistance is more evident at low drain bias, causing the current to curve-off at high gate voltages.



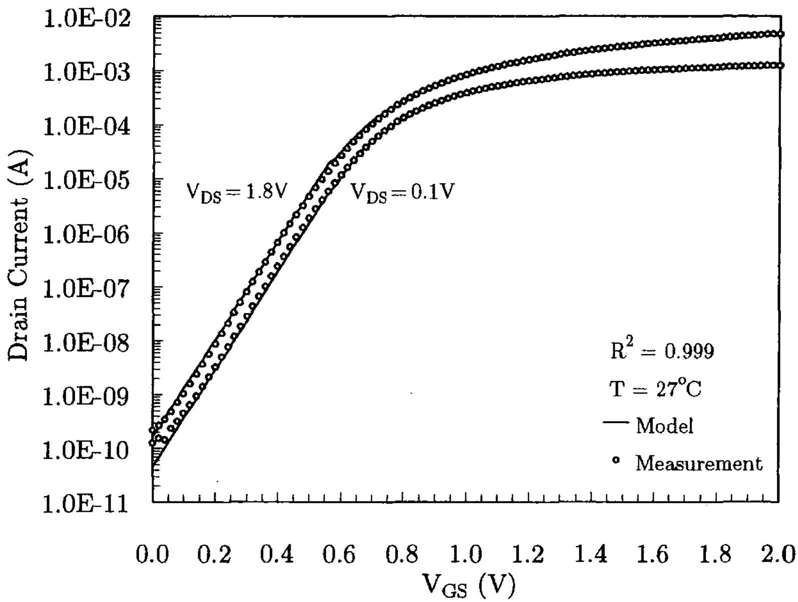
(a) n-channel



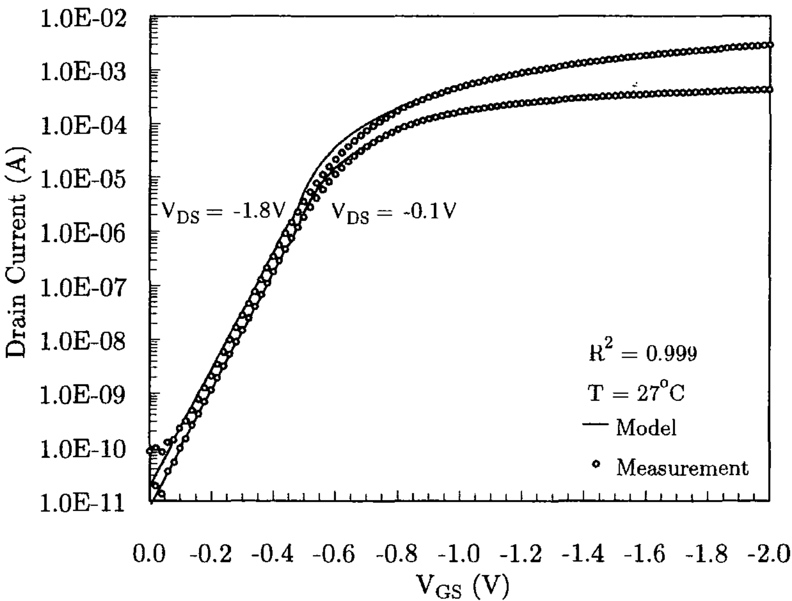
(b) p-channel

Figure 3.11: V_{GS} - I_{DS} characteristic as a function of substrate bias

This verifies the incorporation of the parasitic resistance in the drain current expression (equation 3.25), adding to model accuracy. When a substrate voltage is applied to a MOSFET, it effectively increases the voltage required to turn the device on. This change was introduced with the use of γ, the substrate bias effect coefficient



(a) n-channel



(b) p-channel

Figure 3.12: Logarithmic V_{GS} - I_{DS} characteristic

in equation 3.2. Measurements were taken at four different substrate voltage, and the results compared in figure 3.11. Using γ values of 0.26 and 0.3 for n-channel and p-channel devices respectively, the model results show excellent agreement with experimental data. Figure 3.12 is the final result of this section showing the I-V characteristic for the entire MOSFET operation, from subthreshold to the strong inversion region. Excellent agreement between the model and measured data is seen. In particular the transition between the two regions is very smooth and verifies the use of the transition expression 3.29. Again the underlying aim of this work is met with only two parameters being needed to model the subthreshold region, V_{off} and n, making the modelling process simple and accurate at the same time.

3.4 Temperature Effects

This section describes the models that were developed to account for the effect of temperature on a MOSFET's operation. I have modelled the effect of temperature on five parameters. I believe this is all that is needed to allow accurate simulations and this is justified by the accuracy of the final model as shown later.

3.4.1 Carrier Mobility

Modelling the effects of temperature on carrier mobility is vital for any high temperature model. The following empirical relationship has been used to model the mobility temperature dependence $\mu(T)$.

$$\mu(T) = \mu_{(Tnorm)} - \theta \left(\frac{T}{T_{norm}} - 1 \right)^{\mu_T} \quad (3.32)$$

where $\mu_{(Tnorm)}$ is the mobility at room temperature, μ_T the temperature exponent and θ a fitting parameter.

Figure 3.13 shows the dependence of mobility on temperature, where $\mu_{T'}$ was fitted to be 0.56 and 0.7 for n-channel and p-channel devices respectively.

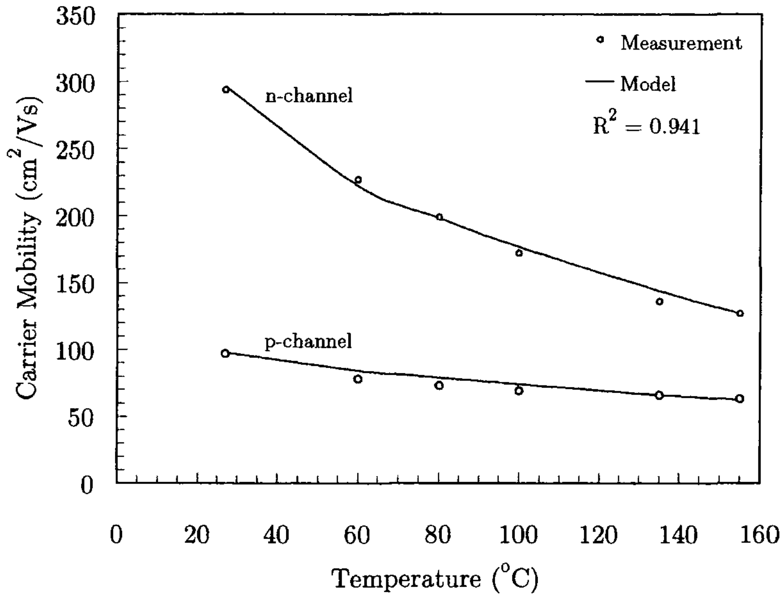


Figure 3.13: Carrier mobility as function of temperature

Mobility values were calculated from the linear $I_{DS} - V_{GS}$ characteristic using the procedure mentioned in section 3.3.4. The decrease seen in carrier mobility at different temperatures is due to scattering effects in the channel. At room temperature the mobility is stable due to there being no scattering of free carriers, and subsequently the carriers do not interchange energy with the stationary lattice structure. As the temperature rises though, vibrations in the lattice cause energy to be transferred between the carriers and lattice. This increase in lattice scattering results in a decrease in carrier mobility.

3.4.2 Velocity Saturation

The temperature dependence of v_{sat} is usually ignored in the literature. However, the author believes the use of a simple model expression enhances the model's accuracy.

As the electric field in the channel increases so does the velocity of the carriers. At high fields the carriers reach a velocity saturation and the relationship with the electric field is no longer linear. This is due to the scattering of highly energetic electrons at high fields, increasing the transit time of carriers through the channel. Hence it effects the voltage at which current saturation occurs and for accurate modelling any v_{sat} temperature effects must be considered. The following empirical expression was found to adequately simulate $v_{sat}(T)$.

$$v_{sat}(T) = v_{sat_0} - \sigma_{vsat}\left(\frac{T}{T_{norm}} - 1\right) \tag{3.33}$$

where v_{sat_0} is the velocity saturation at room temperature and σ_{vsat} the velocity saturation temperature factor. Figure 3.14 illustrates a comparison between the

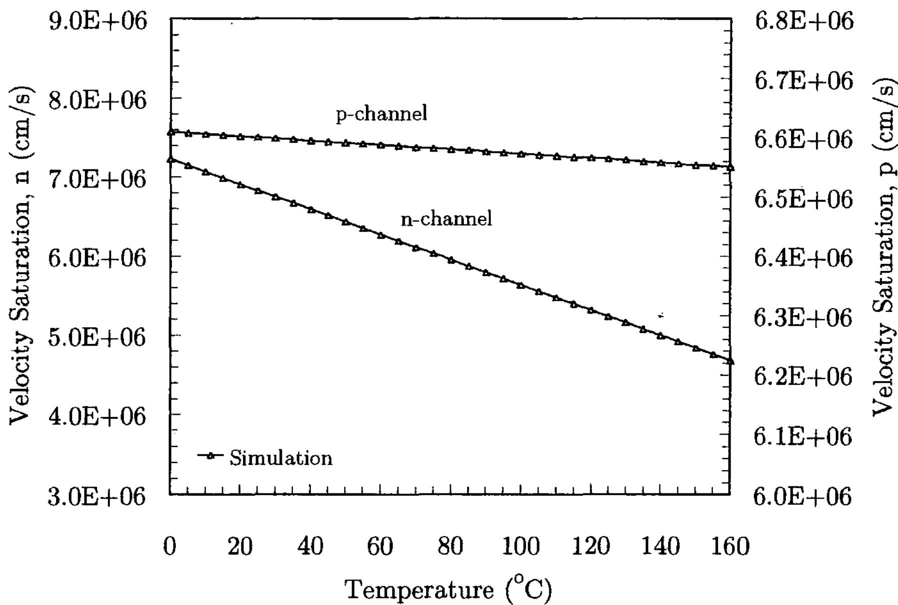


Figure 3.14: Velocity saturation as function of temperature

simulated temperature dependence of v_{sat} for n-channel and p-channel devices. It can be seen that for both types of devices v_{sat} reduces with increased temperature. This is believed to be due to the reduction in carrier mobility at higher temperatures highlighted in section 3.4.1.

3.4.3 Threshold Voltage

Given that V_{th} is dependent on ϕ_s the surface potential, it is this component that is the main cause of V_{th} temperature variations [100]. As temperature increases, so does the intrinsic carrier concentration n_i, whilst the energy band gap is reduced. The combination of these effects brings the Fermi potential ϕ_f towards the mid-gap. One can solve ϕ_f as:

$$\phi_f = k_b T \, ln \left(\frac{N}{n_i} \right) \tag{3.34}$$

where k_b is Boltzmann's constant[5], T temperature, and N the numbers of ionised donors and acceptors in n-channel and p-channel devices respectively. Figure 3.15 illustrates the influence of temperature on the Fermi potential for silicon with donor impurity concentration as the parameter. As temperature increases, ϕ_f approaches zero meaning the extrinsic semiconductor will become intrinsic at sufficiently high

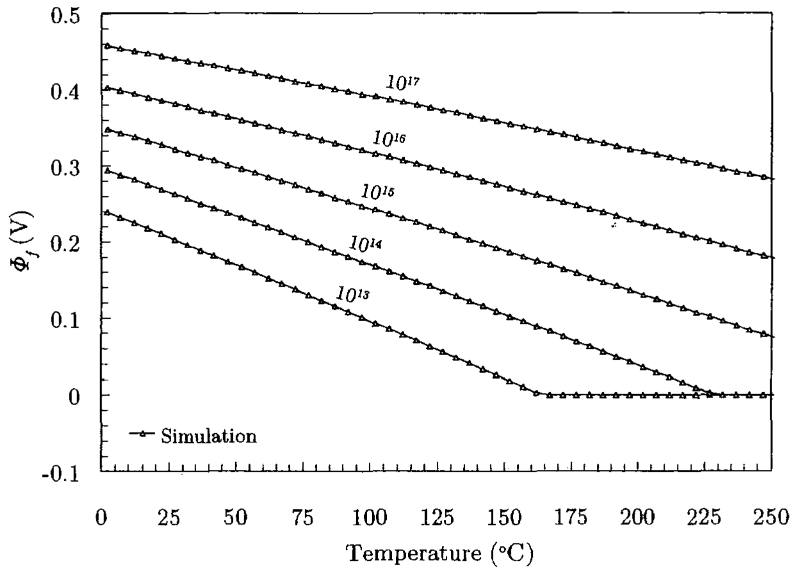


Figure 3.15: Fermi potential as a function of temperature

temperatures. Figure 3.15 also shows how semiconductors with greater doping levels maintain extrinsic behaviour to higher temperatures. It should be noted that the gate oxide charge has been shown to be essentially temperature independent [101].

[5]Relates temperature to energy = 1.38x10^{-23}J K^{-1}

Given that $\phi_s = 2 \cdot \phi_f$, at higher temperatures the surface potential is reduced and V_{th} would be expected to follow a similar trend. Experiments were carried out on both n-channel and p-channel devices at various temperatures and the results are shown in the figure below.

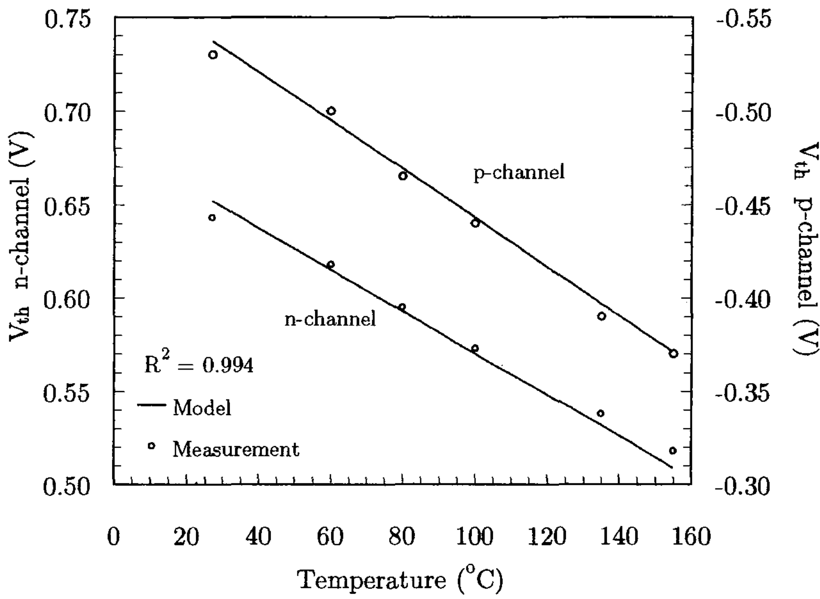


Figure 3.16: Threshold voltage as a function of temperature

Figure 3.16 shows that V_{th} changes approximately linearly with temperature in the range 27-155°C. Threshold voltage values decrease with increased temperature by 1.1mV/°C and 1.3mV/°C for n-channel and p-channel respectively. To model this I have used a simple temperature relationship as show in equation 3.35.

$$V_{th}(T) = V_{th}(T_{norm}) + T_{Vth} \cdot \left(\frac{T}{T_{norm}} - 1 \right) \tag{3.35}$$

where $V_{th}(T_{norm})$ is V_{th} at room temperature, T_{norm} the room temperature, and T_{Vth} the threshold voltage temperature coefficient. The model gives very good agreement with the experimental data over this temperature range as shown in the above figure. Slight variations between n-channel and p-channel devices are taken into account with T_{Vth} equalling -0.030 and -0.035 respectively.

3.4.4 Subthreshold Swing Factor

In the weak inversion regime the diffusion current displays a temperature dependence due to increased carrier concentration. I have used a very simple relationship to model the temperature effects seen in the subthreshold region.

$$n(T) = n_0 + n_T \cdot \left(\frac{T}{T_{norm}} - 1 \right) \tag{3.36}$$

where n_0 is the swing at room temperature and n_T, the temperature factor. Figure 3.17 illustrates how well this simple expression agrees with measured data over a wide range of temperatures. n_0 was extracted as 1.54 and n_T as 0.11. At higher temperatures it can be seen that for p-channel devices there is an increase in drain current. The ability of the model to predict how higher temperatures influence the subthreshold current is also shown. The importance of incorporating $n(T)$ is illustrated with a 155°C model response with a constant n value. Similar results were obtained for n-channel devices.

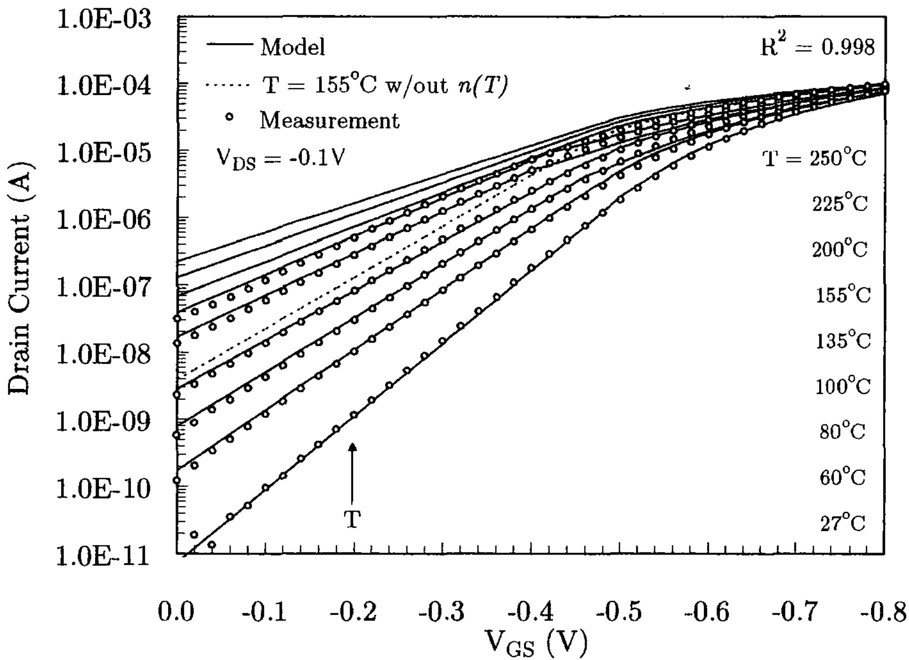


Figure 3.17: Temperature dependence of the subthreshold region (p-channel)

3.4.5 Drain Source Resistance

The fifth and final temperature dependent parameter included is the drain/source resistance R_{DS}. The existence of a drain/source resistance can cause a reduction in the drain current, because the effective voltages at the source and drain junctions are less than those applied to the terminals. Hu[102] proposed that the drain/source resistance increases almost linearly with temperature. I have used the following expression to model $R_{DS}(T)$:

$$R_{DS}(T) = R_{DS(0)} + C_{RDS} \cdot \left(\frac{T}{T_{norm}} - 1 \right) \tag{3.37}$$

where $R_{DS(0)}$ is the parasitic resistance at room temperature, C_{RDS} the temperature constant, and T_{norm} room temperature. Figure 3.18 shows the simulated drain/source resistance values against temperature. It can be seen that the model predicts R_{DS} to rise linearly with temperature, and hence will result in a reduction of drain current at higher temperatures.

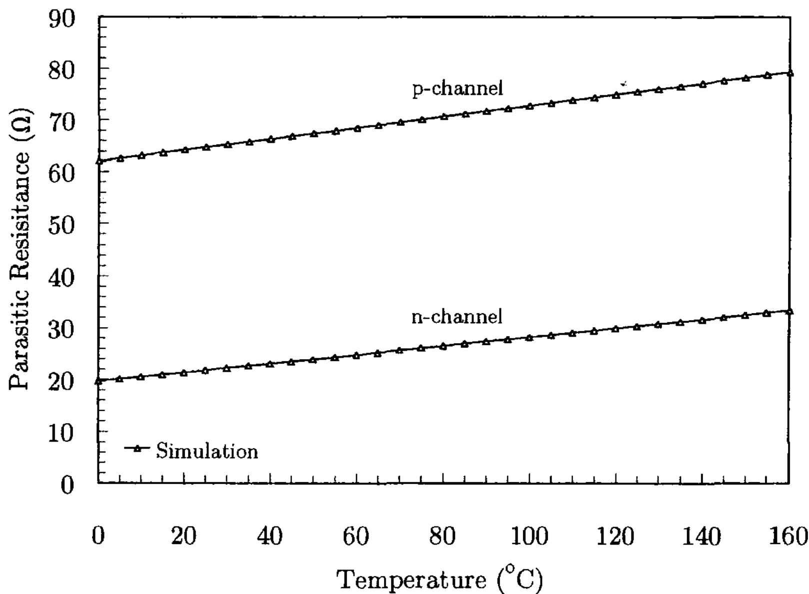
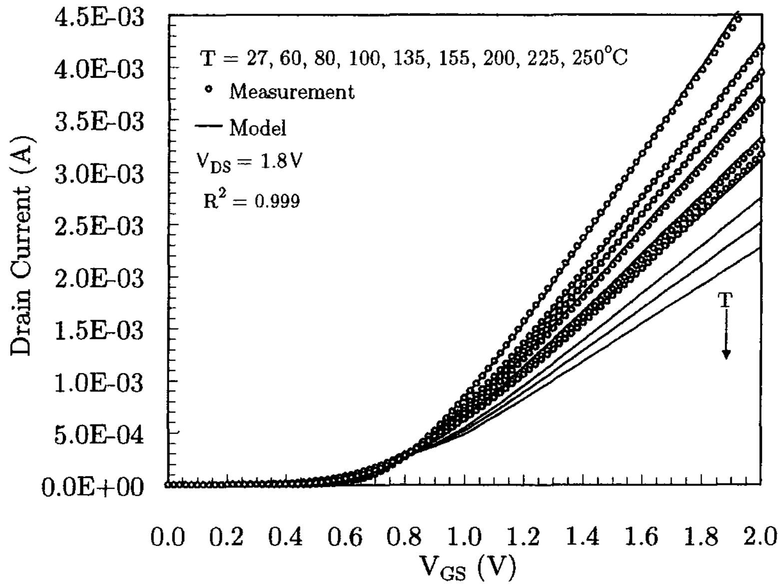


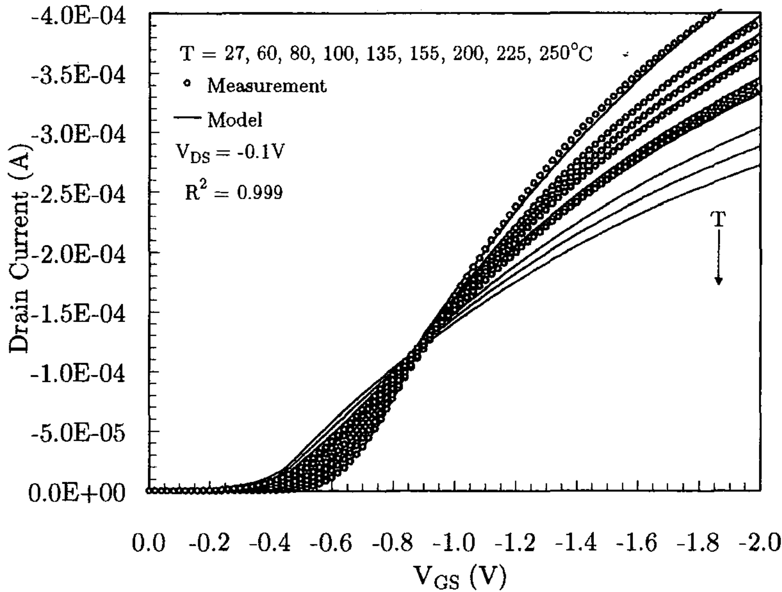
Figure 3.18: Temperature dependence of parasitic resistance

3.4.6 Temperature Results

Having incorporated all the temperature dependent components into the original I-V model, we now compare and analyse the influence of temperature on the MOS-FET operating characteristics. Figure 3.19 shows I_{DS} versus V_{GS} for a number of



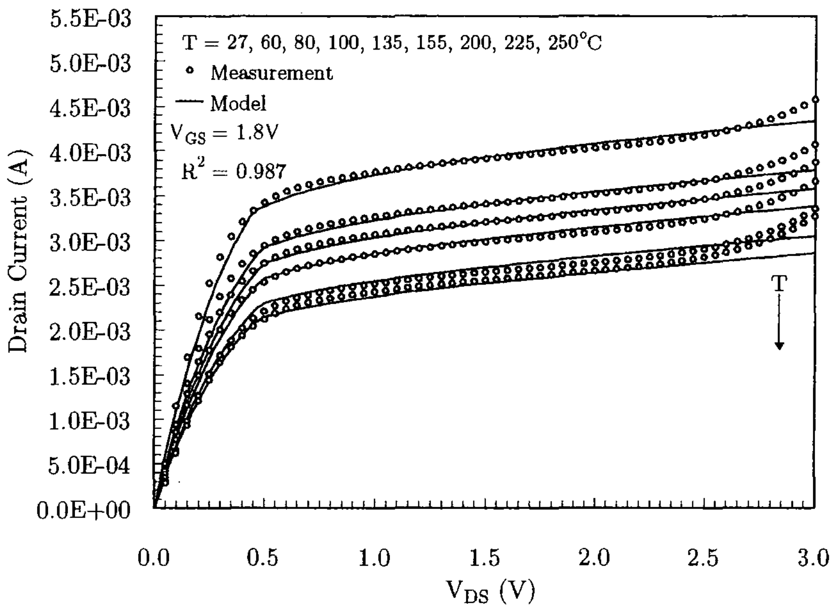
(a) n-channel



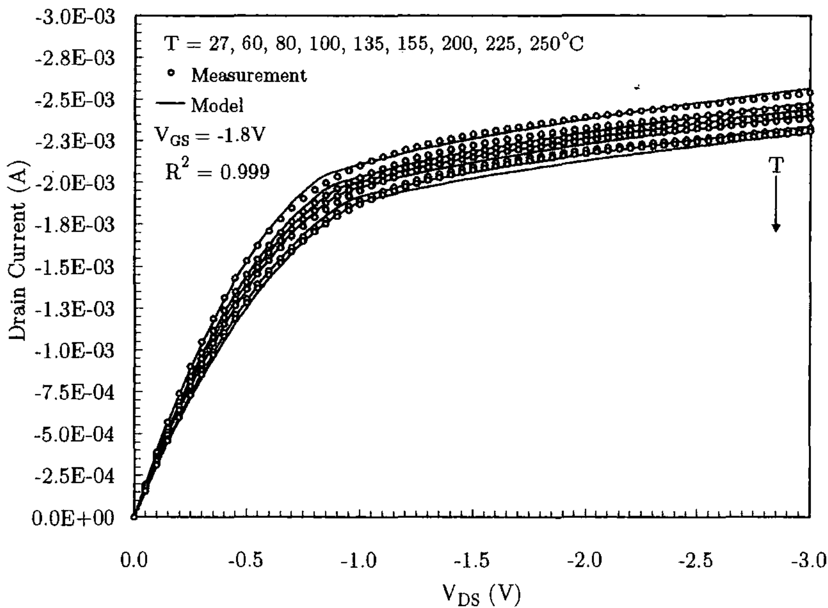
(b) p-channel

Figure 3.19: V_{GS} - I_{DS} characteristic at various temperatures

different temperatures. 'Figure 3.19(a) illustrates the influence of temperature on the saturation region of a n-channel device. The same characteristic is shown is figure 3.19(b) for p-channel, but this time biased in the linear region. In both figures the the influence of higher temperatures (>155°C) is also predicted using the model.



(a) n-channel



(b) p-channel

Figure 3.20: V_{DS} - I_{DS} characteristic at various temperatures

At low fields there is an increase in I_{DS} similar to what we saw in the subthreshold region (see figure 3.17). This creates a point where the drain current is not influenced by temperature. This point is called the Zero-Temperature Coefficient point (ZTC). Below the ZTC point, the channel is weakly inverted and the current increases due to increased carrier concentration. Above the ZTC point, a decrease in carrier mobility causes the change in I_{DS}. The resultant impact of the mobility change on drain current I_{DS} is further illustrated with the $I_{DS} - V_{DS}$ characteristics in figure 3.20.

3.5 Conclusion

The design of a new high temperature MOSFET model has been presented in this chapter. Deviating from the current trend of using a unified current expression, each operational region of the device has been considered separately. This is to ensure simplicity, ease of use, and efficiency.

The basic functional form of a fully physical model is retained whilst important physical phenomena are implemented through a semi-empirical approach. This is to achieve high computational efficiency and ease of parameter extraction. For high temperature characterisation, the model requires specifying the temperature dependence of at most, five parameters. A level 2 PSpice model requires eight parameters. Furthermore the model accounts for major physical effects in state-of-the art MOSFET devices such as velocity saturation, CLM, DIBL, and carrier mobility reduction due to a vertical field. The work strays from the tendency to concentrate on n-channel devices only, with both types always being considered.

Discontinuities at the boundaries between regions are overcome by using a weighting function that leads to a smooth and accurate model for the whole MOSFET operation. Both n- and p-channel simulations show good agreement with measured data. The temperature dependence of μ, v_{sat}, V_{th}, n, and R_{DS} are presented in section 3.4. At elevated temperatures the decrease in carrier mobility and threshold voltage are the main cause of change in drain current. Their competing influence creates a ZTC point where the drain current is stable over a wide range of temperatures.

Accuracy and efficiency were the two words in the introduction to this chapter that I believed should always be considered when designing a new model. This work extends the understanding of device modelling and derives simple formulae. Parameter extraction is very simple and requires only a small number of I-V measurements. However accuracy is not sacrificed and the model provides additional precision for high temperature prediction. Close resemblance between the calculated and measured data validates the model.

This work aims to design an improved high temperature I-V model. I believe it fulfils this, and forms the basis for failure mechanism modelling and the development of a simulation tool in a later chapter.

Chapter 4

NBTI

4.1 Introduction

NEGATIVE BIAS TEMPERATURE INSTABILITY (NBTI) has emerged as one of the biggest reliability concerns for today's CMOS technologies. It predominantly occurs in p-MOSFET devices when stressed at elevated temperatures with negative gate voltages. The effects on device characteristics are an increase in threshold voltage and decrease in drive current, causing a reduction in device speed.

At present there is no existing model that is comprehensive enough to predict all known NBTI effects. In this chapter a new model is developed to physically explain the effect of NBTI stress on device characteristics. The model is designed to accurately predict the influence of temperature and gate voltage on NBTI effects. Each development stage is validated with experimental data. Novel models for the incorporation of recently discovered NBTI phenomenon such as degradation recovery and a cumulative effect are presented. Additionally the impact on device lifetime under dynamic stress conditions is investigated.

4.2 Mechanism Overview

Negative bias temperature instabilities are caused by elevated temperatures and negative gate voltages. NBTI effects have been observed for gate voltages which result in a electric field of less than $6MVcm^{-1}$ [43]. This level of electric field is smaller than those capable of initiating hot-carrier effects. The rate of NBTI degradation is influenced by temperature and becomes more severe within the range of 100-250°C. Devices can encounter such stress conditions when they are in static state operation. An example circuit to demonstrate when devices experience NBTI stress is shown in the below figure.

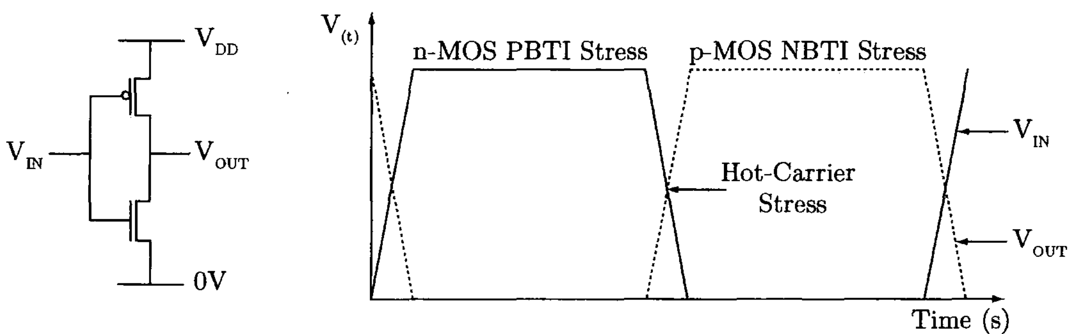


Figure 4.1: Stress time window

Considering the p-MOSFET there are two different stress time windows. The first is when the inverter input is low and output voltage is high. During this period no current flows through the p-MOSFET transistor and only the oxide is stressed by an electric field. This is when NBTI stress occurs. The second stress window is during the dynamic switching period. As the inverter output is pulled up, the p-MOSFET experiences hot-carrier stress. The same situation occurs for the n-MOSFET during the opposite dynamic stage when the inverter output switches to a low voltage level. The n-MOSFET also undergoes a static state stress when the input voltage is high and the output voltage low. Similar to NBTI, instabilities occur for n-MOSFET and p-MOSFET devices. For n-MOSFET devices the degradation

is called Positive Bias Temperature Instability (PBTI). To understand the relative impact of both NBTI and PBTI on p-channel and n-channel devices a comparative test was conducted. Each type of device was stressed under both NBTI and PBTI stress. The voltage and temperature stress conditions were chosen to significantly accelerate degradation. Figure 4.2 shows the results.

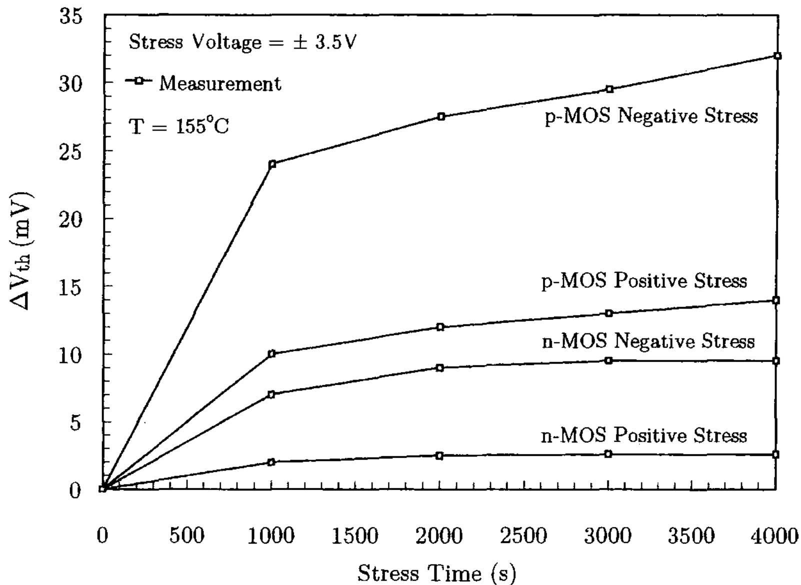


Figure 4.2: NBTI n-channel and p-channel degradation comparison

Both p-channel and n-channel devices under both NBTI and PBTI stress exhibit a change in threshold voltage. However the greatest degradation is seen in the NBTI stressed p-channel device. This is consistent with previous research conducted by Huard *et al* [46]. Huard proposed that NBTI stress is closely related to the tunnelling of holes into the $Si/Si0_2$ interface. Considering there is only a negligible hole density at the $Si/Si0_2$ interface in n-channel devices, degradation is minimal. Whereas p-channel devices exhibit the greatest degradation supporting Huard's proposal [46]. However, today the hypothesis involving the tunnelling of holes is still debated and at present there is still not a strict consensus on the precise physical mechanism behind NBTI.

Following Huard's hypothesis, it is believed that NBTI effects are a result of a build-up of positive charge at the silicon-oxide interface $Si/Si0_2$ and/or in the bulk oxide layer. When a gate voltage is applied, it initiates a field dependent reaction at the interface that generates interface traps. The interface traps are created by the breaking of passivated Si-H bonds at the interface and the subsequent movement of released hydrogen species (H) away, leaving behind positively charged interface states. Figure 4.3 illustrates this reaction-diffusion process.

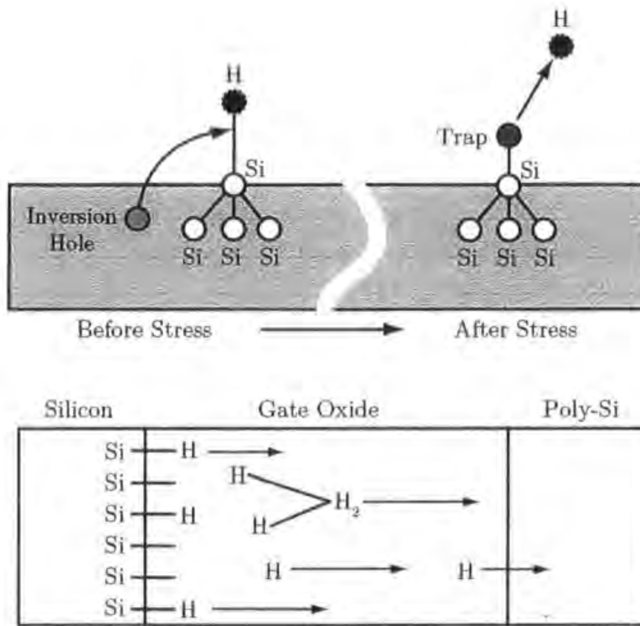
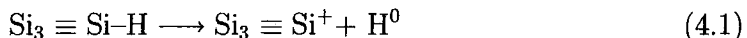


Figure 4.3: Interface trap creation

This Reaction-Diffusion (R-D) process was first proposed by Jeppson and Svensson [50]. It assumes that the dissociation of hydrogen species is caused by tunnelling of inversion layer holes that are captured by Si-H bonds. This weakens the Si-H bond, which is easily broken at higher temperatures. The newly released hydrogen diffuses away from the interface until it meets the poly-Si, leaving a silicon dangling bond (interface trap). Some H species are absorbed into the poly-Si and saturation occurs when all the Si-H bonds are broken.

The process by which interface traps are created can be schematically expressed as:



$$\mathrm{Si_3 \equiv Si\text{–}H \longrightarrow Si_3 \equiv Si^{+} + H^{0}} \tag{4.1}$$

where $\mathrm{Si_3 \equiv Si\text{–}H}$ is a hydrogen terminated trivalent silicon bond, which is electrically inactive in this form. However when the terminated hydrogen $\mathrm{H^0}$ is released we are left with a positively charged interface trap denoted by $\mathrm{Si^+}$. It is this increase in charge that contributes to the increase in threshold voltage seen after NBTI stress.

Although certain microscopic details of the NBTI mechanism are still debated such as how the reaction itself takes place, the R-D model has formed the basis for many NBTI investigations. Throughout this work it is applied as a foundation for the cause of NBTI and experimental results work towards validating its use.

4.3 Experimental Method

All NBTI results were characterised using the procedure outlined in the JEDEC Procedure for Measuring P-Channel MOSFET NBTI document [96]. A negative voltage higher than the nominal supply voltage to accelerate degradation was applied to the gate, whilst the drain, source, and substrate were grounded. Temperatures were varied from room temperature to a maximum 155°C. Periodically the gate stress was interrupted to measure (at nominal voltages) the threshold voltage and saturation current. Delays between stress and measurements were less than 500ms. The shift in V_{th} was monitored over a stressing period (up to 20,000s) to determine device lifetime. This is defined as the stress time when the threshold voltage changes more than the specified failure criteria ($\triangle V_{th}$ = 30mV). All measurements were taken using the apparatus specified in section 3.2.

4.4 NBTI Model

A physical model for NBTI voltage and temperature dependence is formulated in this section. Comparisons between model predictions and experimental results of such stress on device performance are presented. Throughout the design of the model the underlying aim to develop simple but accurate models that can be implemented in a common reliability framework is maintained.

The design of the model had to consider several factors. Firstly it needed to be a function of stress time. It also had to be capable of predicting the influence of electric field and stress temperature. Moreover the model had to be linked to the physical mechanism believed to be behind the cause of NBTI. In section 4.2 it was stated that we believe NBTI is a result of a reaction-diffusion process, and the generation of interface traps to be the main cause of V_{th} shift. Using this as a physical foundation we design a model around an expression for interface trap generation, N_{it} [43].

$$N_{it} = D \cdot {E_{ox}}^{3.5} \cdot t^{0.21} \cdot \frac{1}{t_{ox}} \cdot \exp\left(\frac{-E_a}{k_b T}\right) \tag{4.2}$$

where D is a technology constant, E_{ox} the electric field dependent on stress voltage, t stress time, and E_a the thermal activation energy. The fractional power law time dependence provides a simple solution to the reaction-diffusion process. During the early stress period the disassociated hydrogen species diffuse away from the interface causing a rapid increase in interface traps and threshold voltage shift. When the hydrogen diffusion front meets the poly-Si interface and there are fewer Si-H bonds to be broken the interface trap generation rate slows and V_{th} shift begins to saturate. The voltage dependence arises due to the electro-chemical nature of the NBTI stress and the best fitting model was found to be ${E_{ox}}^{3.5}$, where $E_{ox} = V_{GS_{Stress}}/t_{ox}$. An Arrhenius relationship is used to model the temperature dependence of NBTI.

Using equation 4.2, the net contribution of $\triangle N_{it}$ to the threshold voltage shift can be expressed as follows:

$$\triangle V_{th} = \frac{\triangle Q_{it}}{C_{ox}} \tag{4.3}$$

where $\triangle Q_{it}$ is the effective charge due to interface trap generation $\triangle Q_{it} = \triangle N_{it} \cdot q$. A comparison between model data and experiments is shown below in figure 4.4. A negative gate voltage was applied to a p-channel device whilst the source, drain, and substrate were grounded. The stress temperature was constant at 155°C.

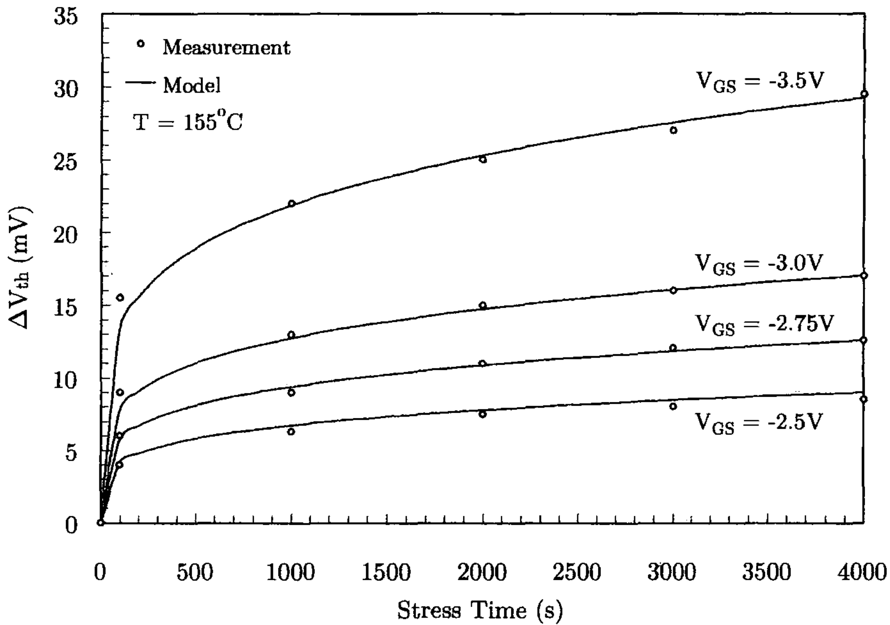


Figure 4.4: Threshold voltage shift under DC stress

Good agreement is seen between experimental and model data. With increased stress voltage, the threshold voltage degradation increases. This is believed to be due to tunnelling holes having greater energy and breaking the Si-H bonds more easily at higher stress voltages. Furthermore we can see the power law time dependence. In this case a power law coefficient of 0.22 was found to best fit experimental data. Different values for this coefficient have been published in the literature [46, 50], although it is commonly extracted to be around 0.25. The degradation is seen

to saturate with stress time. This is due to the hydrogen diffusion front reaching the polysilicon/silicon interface. Also the hole tunnelling probability reduces with time as there are fewer Si-H bonds to be broken. This saturating feature seen in the experimental results supports the use of the theoretical reaction-diffusion process described in section 4.2. As well as voltage experiments fresh devices were stressed at different temperatures. The gate voltage was constant at -2.75V whilst various stress temperatures applied. Periodically the stress temperature was monitored to ensure a constant temperature throughout the experiment. The results are illustrated below.

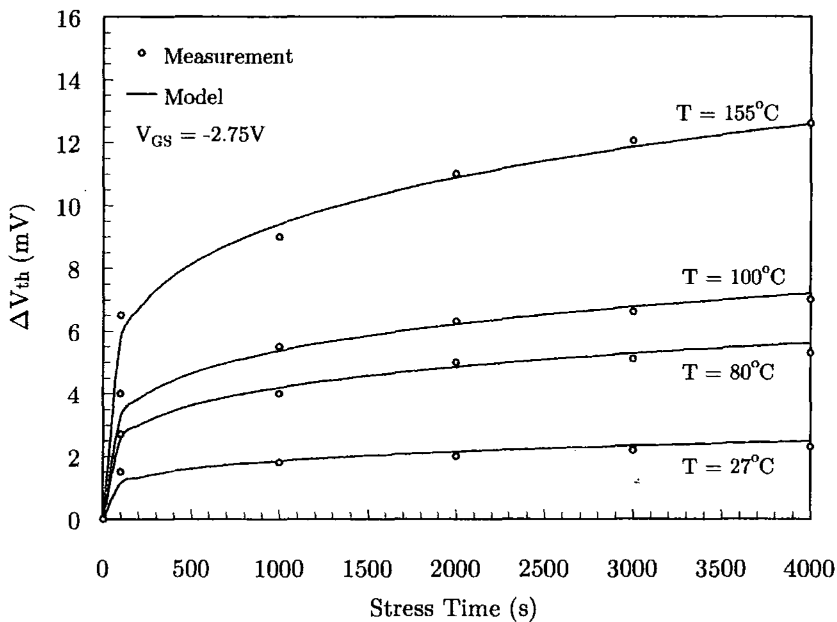


Figure 4.5: NBTI stress as a function of temperature

There is good correlation between experimental and model data, supporting the use of the Arrhenius expression in equation 4.2 to predict NBTI temperature dependence. Higher temperature enhances NBTI degradation, thus the degradation process is thermally activated. This is due to the thermal energy increasing hole tunnelling probability, and enhancing the diffusion of hydrogen species. Hence at higher temperatures interface trap generation, and effective oxide charge is greater.

The activation energy was extracted from the Arrhenius plot shown in figure 4.6. For the p-channel device stressed at -2.75V we obtain $E_a = 0.14\text{eV}$. This is comparable to the range of activation energies (0.12 - 0.16eV) that have been reported in the literature recently [46, 52].

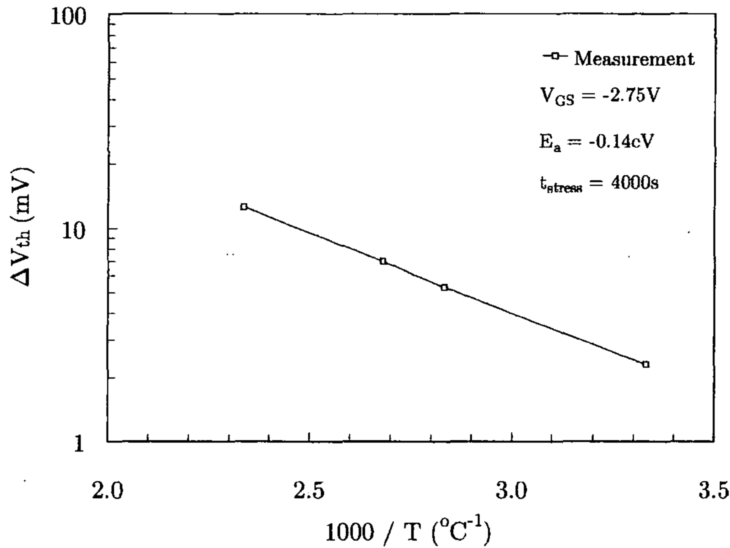


Figure 4.6: Arrhenius plot showing temperature dependence of ΔV_{th}

The temperature and voltage results presented in this section validates the use of equation 4.2. These results provide a basis for DC stress predictions. However it has been observed that after the removal of this stress there is a recovery phase that needs to be incorporated into the model.

4.5 Recovery Effect

Until recently it was assumed that after stress removal, NBTI degradation was permanent. However it has been observed that some of the threshold voltage shift induced by NBTI stress can be recovered. As continuous DC stress is seldom seen, it is essential that this post stress phenomenon is contained in the model for a complete picture of NBTI.

It is believed that when stress is removed, hydrogen species diffuse back to the Si/SiO_2 interface, thereby recovering some of the V_{th} degradation. To explore this, we performed an experiment with measurements taken during and after stress periods. A fresh device was stressed for up to 4000s and then allowed to recover for another 10,000s. The temperature was constant as 155°C. The measurement results are shown below in figure 4.7.

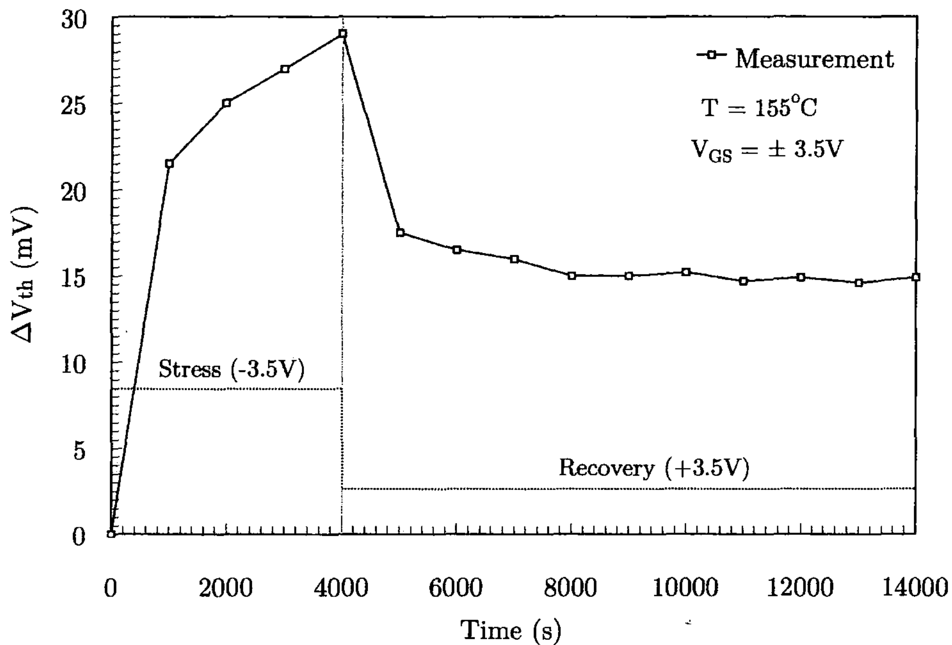


Figure 4.7: Threshold voltage recovery after NBTI stress

The recovery consists of a fast initial transient followed by a slow slope which appears to saturate with time. We presume that these characteristics of the recovery are due to the origin of interface trap generation. When an opposite voltage is applied during the recovery period the hydrogen species diffuse back from the gate oxide bulk towards the Si/SiO_2 interface. Thus the hydrogen can now diffuse back and recombine with silicon dangling bonds restoring them to their passive Si-H state. It is believed that because not all broken Si-H bonds are re-passivated the degradation does not fully recover. Another proposal is that the permanent damage may be

caused by hydrogen species that have reached the poly silicon electrode during the stress phase, which cannot be recovered during the relaxation phase. To further understand this, we also undertook experiments to investigated if the recovery effect was field dependent. During the recovery stage different positive voltages were applied to the gate as well as taking measurements when the gate electrode was grounded. The experimental results are illustrated below.

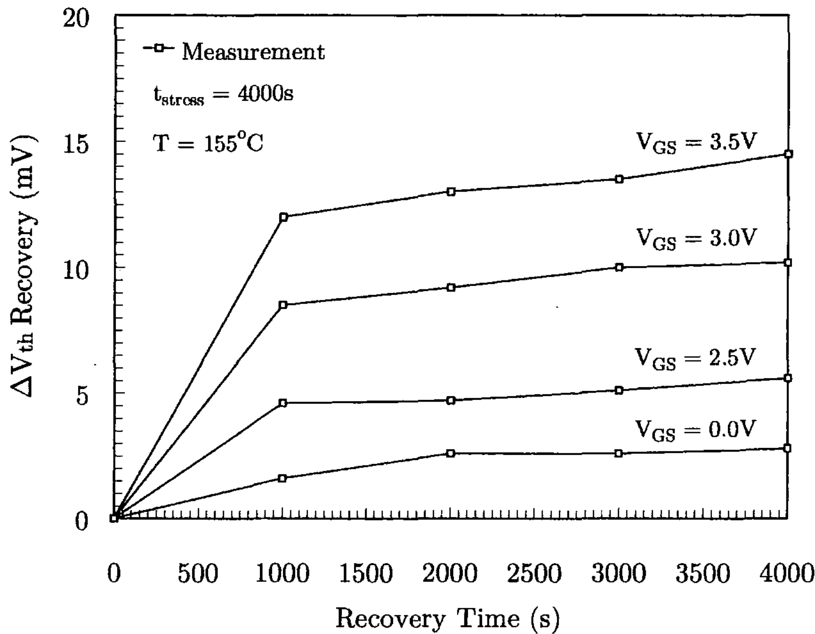


Figure 4.8: Threshold voltage recovery as a function of gate voltage

Several interesting observations can be deduced from the above figure. Firstly greater recovery is seen when more positive voltages are applied, and the rate of recovery is a function of the gate voltage. Secondly even when no voltage is applied during the recovery period, V_{th} recovers a certain amount. The fact that V_{th} recovers over time when no stress is applied is an important result as it supports the reaction-diffusion hypothesis. Under no applied field, the hydrogen species diffuse back to the silicon-oxide interface. This is enhanced when a positive voltage is applied to the gate, comparable to the NBTI stress characteristic. On this basis we adopt a similar model as that of equation 4.2, for NBTI recovery. The coefficient for

the recovery period was extracted from figure 4.8 to be 0.17. Thus the new model proposed for voltage dependency of NBTI recovery is given by:

$$R_{ec} = R_1 \cdot {E_{ox}}^{2.5} \cdot t^{0.17} \cdot \frac{1}{t_{ox}} \cdot \exp\left(\frac{-E_a}{k_b T}\right) \tag{4.4}$$

where R_1 is the recovery fitting constant. To simulate the recovery when the gate is grounded the dependency on oxide field was removed from the expression as shown below in equation 4.5. The time coefficient under 0V recovery conditions also differed slightly. This is because the diffusion process is enhanced when a positive voltage is applied and therefore tends to saturate quicker.

$$R_{ec}(0V) = R_2 \cdot t^{0.28} \cdot \frac{1}{t_{ox}} \cdot \exp\left(\frac{-E_a}{k_b T}\right) \tag{4.5}$$

Using both the stress and recovery models, the figure below shows a comparison to experimental data. Devices were stressed at varying voltages up to 4000s followed by a recovery period. Notice the relative decrease in recovery rate when the device is left to recover at 0V compared to $V_{GS} > 0$.

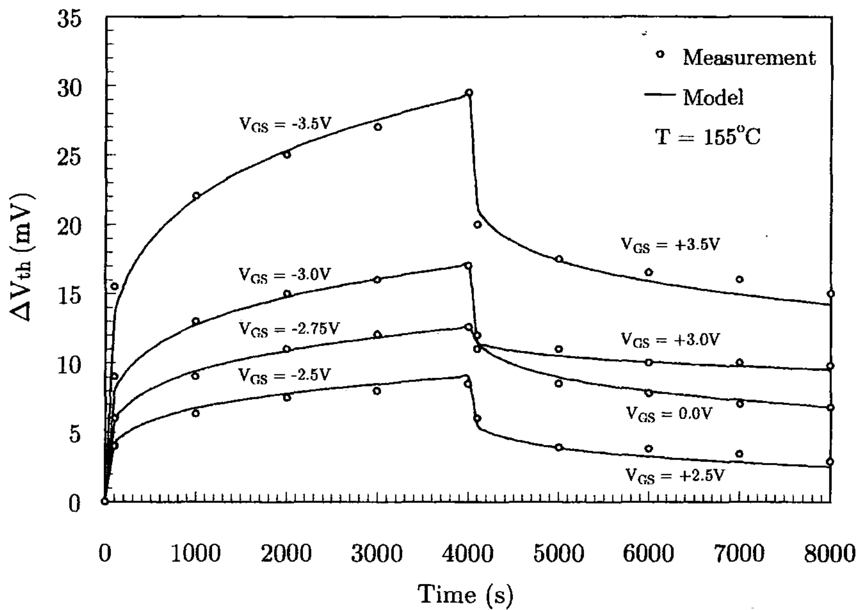


Figure 4.9: Comparison between NBTI model and experimental results

Figure 4.9 shows that the model efficiently predicts the change in threshold voltage during both stress and post-stress phases. These two key features of NBTI are modelled on the physical basis of interface trap creation and re-passivation of Si-H bonds during recovery. We must also study the effect of temperature in V_{th} recovery. The recovery model (see equation 4.4) contains an Arrhenius relationship to account for the temperature dependence. The author is presently unaware of any model that accounts for the temperature dependence of NBTI recovery. To extract their activation energy, the devices were stressed and then allowed to recover ($V_{GS} = 0V$) at various temperatures. The results are illustrated below, showing a comparison to the NBTI stress in an Arrhenius plot.

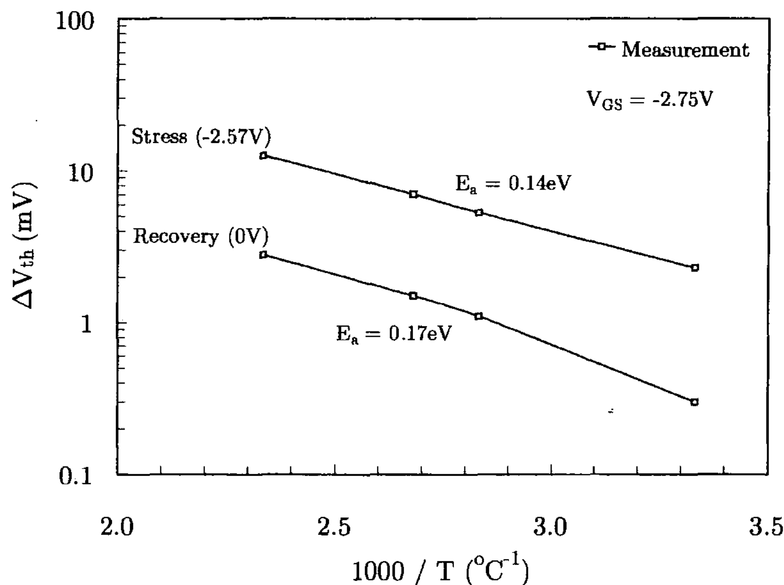


Figure 4.10: Comparison between stress and recovery activation energies

The results indicate that the recovery is enhanced at higher temperatures. We hypothesise that both the diffusion and re-passivation of Si-H bonds increase with elevated temperatures, thus promoting V_{th} recovery. The activation energy for NBTI recovery was extracted to be $E_a = 0.17$eV. This is greater than that during the stress period. This could be due to an electro-thermal interaction enhancing interface trap creation, as the gate was grounded during recovery measurements.

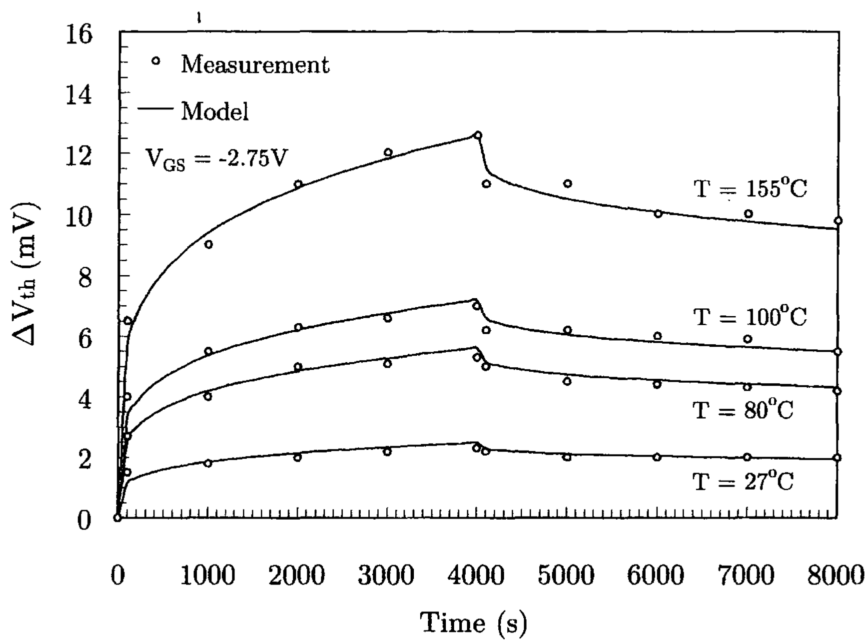


Figure 4.11: Temperature dependence of NBTI stress and recovery

The ability of the model to predict the NBTI temperature dependence during both stress and recovery is shown in the above figure. Good agreement between experimental and model data can be seen. Both the voltage and temperature dependencies are accurately traced by the model predictions. This supports the use of equations 4.2 and 4.4 is the design of this new NBTI model.

4.6 Lock-in

Another new feature built into the model is the lock-in effect. This is sometimes referred to as the cumulative effect and there is no model at present that accounts for its effect. We have seen that NBTI stress can recover a fraction of the V_{th} shift. However if the device is re-stressed the degradation reappears. We have speculated that this permanent damage is because hydrogen has reached the polysilicon electrode during the stress phase and the charge cannot be recovered during relaxation.

To investigate this we stressed a device over 20,000s, periodically removing the stress and allowing the device to recover at $V_{GS} = 0V$.

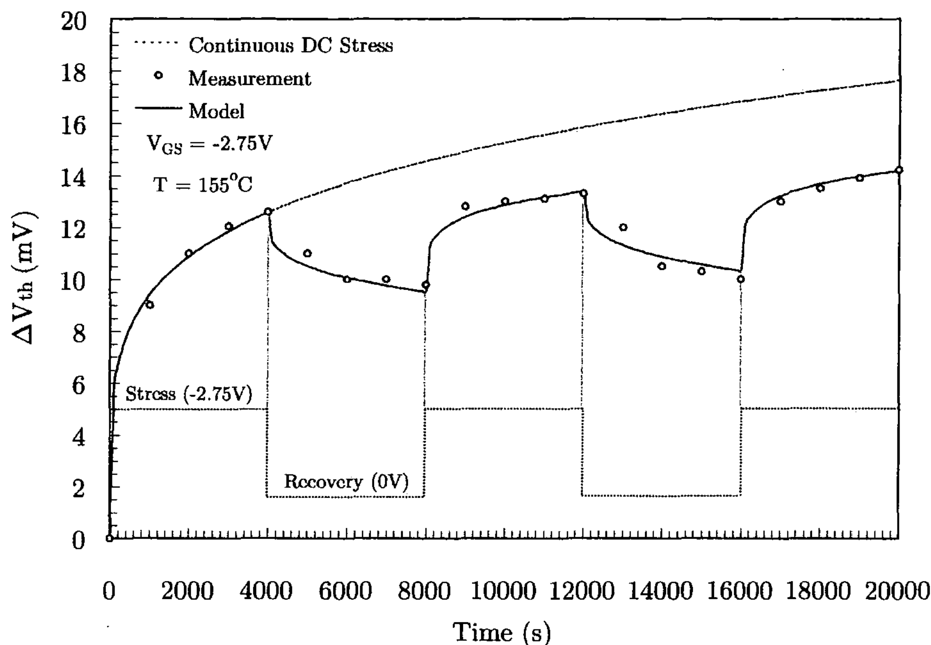


Figure 4.12: NBTI lock-in illustration

In figure 4.12, at each stress interval V_{th} degradation returns quickly at first, then continues to degrade more slowly. Over the total 20,000s test period, the net degradation is less than that for the same DC-equivalent stress. The ratio of the reversible component to the permanent component decreases with time, so the relative importance of the recovery becomes smaller with time. From these results we support the speculation that the lock-in effect is caused by hydrogen that has diffused into the poly-silicon electrode interface and cannot be recovered. To implement this characteristic into the model, a very simple numerical equation is used.

$$\triangle V_{th} = \triangle V_{th}(t = t_{stress}) \cdot (T^{-0.91} + L_0) \tag{4.6}$$

where t_{stress} is the time at which the stress is re-applied, T temperature, and L_0 is a constant that determines the relative increase in $\triangle V_{th}$ at each stress interval.

Figure 4.13 shows comparisons of model data to experimental recovery and re-stress measurements. The model is able to capture the observed experiential trends accurately. Each time the stress is reapplied the degradation increases, but over time this increase reduces. It appears that there is a slight increase in the relative lock-in value at decreasing temperatures. A valid explanation for this is that at higher temperatures the diffusing hydrogen is encouraged to move back away from the poly-Si interface.

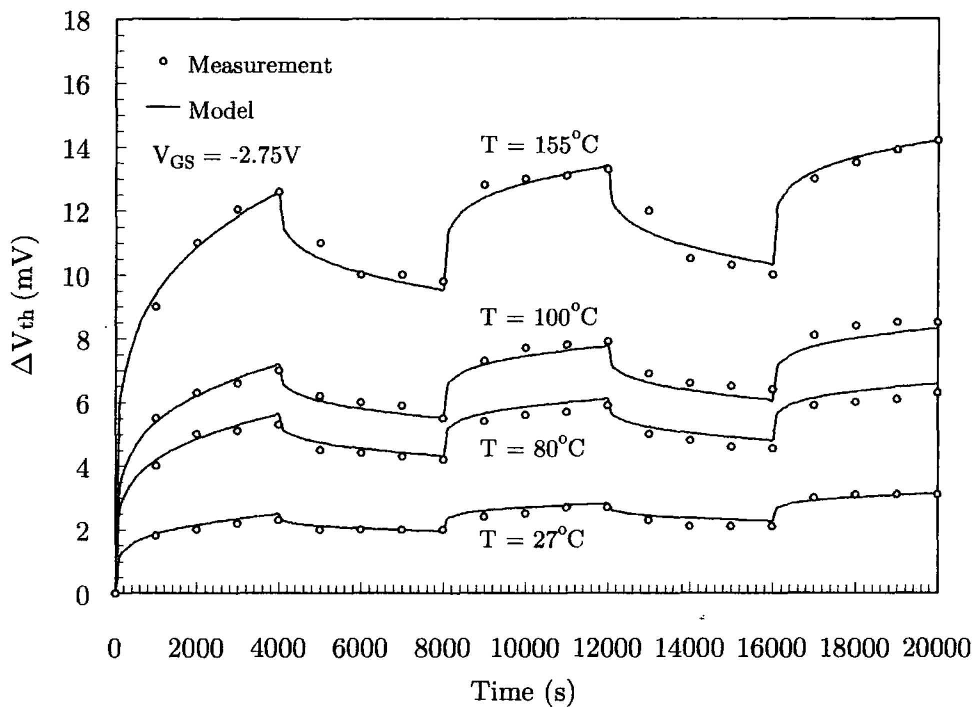


Figure 4.13: NBTI lock-in temperature dependence

At this point in the experimental analysis, the results we had collected, not only gave us greater understanding of NBTI but also guided us into the next area of research. We deduced that if NBTI effects are recovered during a relaxation period then this would imply that NBTI degradation could be an influence in applications where dynamic signals are used. Therefore the impact of dynamic degradation is the next subject of investigation.

4.7 Dynamic Degradation

In practical applications, devices rarely experience continual DC stress. Typically a device may be exposed to periodic AC and DC stress conditions. Given that NBTI is a relatively new failure mechanism, research on NBTI effects under pulsed or dynamic NBTI stress is limited. The literature that is available has shown varied results. Chen *et al* [55] presented results showing dynamic NBTI degradation (DNBTI) was less than that of equivalent static NBTI (SNBTI) stress and DNBTI effects were frequency independent. In contrast Mahapatra *et al* [49] proposed that interface trap creation and thus $\triangle V_{th}$ decreased at higher frequencies. In this section we further the research needed in this area by presenting new findings on NBTI frequency dependency. We show that device lifetime is substantially enhanced under DNBTI conditions. We also present a model that can accurately predict device degradation due to DNBTI stress and can be easily incorporated into the complete NBTI framework.

To analyse DNBTI effects we simulated the stress conditions a p-channel device is exposed to in a CMOS inverter. The test configuration is show in figure 4.14.

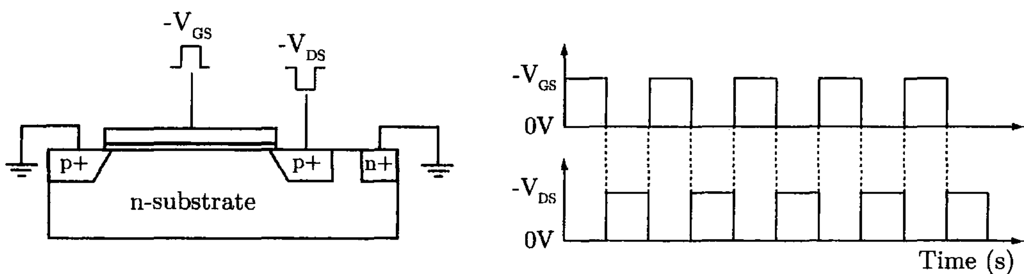


Figure 4.14: Dynamic stress configuration

Pulsed signals are applied to both gate and drain of the device. They are supplied out of phase to simulate actual p-MOSFET bias operating conditions. Thus in this configuration the device will alternate between ON and OFF stressing modes.

Figure 4.15 shows a comparison between DNBTI and SNBTI measurements. The stress voltage for gate and drain junction was -2.75V. The temperature was constant throughout the experiment at 155°C. For all DNBTI stress the duty cycle was 50%.

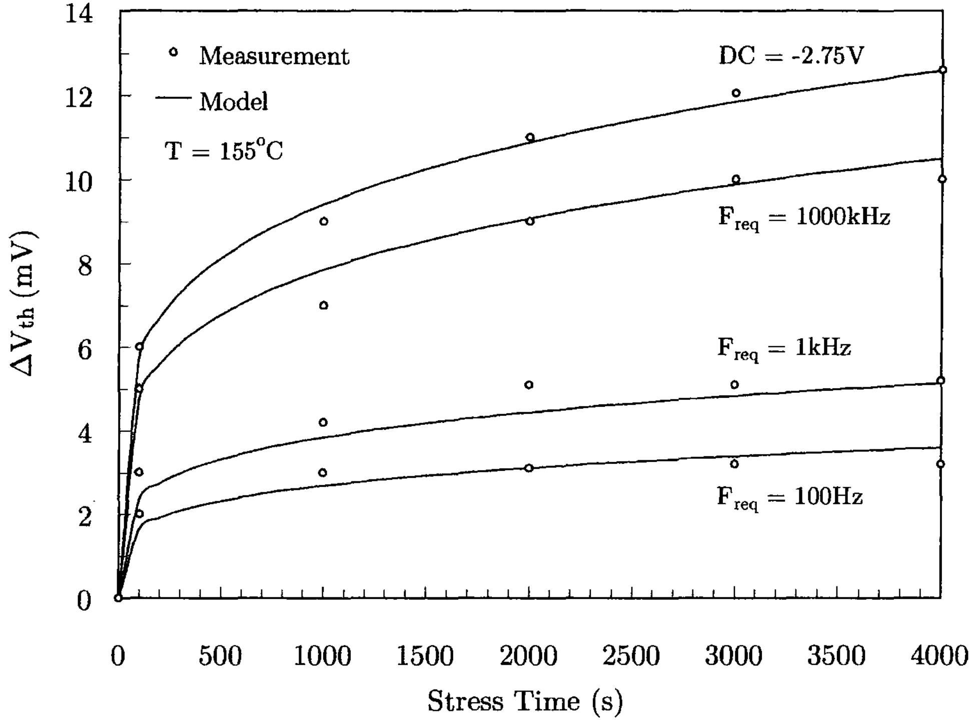


Figure 4.15: NBTI frequency dependence

For all DNBTI stress conditions V_{th} shift is reduced at any given time, compared to the worst case DC condition. Additionally as frequency is increased, the degradation increases, disagreeing with the results presented by Mahapatra *et al* [49]. We believe that at higher frequencies the recovery effects seen in section 4.5 do not have sufficient time to take place. At lower frequencies the relaxation period is greater and therefore allows time for the diffusion of hydrogen species back towards the Si-Si0$_2$ interface and re-passivation of Si-H bonds. Therefore for longer t_{off} during DNBTI stress (lower frequency) it is hypothesised that the hydrogen species have more time to recover and thus less overall degradation occurs.

From these results we developed a simple numerical expression to predict DNBTI frequency dependence. We propose that V_{th} degradation follows a power law dependence on frequency and can be given by:

$$\triangle V_{th} = \triangle V_{th}(DC) \cdot R \cdot \left({F_{req}}^{0.15}\right) \quad (4.7)$$

where R is a process constant, that can be obtained along with the frequency exponent from a small number of frequency experiments. From figure 4.15 it can be seen that the model is able to capture the observed experimental trends accurately.

Further comparisons were made between SNBTI and DNBTI degradation, analysing device lifetime. Figure 4.16 shows a plot of lifetime as a function of operating voltage V_{DD} where lifetime is defined as the time needed for $\triangle V_{th}$ to reach 30mV.

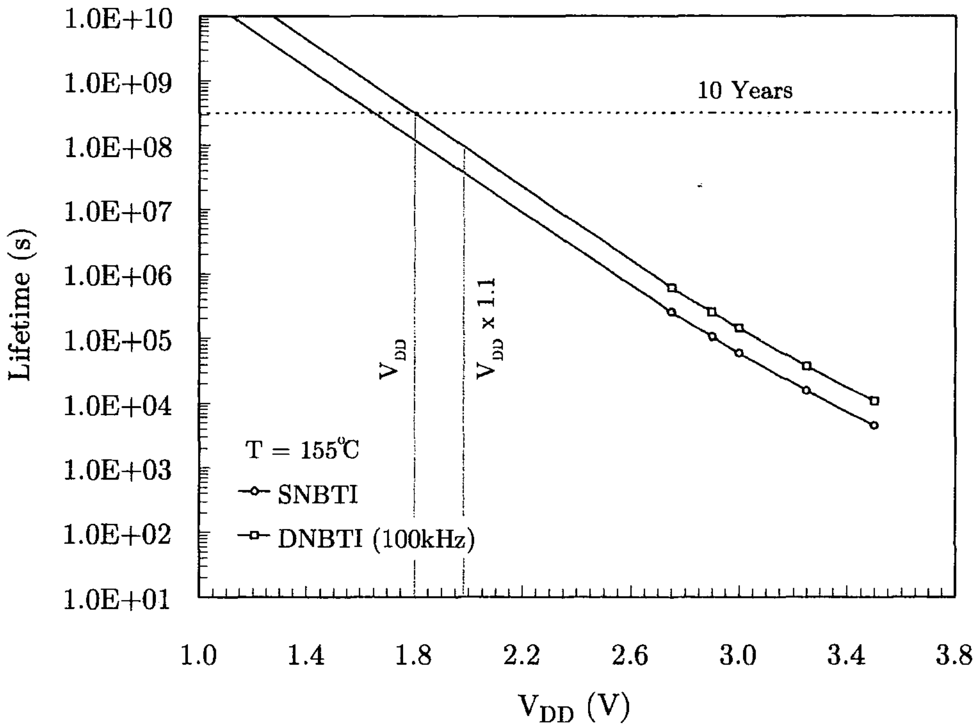


Figure 4.16: NBTI lifetime comparison

Extrapolating lifetime to normal operating conditions (V_{DD} x 1.1 for safeguarding) device lifetime is just under 2 years for SNBTI, whereas it reaches $4\frac{1}{2}$ years for DNBTI. The ten year operating voltage is 1.62V and 1.8V for SNBTI and DNBTI respectively. The enhanced lifetime for DNBTI stress is due to the longer relaxation period and re-passivation of Si-H bonds described earlier in this section. As a result, if DNBTI effects are not considered the lifetime of devices that experience dynamic signals could be overestimated if predicted under only SNBTI stress conditions.

4.8 Conclusion

In this chapter we have developed a novel model that encompasses several new NBTI phenomena. We began with experimental results comparing NBTI and PBTI stress in both n-MOSFET and p-MOSFET devices. Greatest degradation was found to occur in p-MOSFET devices stressed with a negative gate voltage. It was hypothesised that this degradation was due to the breaking of Si-H bonds at the $Si/Si0_2$ interface and creation of interface traps. On this basis a model was designed to predict NBTI voltage and temperature dependencies. At both higher voltages and temperatures V_{th} degradation was found to increase due to increased interface trap generation. Recovery effects, as well as the lock-in effect were included in the model. Experimental results demonstrated that in practical operating conditions where devices experience dynamic signals, lifetime is prolonged. This is due to hydrogen diffusing back to the $Si/Si0_2$ interface during the relaxation periods. At all stages the model predictions were supported by experimental results.

In conclusion, several relatively new NBTI effects have been researched. A simple but accurate and comprehensive NBTI model has been designed that can be easily implemented into a complete reliability framework.

Chapter 5

GIDL

5.1 Introduction

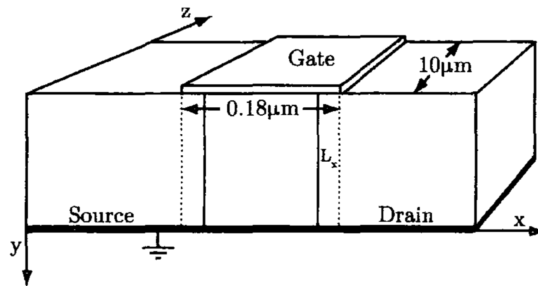
IN THIS CHAPTER we investigate the temperature dependence of an OFF-state leakage current that is posing serious limitations on future device scaling. GIDL is the Gate Induced Drain Leakage. This leakage current is caused by a gate-induced high electric field in the gate-to-drain overlap region which many researchers have attributed to band-to-band tunnelling [28, 32, 35]. The relentless reduction of gate oxide thickness mean suppression of GIDL is becoming an increasingly difficult technological challenge, and vital new models are developed as a result.

The need for detailed models of GIDL is particularly important for ULSI[1] applications, where any increase in OFF-state power dissipation is detrimental. Very little work on GIDL has focused on the effect of temperature [37]. We therefore emphasise the analysis of this aspect and present a new temperature dependent GIDL model that accurately follows experimental trends.

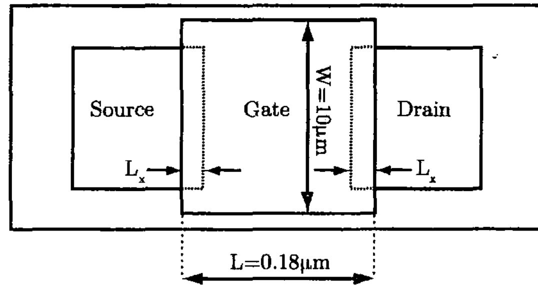
[1]Ultra Large Scale Integration

5.2 Mechanism Overview

GIDL has been identified as a major drain leakage phenomenon in OFF-state MOSFET's. GIDL has been detected in DRAM and EEPROM cells and is the dominant leakage mechanism responsible for discharging the storage node [35]. There has been considerable interest in the study of the mechanisms responsible for GIDL current [28, 29, 35]. Several mechanisms have been proposed to describe the behaviour of the leakage current and many have attributed it to band-to-band tunnelling taking place in the gate-to-drain overlap region.



(a) Cross-section (not to scale)

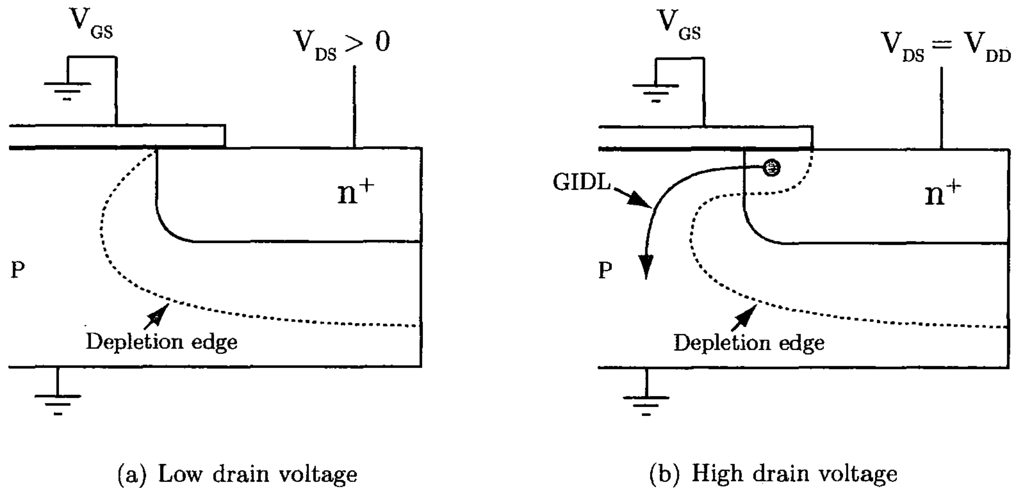


(b) Plan view (not to scale)

Figure 5.1: Illustration of gate-to-drain overlap region

Figure 5.1(a) illustrates the cross-section of a MOSFET device, where the gate length L, gate width W, and gate-to-drain overlap region L_x are indicated. Figure 5.1(b) shows the equivalent device from above. From figure 5.1(a) we can see that during device fabrication the gate length does not equal the physical dimensions of

the gate, but rather the distance between the source and drain regions underneath the gate. The reason for this is that an overlap between the gate and source/drain regions is required to ensure that the inversion layer forms a continuous conducting path between the drain and source. The overlap is made as small as possible in order to minimise its parasitic capacitance. Typically for devices smaller than 0.35μm the gate-to-drain overlap L_x < 150nm.



(a) Low drain voltage

(b) High drain voltage

Figure 5.2: GIDL mechanism (n-channel)

Figure 5.2 illustrates how the GIDL mechanism occurs in the gate-to-drain overlap region. In this case we describe GIDL in an n-channel device, but the process is the same for p-channel devices except the roles of the electrons and holes are reversed. When a voltage is applied to the drain with the gate grounded (figure 5.2(a)), an accumulation layer is formed at the silicon surface. Due to the presence of accumulated holes at the surface, it behaves like a p-type region more heavily doped than the substrate. This causes the depletion layer at the surface to be narrower than anywhere else in the channel. As a consequence there is field crowding and an increase in the local electric field in this region. If the drain voltage is increased further or the gate voltage made more negative, the n+ region can become depleted and at sufficiently high voltages even inverted as shown in figure 5.2(b). Hence if

the field crowding and peak electric field is high enough, there is an increase in both the tunnelling probability and the generation of electron-hole pairs. Due to the substrate being at a lower potential for minority carriers, any minority carriers created (holes in n-MOSFET's and electrons in p-MOSFET's) in the drain region flow to the substrate due to the lateral field. Any electrons (holes in p-MOSFET's) generated are collected by the drain resulting in the increase of the GIDL current.

5.3 Experimental Method

To measure the GIDL current, both the gate and source electrodes were grounded whilst the drain voltage was swept from 0 to ±5V. Current measurements were taken every 0.05V. Normal $V_{GS} - I_{DS}$ measurements were also taken at various drain voltages to analyse the impact of GIDL on operating characteristics. To compare n-channel and p-channel devices the same measurement voltages and conditions were applied to both types of devices. The maximum drain voltages were chosen so as not to cause rapid breakdown at the gate oxide. For the devices under test the breakdown voltage specified by the manufacture was 3.8V and -5.0V for n-channel and p-channel devices respectively. The gate oxide thickness was 4.2nm and the channel length 0.18μm. The gate to drain overlap region was estimated to be 0.1μm. The OFF-state current for the devices was specified by the manufacturer to be less than 50pA. The temperature was varied between room temperature and a maximum of 155°C. All measurements were carried out in electrically shielded conditions using the Keithley 2400 and 6487 precision pico-ammeters. These ammeters are capable of measuring in the pico-amp range with a measurement resolution of ±10pA. For further details of the experimental set-up please see Appendix A.

5.4 GIDL Model

In this section we develop a new GIDL model. We begin by highlighting the inaccuracies in present models, and then proceed to describe each design stage of our model. The model is designed to provide accurate predictions over a wide range of gate and drain voltages and also the capability of predicting the influence of temperature. Its design also allows for easy incorporation into a complete I-V model.

To date a number of GIDL models based on tunnelling theory have been proposed in the literature [28, 35, 36]. All of the models are very similar, calculating the GIDL current using a simple 1-D band-to-band tunnelling expression given by:

$$I_{GIDL} = A \cdot E_s \exp\left(\frac{-B}{E_s}\right) \tag{5.1}$$

where A and B are fitting constants and E_s the electric field at the Si-SiO_2 interface in the gate-to-drain overlap region. Figure 5.3 shows a comparison between the tun-

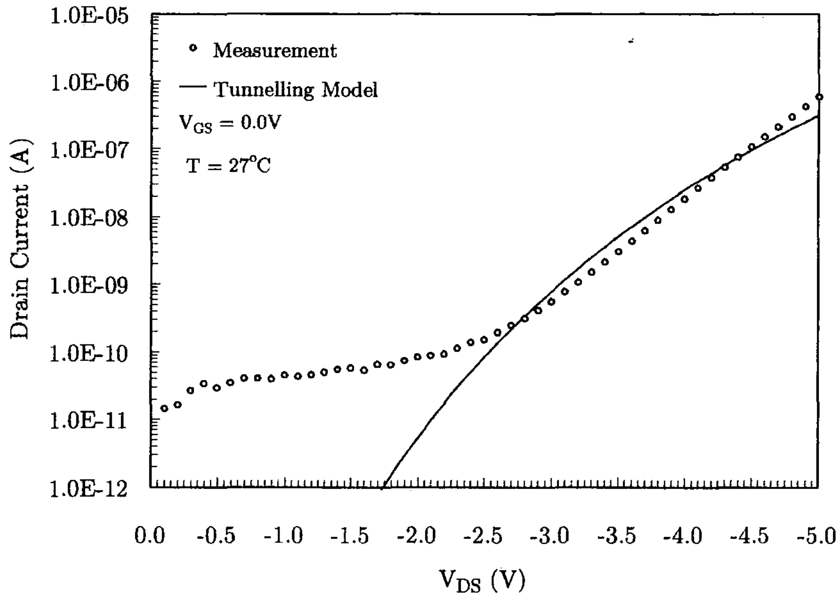


Figure 5.3: GIDL characteristic (p-channel)

nelling model and experimental data for a p-channel device. The gate was grounded and the drain swept from 0V to -5V to distinguish the GIDL characteristic. Despite carefully fitting with constants A and B, the tunnelling model is not precise. It does follow the trend above $\approx$ -2.5V which is expected as this is the region where tunnelling may occur, but because this is a model that describes tunnelling any leakage current below $\approx$ -2.5V is not accounted at all. Furthermore, no direct temperature dependence is provided for, and its simple closed form makes it difficult to implement into a current-voltage model. Currently there is only one GIDL model for use in a SPICE like simulator, which is a complicated expression in the BSIM4 model [42]. Our aim from the outset was to provide simple, easy to use models.

The first consideration in our model design was that we wanted to provide an expression that could model GIDL for the entire range of drain voltage. In our first design stage we used a simple exponential expression and figure 5.4 shows a comparison with the measured data and existing tunnelling model.

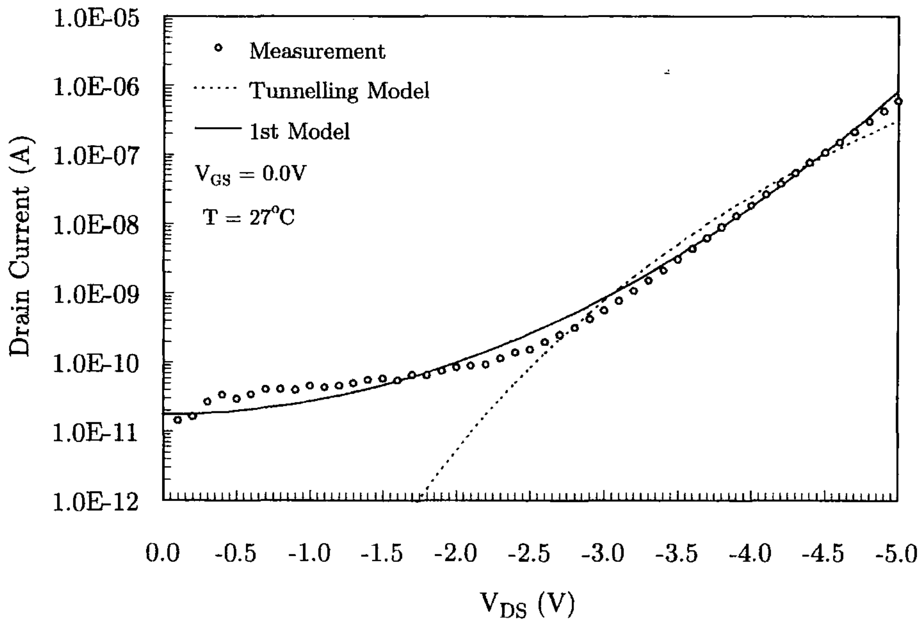


Figure 5.4: Comparison of first GIDL model with measured data

The exponential equation used in the first model was given by:

$$I_{GIDL} = e^{\left(S \cdot V_{DS}{}^{2}\right) - I_0} \tag{5.2}$$

where I_0 is the zero offset current and S the plot gradient. We can see from figure 5.4 that the calculated results agree well for a range of drain voltages, but it fails to accurately match the measured data between -1.5 and -3.0V. This is because in this region the physical mechanism producing the leakage current is changing. In the low field region ($V_{DS} < 2.0$V) leakage is caused by diffusion but in the high field region ($V_{DS} > 3.0$V) tunnelling is the main contributor to increased current.

In order to develop an accurate model for the complete GIDL characteristic we decided that each region should be modelled by separate functions. Given that the GIDL effect consists of two linear regions on the log plot, equation 5.2 was developed further separately for each region. To describe the GIDL current in the low field region we used the following expression:

$$I_{GIDL(L)} = e^{(S_L \cdot V_{DS}) - I_0} \tag{5.3}$$

where I_0 is the zero offset current and S_L the gradient in the low field region. Using a similar expression the GIDL current in the high field region was given by:

$$I_{GIDL(H)} = e^{[S_H \cdot (-V_{DS}+V_{MAX})] - I_T} \tag{5.4}$$

where I_T is the current tunnelling point, S_H the gradient in the high field region, and V_{MAX} the maximum drain voltage. The parameters I_0, I_T, S_L, S_H and V_{MAX} can be extracted from one simple device measurement to allow accurate fitting to the experimental data.

To achieve a smooth transition between the two expressions a transition function G_{Tran} is used to move between $I_{GIDL(L)}$ and $I_{GIDL(H)}$ through a mid-point G_{MP}.

$$G_{Tran} = \left[w \cdot I_{GIDL(H)}\right] + \left[(1-w) \cdot I_{GIDL(L)}\right] + \left[2\,(1-w) \cdot w \cdot G_{MP}\right] \qquad (5.5)$$

where w is the weighting function that goes between 0 and 1 during the transition. The transition range is defined by the $V_{DS(low)}$ and $V_{DS(high)}$ drain voltage values.

$$w = V_{DS} - V_{DS(low)} \cdot \left(\frac{1}{V_{DS(high)} - V_{DS(low)}}\right) \qquad (5.6)$$

Figure 5.5 illustrates how this transition function is implemented to establish a complete GIDL characteristic. Beyond the transition point $V_{DS(low)}$, the model moves from the low field expression to the transition function and then onto the high field expression above $V_{DS(high)}$ providing a smooth function for the whole range of drain voltage values.

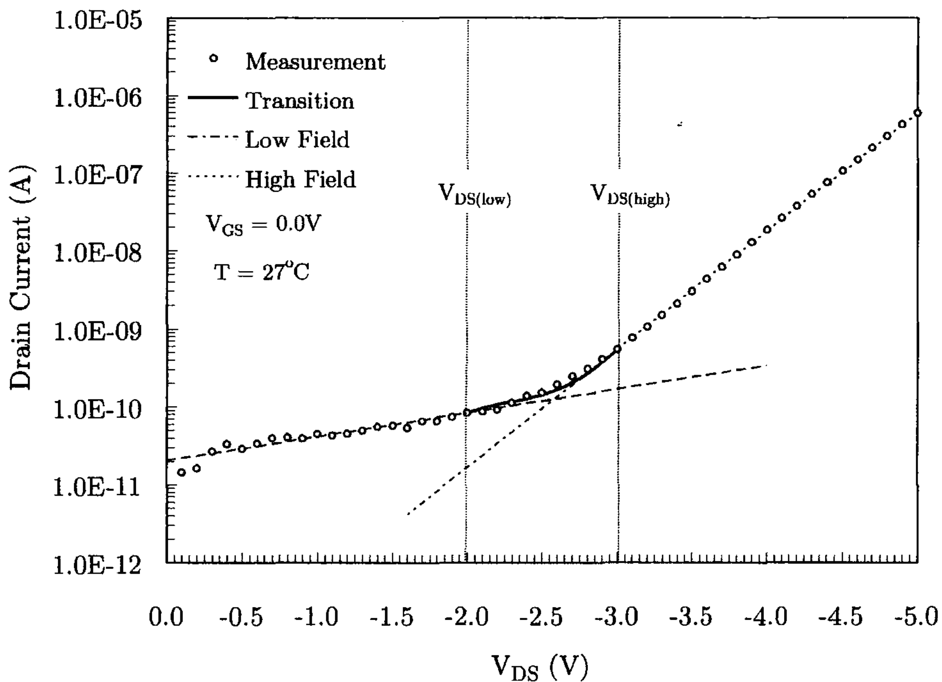


Figure 5.5: Illustration of GIDL transition function

By using separate equations to describe each region of the characteristic the accuracy has been improved. It can be seen in figure 5.6 that during the transition from diffusion to tunnelling current (-2.0V $< V_{DS} >$ -3.0V) and also in the low field region the modelling accuracy is particularly refined. Additionally, in comparison to the tunnelling expression, our new GIDL model provides for prediction over the total range of operating drain voltages.

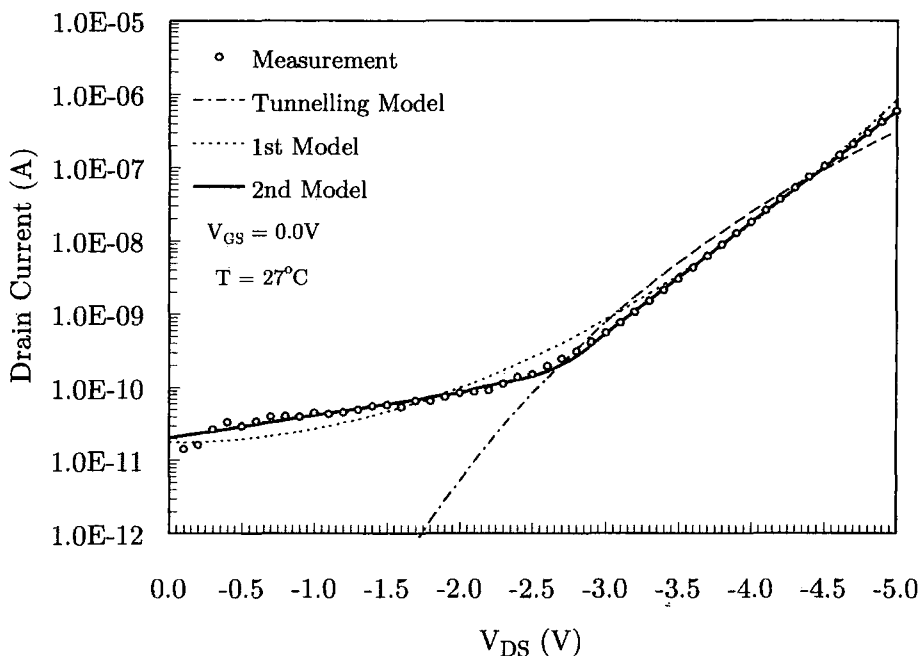
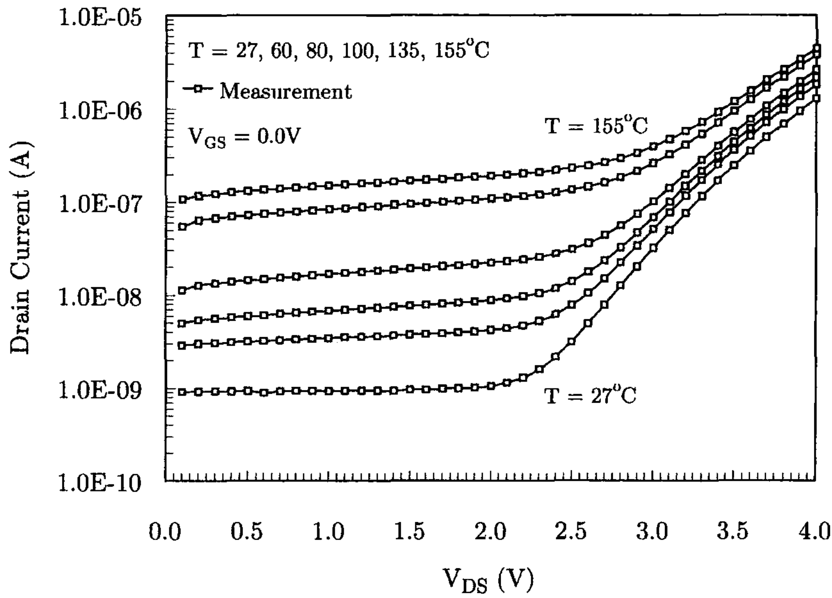


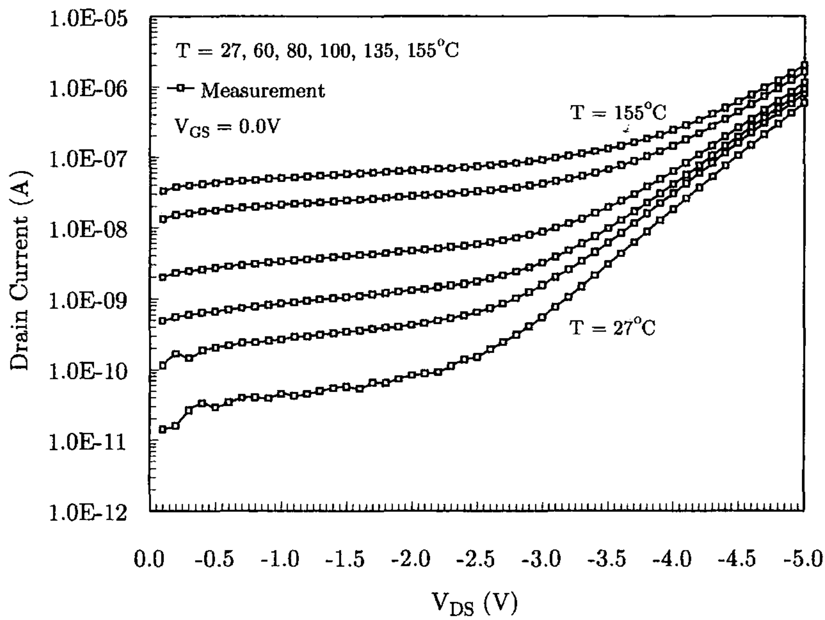
Figure 5.6: Comparison of GIDL model with 1st design

Having considered GIDL at room temperature the next stage of the investigation was to research the influence of temperature. Limited data is available on how temperature affects GIDL current. It was proposed in a report by Slisher *et al* [38] that GIDL is independent of temperature whereas both Lopez *et al* [36] and Bouhdada *et al* [39] presented results showing an increase in GIDL at elevated temperatures. In all cases no physical explanation was provided for why the GIDL mechanism stayed constant or changed at different temperatures.

GIDL measurement wel̈e taken over a range of temperatures using the experimental set-up detailed previously. The measured data is illustrated in figures 5.7(a) and 5.7(b) for n-channel and p-channel devices respectively.



(a) n-channel



(b) p-channel

Figure 5.7: Measured GIDL current at various temperatures

We can see from figure 5.7 that as the temperature is elevated, the GIDL current for both n-channel and p-channel devices increases. For both n- and p-channel devices there is however a different temperature dependence for each of the GIDL regions. In the low field region the leakage is strongly dependent on temperature whereas in the high field region the temperature dependence weakens. The relative difference between each region is illustrated in figure 5.8.

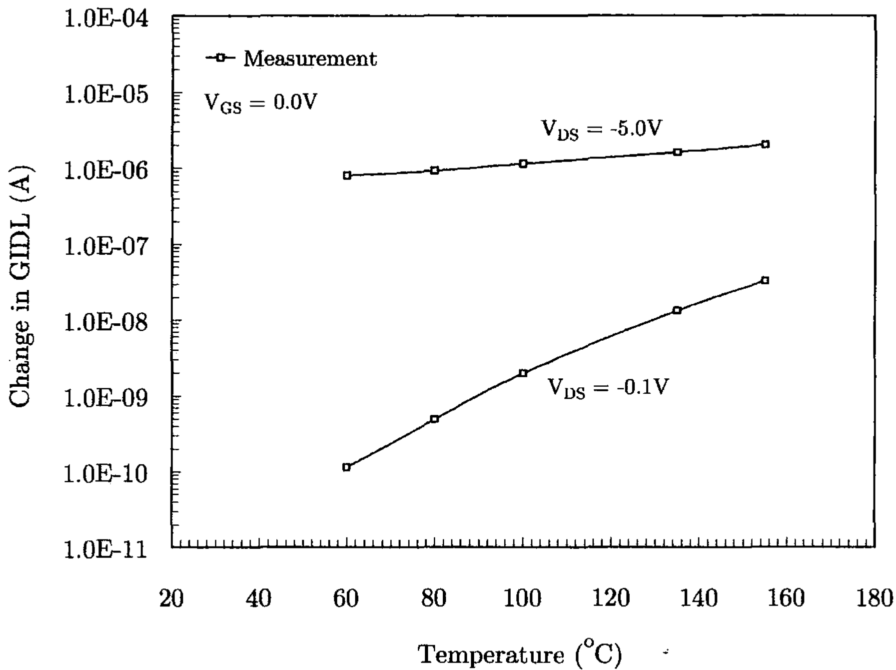


Figure 5.8: Change in GIDL in different operating regions (p-channel)

In the low V_{DS} region leakage is highly sensitive to temperature, increasing exponentially with temperature and $\triangle I_{DS}$ can be attributed to thermal electron-hole pair generation. In the high field region (V_{DS} = -5.0V) $\triangle I_{DS}$ is almost insensitive to temperature since the tunnelling probability and the electric field across the oxide does not strongly depend on temperature. This finding substantiates the hypothesis that the GIDL characteristic is caused by two different conduction mechanisms and supports our decision to model each region separately.

The nature of the model developed allows for easy modification to include temperature dependence. The parameters I_0, I_T, and S_L, S_H were used in the model to define the offset and slope respectively. Therefore to follow the trends illustrated in figure 5.7 requires that these parameters become functions of temperature. The low field expression is then given by:

$$I_{GIDL(L)} = e^{\,(S_L(T)\cdot V_{DS})\ -\ I_0(T)} \tag{5.7}$$

where $I_0(T)$ and $S_L(T)$ are the temperature dependent zero offset and model gradient in the low field region. These parameters were found to fit most accurately to measured data when taking the following form:

$$S_L(T) = (0.0580 \cdot T) - S_L \tag{5.8}$$

$$I_0(T) = (0.0029 \cdot T) - I_0 \tag{5.9}$$

Similarly, in the high field region $I_{GIDL(H)}$ is expressed as:

$$I_{GIDL(H)} = e^{\,[S_H(T)\cdot(-V_{DS}+V_{MAX})]\ -\ I_T(T)} \tag{5.10}$$

where parameters $I_T(T)$ and $S_H(T)$ were expressed as:

$$S_H(T) = (0.009 \cdot T) - S_H \tag{5.11}$$

$$I_T(T) = (0.009 \cdot T) - I_T \tag{5.12}$$

These modifications to the model now allow for the low and high field regions to be predicted at various temperatures. However we can see from figure 5.7 that we must also allow for changes in the transition points at different temperatures. We must also incorporate the change in the transition mid-point if the model is to accurately follow the trends at higher temperatures.

As temperature is increased the transition points were also seen (figure 5.7) to increase. This effect could be modelled simply for the low transition point as:

$$V_{DS(low)}(T) = V_{DS(low)} - 0.19\left(\frac{T}{T_{norm}} - 1\right) \tag{5.13}$$

and for the high transition point as:

$$V_{DS(high)}(T) = V_{DS(high)} - 0.28\left(\frac{T}{T_{norm}} - 1\right) \tag{5.14}$$

where $V_{DS(low)}$ and $V_{DS(high)}$ are the low and high drain transition voltage points at room temperature. The mid-point value used in the transition function (equation 5.17) was expressed as a function of temperature as:

$$G_{MP}(T) = G_{MP} + \left[0.00145\left(\frac{T}{T_{norm}} - 1\right)\right]^{3.5} \tag{5.15}$$

where G_{MP} is the mid-point value at room temperature. The resultant effects on the transition range are illustrated in the figure below.

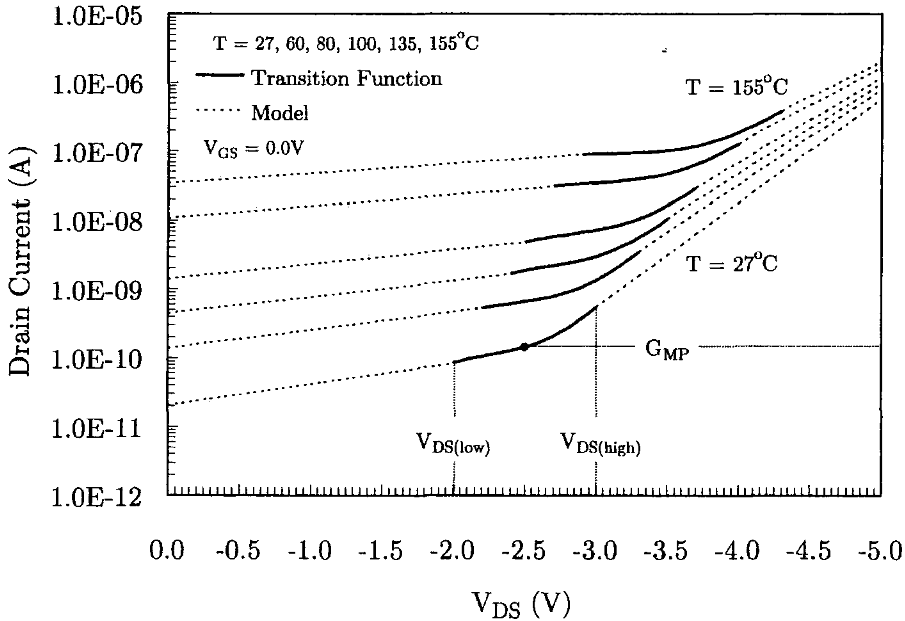
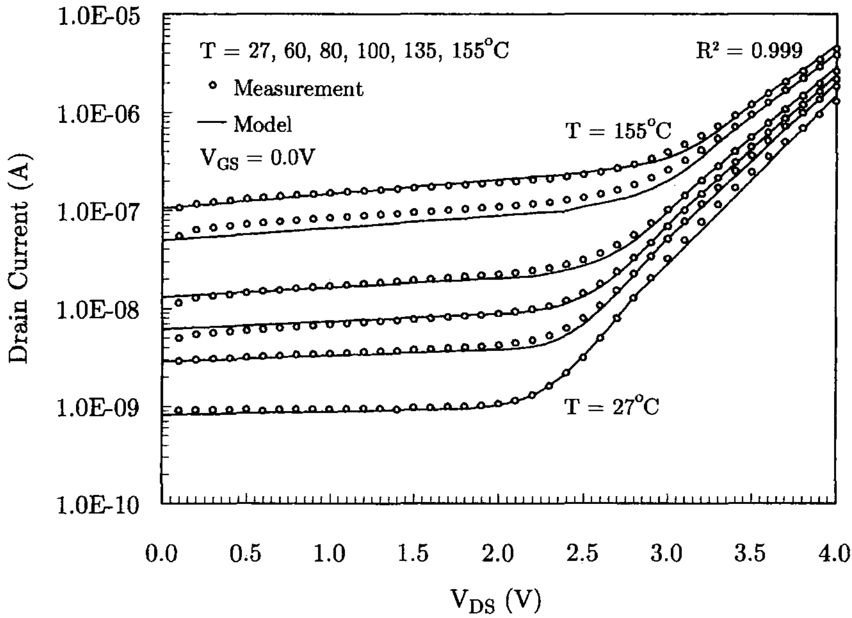
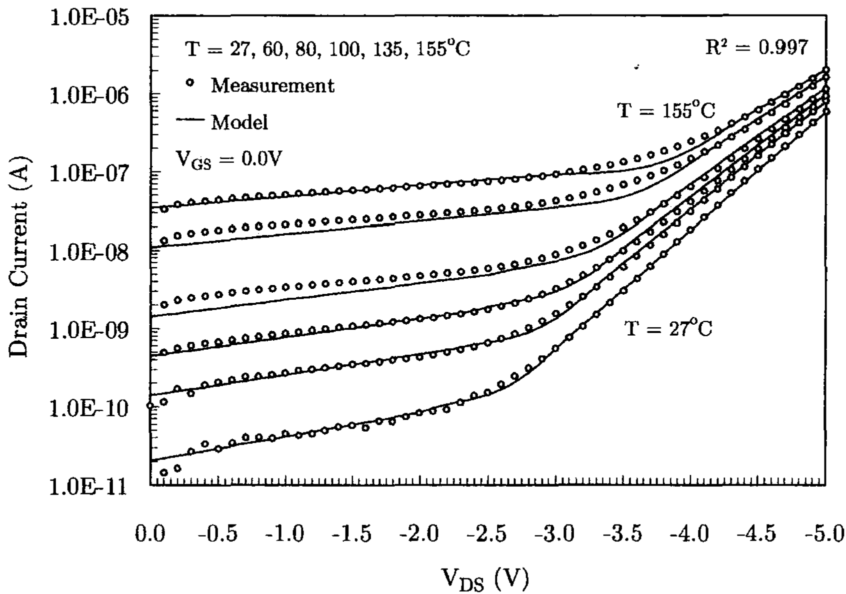


Figure 5.9: Change transition points with temperature (p-channel)

Figure 5.10 shows a comparison between measured data and the model predictions after incorporating the temperature effects for both low and high field regions, along with the changes to the transition function. The calculated results agree well with the measurement data throughout the range of drain voltages for both n-channel and



(a) n-channel



(b) p-channel

Figure 5.10: Model predictions at various temperatures

p-channel devices. Good agreement is seen for low drain voltages, where diffusion is the cause of GIDL current. Predicted tunnelling current for higher drain voltage ($V_{DS} > 3.0V$) is also seen to be accurate. The model predictions also match the increased current at higher temperatures. At low temperatures the transition function follows the measured data well but there is a slight error at higher temperatures, which could possibly be eliminated with further work on defining the transition mid-point function. The advantage of the GIDL expressions in their present state is that with a simple set of measurements the temperature dependence of only a small number of parameters can be extracted.

Model Parameters		n-channel	p-channel
Low field zero offset	I_0 (A)	21.948	26.17
High field offset (@V_{MAX})	I_T (A)	13.67	14.61
Low field slope	S_L	0.02	-0.78
High field slope	S_H	4.20	-3.73
Maximum drain voltage	V_{MAX} (V)	4	5
Low transition point	$V_{DS(low)}$ (V)	1.8	-2.0
High transition point	$V_{DS(high)}$ (V)	2.8	-3.0
Transition mid-point	I_{MP} (A)	5x10^{-10}	7x10^{-11}

Table 5.1: Extracted GIDL model parameters

Table 5.1 shows the eight parameters needed to complete the fitting to measurement data. With a single set of temperature measurements the model allows for easy calibration. Although our model provides an empirical expression for GIDL, we believe it provides a simple and accurate way of taking into consideration the different physical mechanisms behind the effect. Furthermore the model we have developed can be incorporated easily into a larger simulation tool.

5.5 Integration into I-V Model

Having developed the present GIDL model over a range of drain voltages, the final design stage of this work involved integrating it with the complete current-voltage model. Figure 5.11 illustrates the complete gate voltage characteristic of a n-channel device. Only measurement data is plotted.

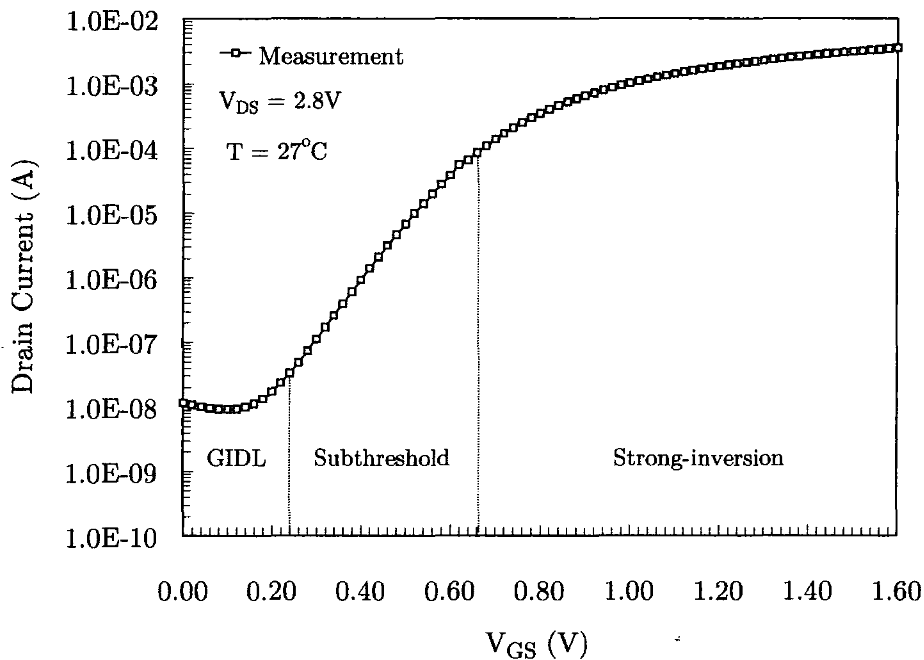


Figure 5.11: Illustration of device operating regions (n-channel)

During the measurement the drain voltage was set above the normal operating voltage (V_{norm} = 1.8V) to enhance the GIDL current. We can see that the total characteristic is divided into three regions. The strong inversion region where drift current contributes to the drain current was described in section 3.3 of chapter 3. Below the threshold voltage the characteristic moves into the subthreshold region. The transition between these two regions was also covered in chapter 3 in section 3.3.7. A means of implementing the GIDL model is now needed.

To do this we used a weighting function similar to that used previously. This will provide a smooth transition between the subthreshold and GIDL model and will allow for relatively easy implementation into the complete simulation tool. Thus the transition between the subthreshold and GIDL model can be defined as:

$$I_{DS(Sub)} = \begin{cases} I_{GIDL} & \text{if } V_{GS} = V_{GS(low)} \\ I_{Tran} & \text{if } V_{GS(low)} > V_{GS} < V_{GS(high)} \\ I_S & \text{if } V_{GS} > V_{GS(high)} \end{cases} \tag{5.16}$$

where the transition function I_{Tran} is expressed as:

$$I_{Tran} = [w_I \cdot I_S] + [(1 - w_I) \cdot I_{GIDL}] + [2\,(1 - w_I) \cdot w_I \cdot I_{MP}] \tag{5.17}$$

The weighting function w_I was defined as:

$$w_I = V_{GS} - V_{GS(low)} \cdot \left(\frac{1}{V_{GS(high)} - V_{GS(low)}} \right) \tag{5.18}$$

The low transition voltage $V_{DS(low)}$ was set to equal 0V. It was found that the high transition point $V_{GS(high)}$ was a function of drain voltage, which we defined as:

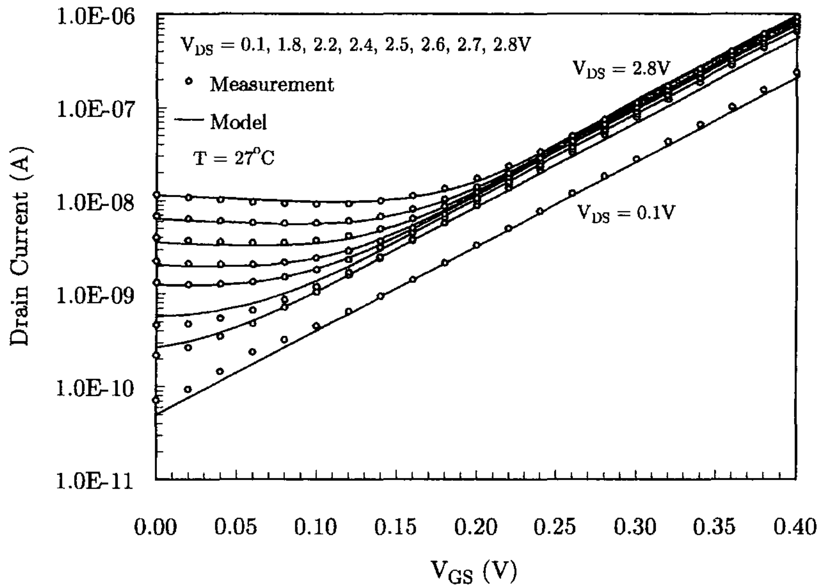
$$V_{GS(high)} = V_{high} + (0.16 \cdot V_{DS}) \tag{5.19}$$

For n-channel and p-channel devices the high transition voltage V_{high} was found to be 0.15 and -0.15V respectively. The mid-point value I_{MP} was found to increase exponentially with drain voltage and was expressed as:

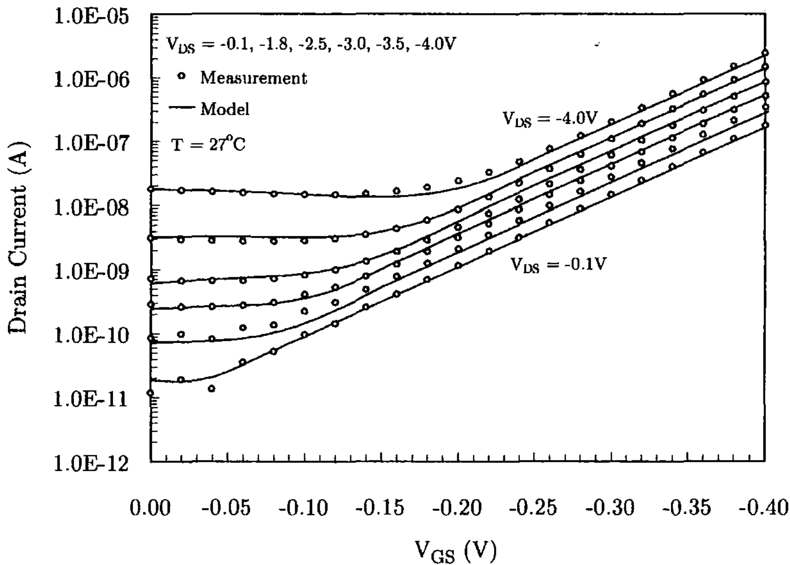
$$I_{MP} = e^{\,(MP_1 \cdot V_{DS}) \; - \; MP_2} \tag{5.20}$$

where MP_1 and MP_2 are fitting constants.

Figures 5.12(a) and 5.12(b) compare the model against measured data in the subthreshold region. In section 3.3.6 we highlighted an error in the subthreshold plot at very low gate voltages. Now with the integration of the GIDL model this error has been removed and good agreement is seen for both n-channel and p-channel devices.



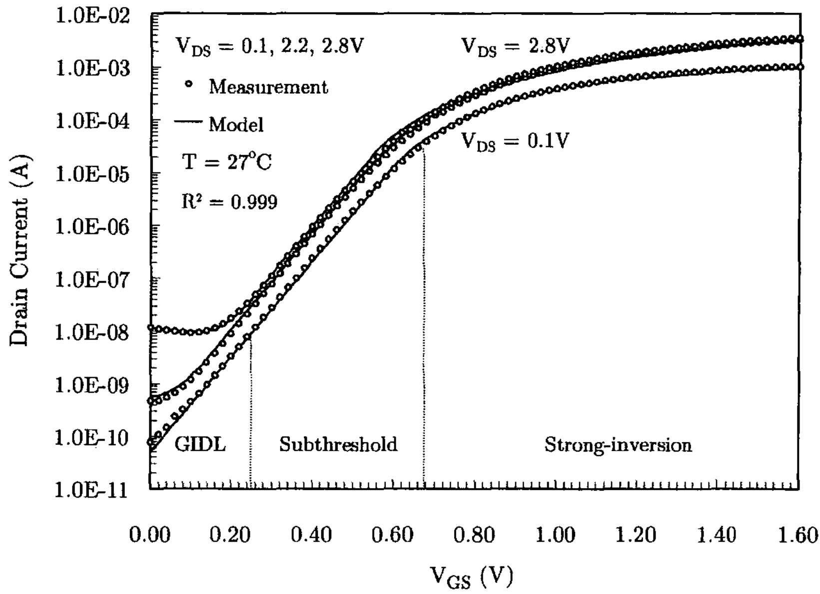
(a) n-channel



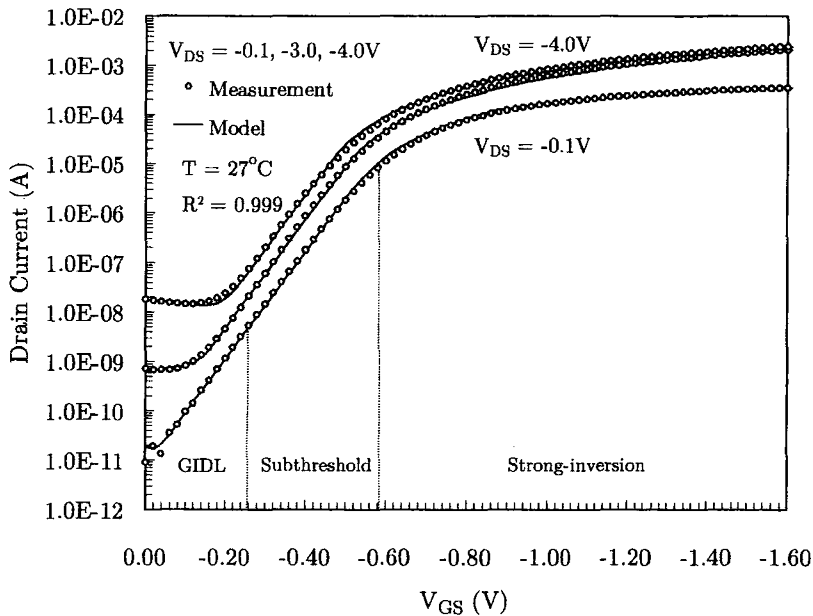
(b) p-channel

Figure 5.12: Subthreshold region at various drain voltages

A smooth transition is seen from the subthreshold to the GIDL model. At higher drain bias we can see the importance of making the high transition point $V_{DS(high)}$ a function of drain voltage. Further illustration of how the GIDL model has been incorporated into the complete I-V characteristic is shown in figure 5.13.



(a) n-channel



(b) p-channel

Figure 5.13: Complete I-V characteristic

As V_{GS} approaches zero we can see that for high drain voltages ($V_{DS} > \pm 1.8$V) there is a significant increase in the OFF-state leakage which is representative of GIDL. The higher drain voltage results in a higher electric field in the gate-to-drain overlap region increasing the chances of tunnelling and hence increased GIDL current.

5.6 Conclusion

A new voltage and temperature dependent model for GIDL has been presented in this chapter. Initially we highlighted the inaccuracies in existing models. We attempted to predict GIDL with a single expression but found that this poorly described the transition between low and high field regions. Separate exponential functions were used to overcome this problem and the influence of temperature was also investigated. In the low field region leakage is highly sensitive to temperature attributed to thermal emission; in the high field region leakage was almost independent of temperature since the tunnelling probability and electric field across the oxide does not strongly depend on temperature. This finding supported the hypothesis that the GIDL characteristic is caused by two different conduction mechanisms and validated our decision to model each region separately.

Investigations into the effect of scaling on the GIDL mechanism would provide for further development of the model. Chung *et al* [40] found that GIDL is enhanced by decreasing oxide thickness. In the same study GIDL was seen to be virtually independent of channel length because the tunnelling depends only on conditions in the immediate gate-to-drain overlap region. However the aim of this work was to develop a simple but accurate model that could be implemented into a I-V model and we believe this has been demonstrated.

Chapter 6

Hot-Carrier Effects

6.1 Introduction

HOT-CARRIER EFFECTS have been a major reliability concern for over 20 years, and a great deal of research has aided the understanding of the mechanism and its effects. Nonetheless, continued device scaling and increases in channel electric fields has required the research community to extend the understanding of hot-carrier effects in deep sub-micron devices.

In this chapter we examine the influence of temperature on hot-carrier effects for deep sub-micron devices. For many years it has been believed that hot-carrier degradation is reduced at elevated temperatures, and for this reason research on this topic has been minimal. We aim to further the understanding of temperature on hot-carrier effects and present new results to clarify if this belief is correct for today's technologies. We propose a new temperature dependent substrate current model, and demonstrate how it can be used to predict device lifetime. The model is supported by experimental results at each stage of its development.

6.2 Mechanism Overview

Hot-carrier effects are the result of high lateral electric fields in the channel near the drain region. This is the main reason why aggressive scaling of devices has resulted in an increase of hot-carrier degradation. While higher electric fields provide for increased carrier velocity, and hence switching speed, these also present a serious concern for the long term reliability of the devices.

There has been substantial research into hot-carriers effects and the cause of the degradation [57, 58, 59]. A strict consensus exists that under high electric fields carriers can gain enough energy to cause impact ionisation. This results in the generation of electron-hole pairs causing an increased substrate current. Moreover if the electric field is high enough, carriers gain sufficient energy to be injected and trapped in the oxide generating interface states and oxide charges.

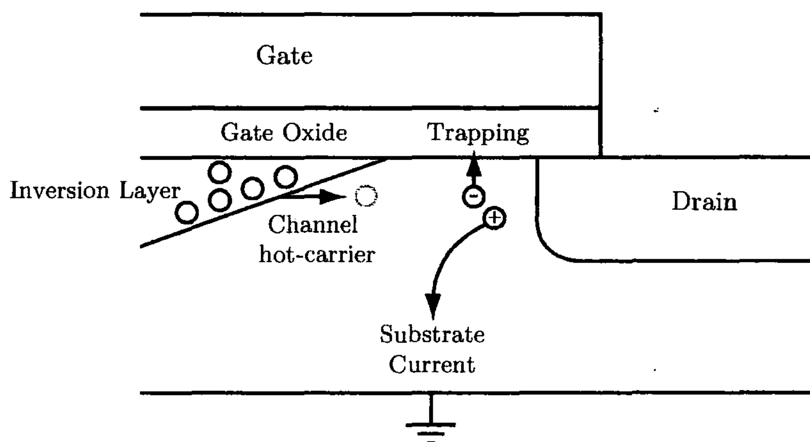


Figure 6.1: Illustration of hot-carrier mechanism (n-channel)

Figure 6.1 illustrates the hot-carrier mechanism in an n-channel device. Under high drain voltages, a pinch-off region is formed between the inversion layer and drain. The large voltage drop across this region causes a high lateral electric field close to

the drain junction. This accelerates carriers travelling from the source such that they reach energies which are considerably higher than the thermal equilibrium energy in the silicon lattice. This is why the carriers are referred to as hot-carriers. They loose their energy via impact ionisation generating electron-hole pairs. The majority of the carriers generated are swept to the substrate resulting in substrate currents. Hence the substrate current serves as an excellent measure of hot-carrier degradation. Any carriers generated with energies large enough to overcome the potential barrier between the Si/SiO_2 interface can be injected into the gate oxide.

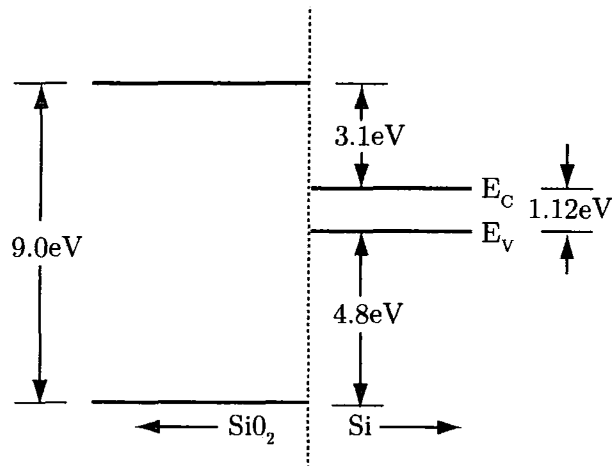


Figure 6.2: Illustration of silicon-oxide interface energy bands

As we can see from the above figure, silicon in conventional devices has a band-gap of approximately 1.12eV and silicon dioxide has a band-gap of approximately 9.0eV. The band alignment at the silicon-oxide interface results in an energy barrier of $\approx$ 3.1eV for electrons and $\approx$ 4.8eV for holes. Due to the energy barrier for electrons being smaller than that of holes, the probability of electron injection is greater than that of hole injection. This is why hot-carrier degradation is different in n-channel than in p-channel devices. In both types of devices any injected carriers cause a change in the oxide charge distribution, and this acts as the main mechanism for degradation of device parameters. Increases in threshold voltage subsequently lead

to a transconductance degradation, and decrease in drain current drive capability. An example of parameter degradation is shown in figure 6.3. It shows a comparison between the transconductance of a n-channel device before and after hot-carrier stress. Measurement data only is shown in the figure.

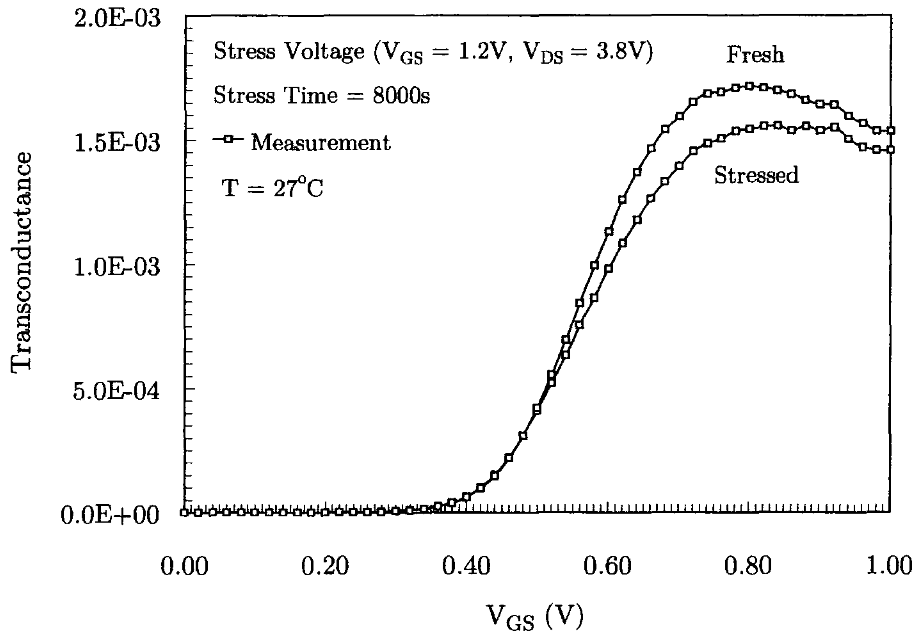


Figure 6.3: Transconductance degradation after hot-carrier stress

After stressing, there is a significant decrease in the transconductance of the device. This will result in a reduction in device operating speed and can have detrimental effects on circuit functionality. This experiment was conducted at room temperature and until now the effects of high temperatures on the hot-carrier mechanism has seen limited research. From the literature, it is believed that at higher temperatures impact-ionisation is decreased and hence hot-carrier degradation is reduced [67]. Recently though there has been some concern that this assumption may not be true for deep sub-micron devices [68]. Hence a part of this work aims to further the understanding of the hot-carrier mechanism, clarifying its temperature dependence.

6.3 Experimental Method

Hot-carrier stress was applied to single n-channel and p-channel devices for different periods of time, and measurement were taken at periodic time intervals. Each test was automated using a LabVIEW™ program. All test measurements followed the JEDEC procedure for hot-carrier measurements [103]. The drain current characteristics were utilised to provide information about the degradation process during hot-carrier stress. The threshold voltage was measured using the constant current method describes in section 3.3.1 of Chapter 3. Additionally for hot-carrier analysis substrate current measurements were taken.

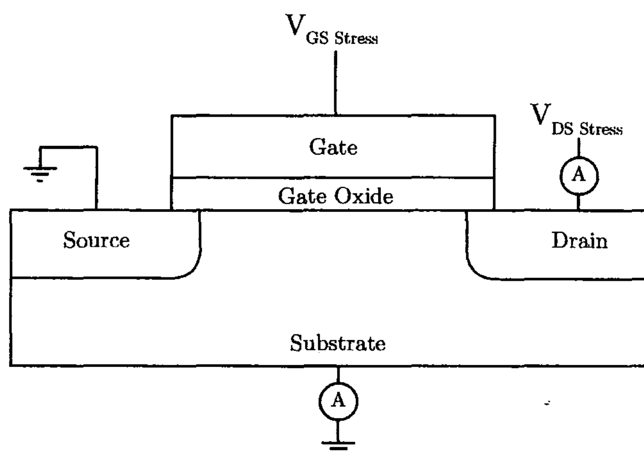


Figure 6.4: Hot-carrier stress configuration

Figure 6.4 illustrates the test configuration used throughout hot-carrier tests. Before stressing, characteristic measurements were taken to determine the gate bias at which maximum substrate current occurs. Stresses were applied at $I_{Sub(max)}$ and various drain voltages above the normal operating voltage (1.8V) to accelerate the degradation. Stress tests were performed at various temperatures in the range of 27-155°C. Hot-carrier lifetime was defined as the stress time required to produce a 10% shift in the saturation drain current from its fresh value.

6.4 Hot-Carrier Model

It is the high electric field at the drain end of the channel that causes impact ionisation, generation of electron-hole pairs and hence device degradation. However, it is difficult to measure the peak electric field and verify it experimentally. The substrate current I_{Sub} which is caused by hot-carriers generated by the same electric field is a simple measurable quantity that can be correlated with device performance and can be used to monitor device degradation. Hence the majority of models developed to predict hot-carrier lifetime are based on substrate current generation [62, 63]. It is therefore very important to have accurate substrate models to provide a good estimate of individual hot-carrier device lifetime. In this study we improve an existing substrate current model to accurately predict the substrate current for deep sub-micron devices for a range of voltages. We also develop the model further to encompass temperature dependence.

Within the literature several analytical models for substrate current behaviour have been reported [57, 62, 63]. However there is concern over the accuracy of these models for today's deep sub-micron devices as substrate current has been expected to increase dramatically with reductions in device dimensions. Our work initially used the following expression for substrate current [57].

$$I_{Sub} = I_{DS} \cdot \frac{A_i \lambda_m}{\phi_i} \cdot (V_{DS} - V_{DSsat}) \cdot \exp\left(\frac{-\phi_i}{\lambda_m E_M}\right) \qquad (6.1)$$

where $\phi_i = 1.3$eV is the critical energy for a hot-carrier to cause impact ionisation, A_i a technology dependent constant, and λ_m is the electron-phonon scattering mean free path. E_M is the peak electric field near the drain region and can be approximated using two-dimensional analysis as $E_M = (V_{DS} - V_{DSsat})/l_d$. The saturation voltage V_{DSsat} is the potential at the pinch-off point defined by equation 3.18 in chapter 3.

The characteristic length l_d is the effective length of the channel pinch-off region and is a function of oxide thickness t_{ox}, channel length L, and drain junction depth X_J. For the devices used in this study X_J was approximated as $0.1\mu m$. For deep sub-micron MOSFET's with thin gate oxide, l_d has been formulated as [67]:

$$l_d = 0.017 \cdot {t_{ox}}^{\frac{1}{8}} \cdot {X_J}^{\frac{1}{3}} \cdot L^{\frac{1}{5}} \tag{6.2}$$

From equation 6.1 it is expected that the substrate current is proportional to the drain current and the probability for impact ionisation. Figure 6.5 shows a plot of I_{Sub} versus V_{GS} at varying drain voltages comparing measurements and model predictions.

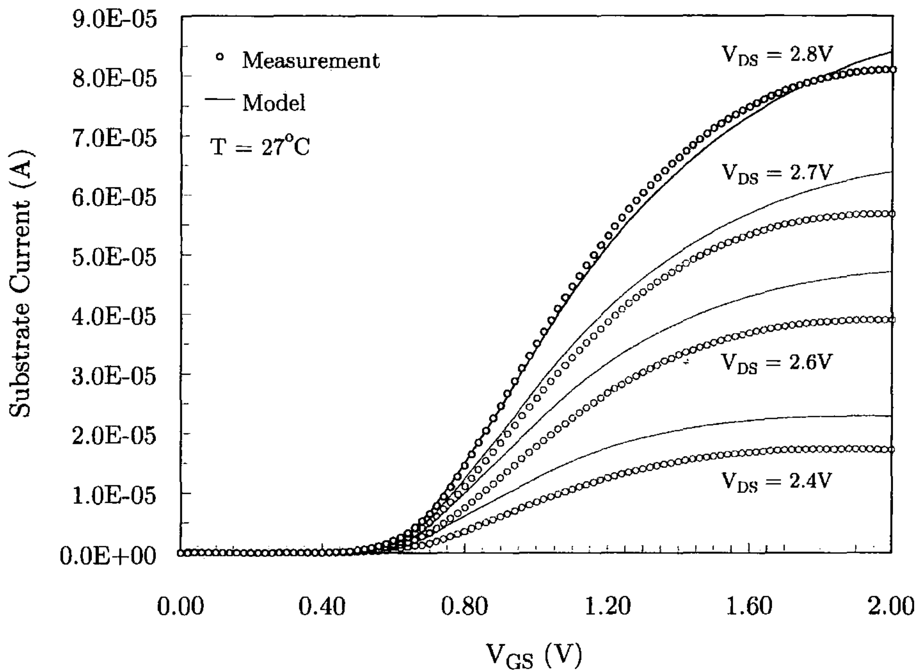


Figure 6.5: Inaccuracy in substrate current prediction (n-channel)

As the gate voltage increases from zero the supply of channel carriers increases resulting in an increase in the I_{Sub}. However, as the gate bias approaches the drain voltage, the lateral electric field in the channel decreases resulting in a decrease in

substrate current. Godd agreement can be seen in figure 6.5 between model and device measurement at $V_{DS} = 2.8$V which was the initial condition used for model calibration. However despite trying to optimise the parameters used in equation 6.1 the drain voltage dependency of the substrate current was found to be inaccurate for the deep sub-micron devices used in our tests. We therefore propose a simple way to develop the existing substrate current expression.

$$A_i = A_0 \cdot \left(C_A \cdot {V_{DS}}^{1.9}\right) \tag{6.3}$$

Defining the drain voltage dependency of the technology constant A_i, with fitting parameters A_0 and C_A, the drain voltage dependency of I_{Sub} can be calibrated for different technologies. With the incorporation of equation 6.3 into the substrate current expression, figure 6.6 illustrates the improvement in accuracy.

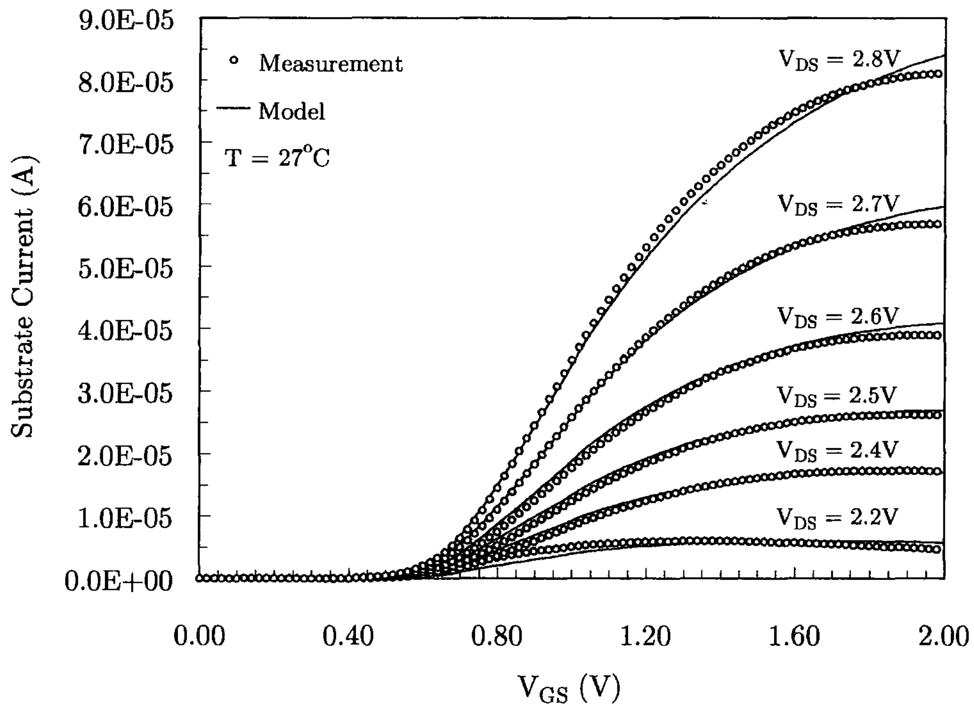


Figure 6.6: Substrate current versus gate voltage (n-channel)

Similarly the model was used to predict I_{Sub} in p-channel devices as shown below in figure 6.7. Good agreement can be seen and also shown is the common bell shaped characteristic resulting from the relative variation between carrier supply and the lateral electric field. Furthermore the refinement of the drain current model in the subthreshold region provides for better I_{Sub} accuracy in the low gate voltage range.

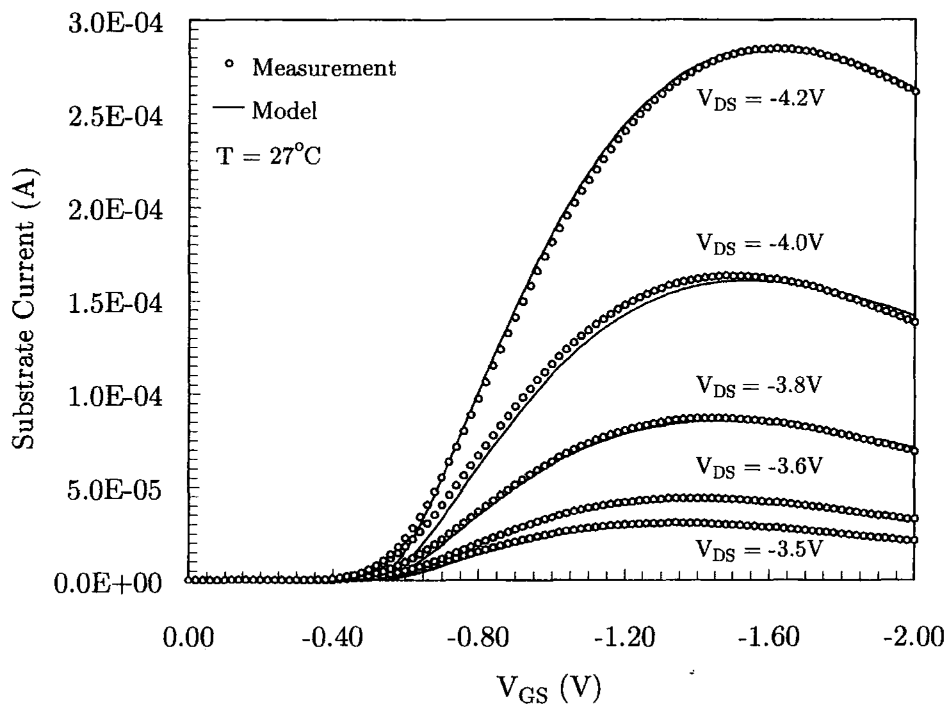


Figure 6.7: Substrate current versus gate voltage (p-channel)

Most studies concentrate on hot-carrier effects in n-channel devices only, but we conducted studies of both n-channel and p-channel devices. The reason why hot-carrier effects in p-channel devices are often neglected is because the mean free path of the majority carriers is shorter in p-channel type devices. This means hot-carrier induced problems are more severe in n-channel devices due to the longer mean free path and hence higher energy of electrons. However concerns have arisen over the hot-carrier reliability of p-channel devices in deep sub-micron devices [69].

Figure 6.8 compares the relative substrate currents for both n-channel and p-channel devices. We can see that the peak substrate current in n-channel devices is about 2 to 3 orders of magnitude larger than in p-channel devices.

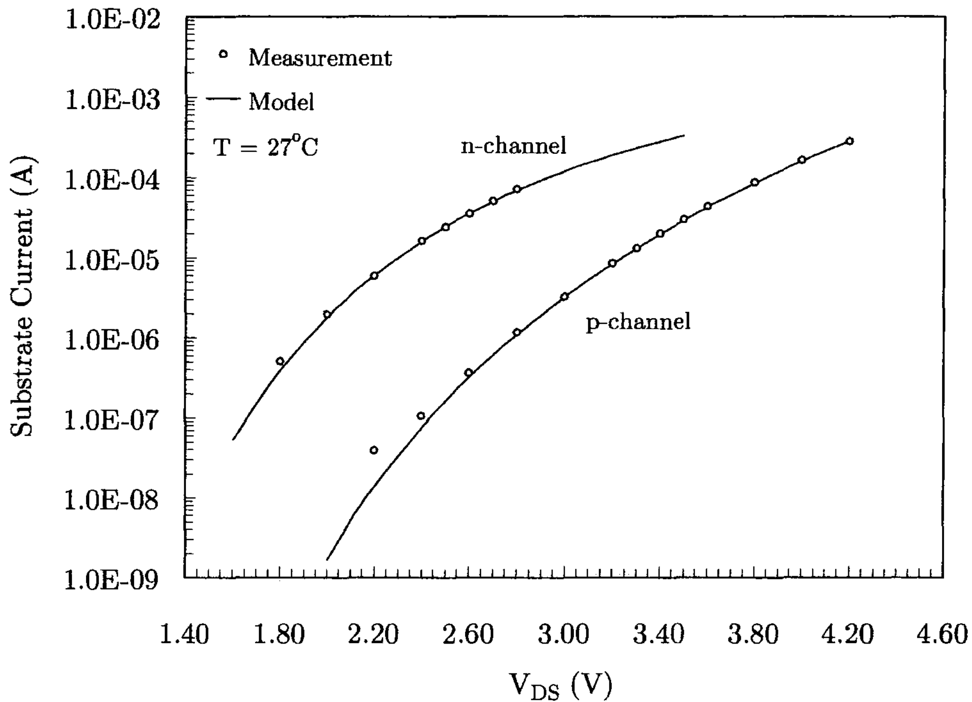


Figure 6.8: Comparison between n- and p-channel substrate currents

Under the same bias conditions there are fewer hot-carriers and hence lower substrate current for p-channel devices compared to that of n-channel devices. Hence in a CMOS inverter it would be expected that the n-channel device would undergo greater stress and fail before the p-channel device.

In this section we have developed a semi-empirical model that captures the basic physical theory of hot-carrier effects, but whose parameters are based on measurable quantities such as I_{Sub}. The next stage in the study investigated how temperature influenced the substrate current and subsequently hot-carrier degradation.

6.5 Influence of Temperature

Until recently it has been assumed that at higher temperature hot-carrier effects are reduced. For this reason the majority of hot-carrier research has concentrated on the influence of increased electric field and voltage dependency at room temperature. However conflicting results on I_{Sub} temperature dependence have been recently reported [67, 68]. In this section we aim to accurately establish the temperature dependence of substrate current for deep sub-micron devices and extend the room temperature model to encompass temperature effects.

Figure 6.9 shows the temperature dependence of the substrate current for a p-channel device. We can see that at higher temperatures the substrate current is reduced. The drain voltage during this experiment was relatively high at -4.2V.

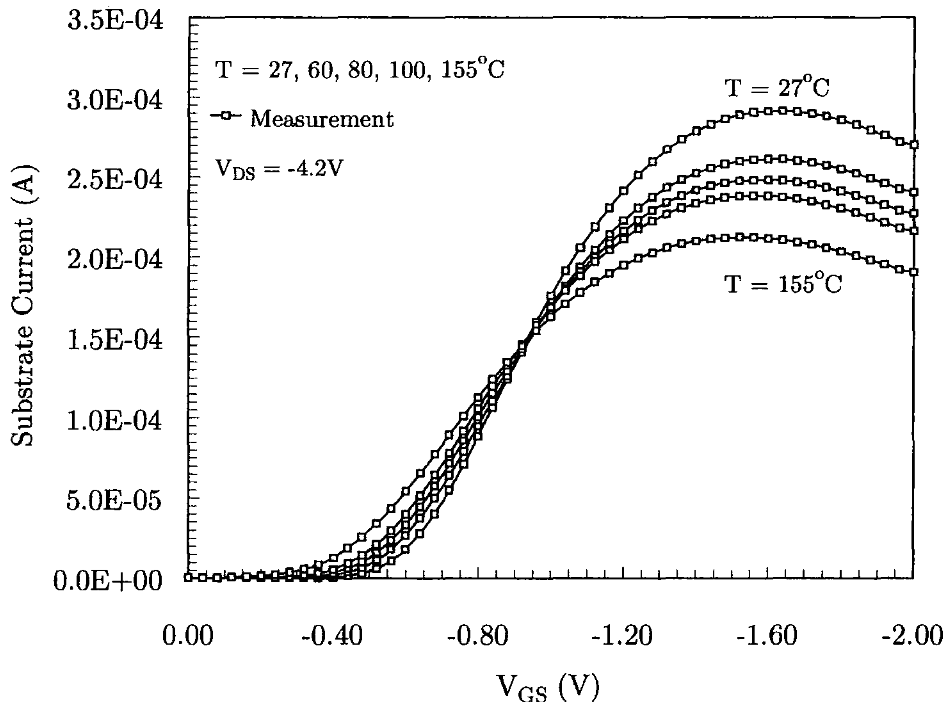


Figure 6.9: Substrate current at various temperatures (high V_{DS})

A comparable trend was also found for n-channel devices, with I_{Sub} decreasing at elevated temperatures at high drain voltages. The reasoning behind the decrease in I_{Sub} is as the thermal energy increases at higher temperatures, the Si lattice vibrates more, and hence the electron-phonon scattering is increased. As a result the hot-carrier mean free path is reduced and the electrons have less time to gain energy. Subsequently impact ionisation and electron-hole pair generation both reduce.

From the trend shown in figure 6.9 it could be assumed that substrate current is reduced as higher temperatures and hot-carrier effects reduced. This is the assumption that has been made in the past. We proposed though that there is a transition in the temperature dependence of I_{Sub} and it does not always decrease at elevated temperatures. Consider figure 6.10 illustrating the temperature dependence of a p-channel device at V_{DS} = -2.6V.

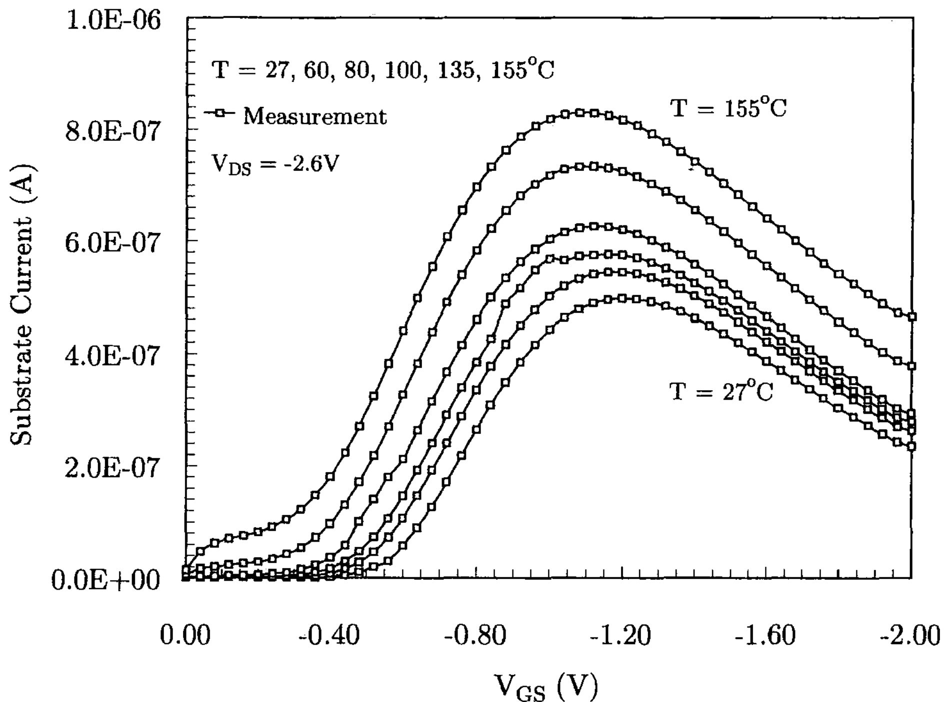


Figure 6.10: Substrate current at various temperatures (low V_{DS})

Clearly at lower drain voltages the substrate current increases with elevated temperatures, contradicting the common belief. This finding prompted further investigations. Experiments were conducted to study the temperature dependance of I_{Sub} at various drain voltages. The results are shown in figure 6.11.

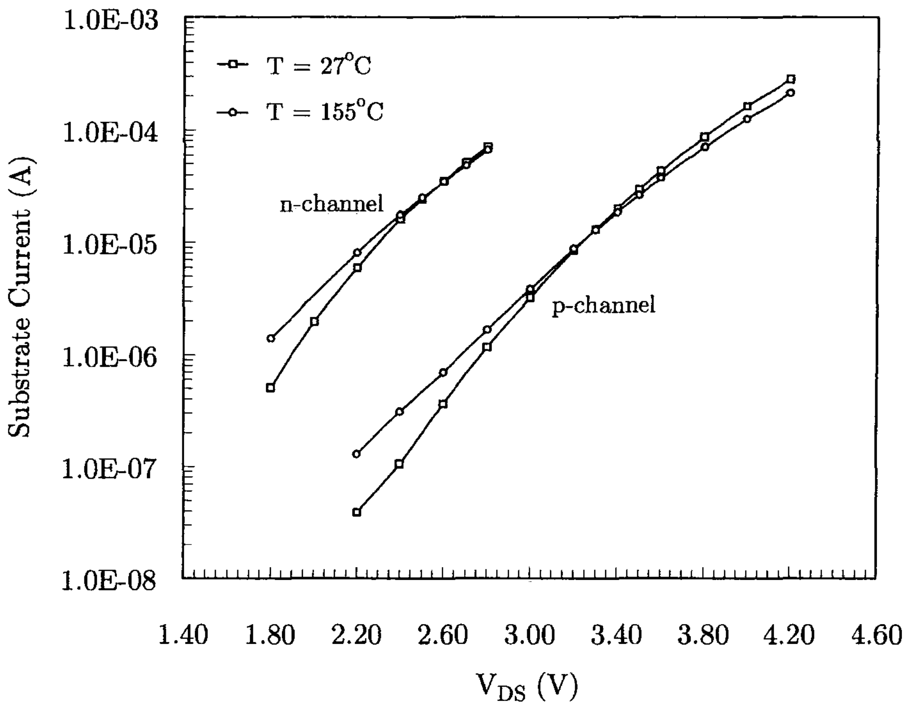


Figure 6.11: Substrate current crossover point

From figure 6.11 it can be seen that there is a transition point in the temperature dependence of the substrate current at a certain drain voltage. This was found to be 2.6V and -3.3V for n-channel and p-channel devices respectively. The measurements were taken at the maximum substrate current values.

To understand this let us first consider the high field region where I_{Sub} decreases with temperature. The dominant cause of the decrease in I_{Sub} is the reduction in electron mean free path λ_m. This reduces the mobility of the hot-carriers and results in a decrease of the impact ionisation rate. At the same time the electron relaxation

length λ_e also decreases at higher temperatures. This is the average distance over which an electron will lose its energy after the impact ionisation event has occurred. At high drain voltages the temperature dependence of λ_e is overshadowed by the reduction in λ_m. However at lower drain voltages its effect is not negligible. In this region the decrease in λ_e with temperature causes an increase in the effective electric field which dominates over the relative reduction in electron-phonon mean free path. Thus the increase in electric field at higher temperatures results in an increased substrate current and subsequently the cause of the transitions seen in figure 6.11. This effect has not been noticed in previous long channel technologies because the increase in electric field at lower drain voltages and elevated temperatures was not sufficiently significant to dominate over the decrease in mean free path. Therefore it has not been included on any previous substrate current models making them inaccurate for deep sub-micron devices at low V_{DS}.

To extend the room temperature substrate current model previously described in section 6.4 we propose the following expression for the temperature dependence of the electron-phonon mean free path λ_m.

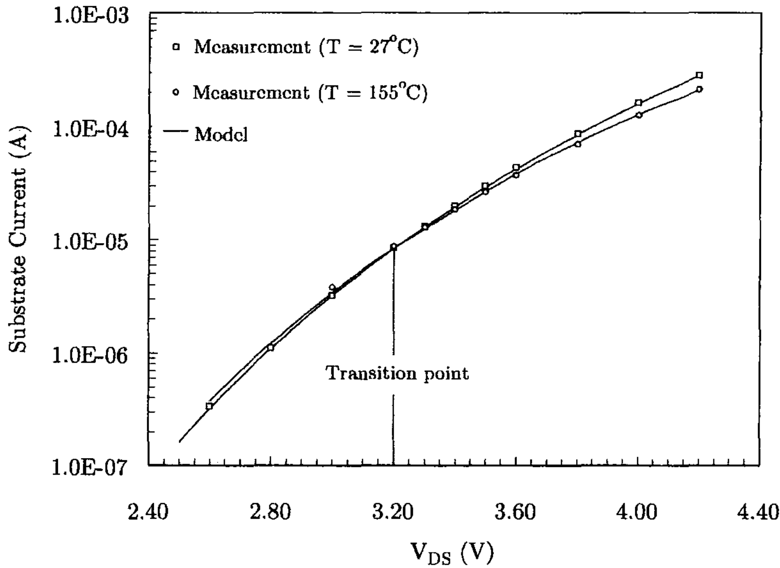
$$\lambda_m = \lambda_{m0} + \left[TP \cdot \left(\frac{T}{T_{norm}} - 1\right)\right] \tag{6.4}$$

where λ_{m0} is the mean free path at room temperature. The transition point is implemented into the model by the parameter TP in equation 6.4. This is a function of drain voltage and is expressed as:

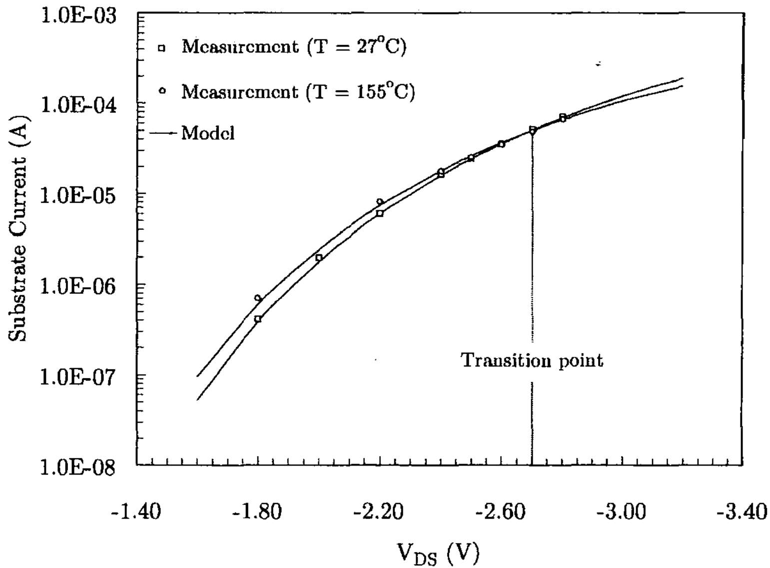
$$TP = TP0 - (V_{DS} \cdot C_T) \tag{6.5}$$

where $TP0$ and C_T are fitting constants that determine the transition voltage. To the author's knowledge this transition effect has not been previously modelled.

Figure 6.12 shows the drain voltage dependency of I_{Sub} at different temperatures. Having incorporated the transition effect into the substrate current model there is good agreement between model and measurement data for a range of voltages with the transition point accurately predicted for both n-channel and p-channel devices.



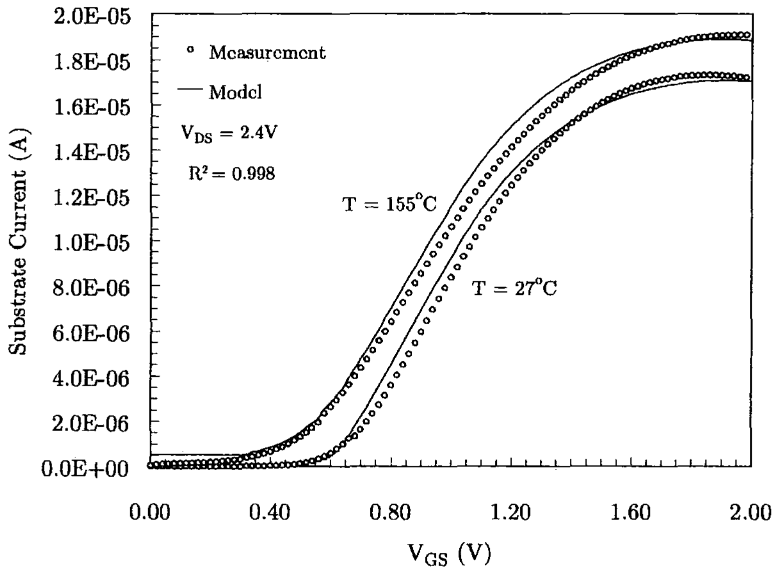
(a) n-channel



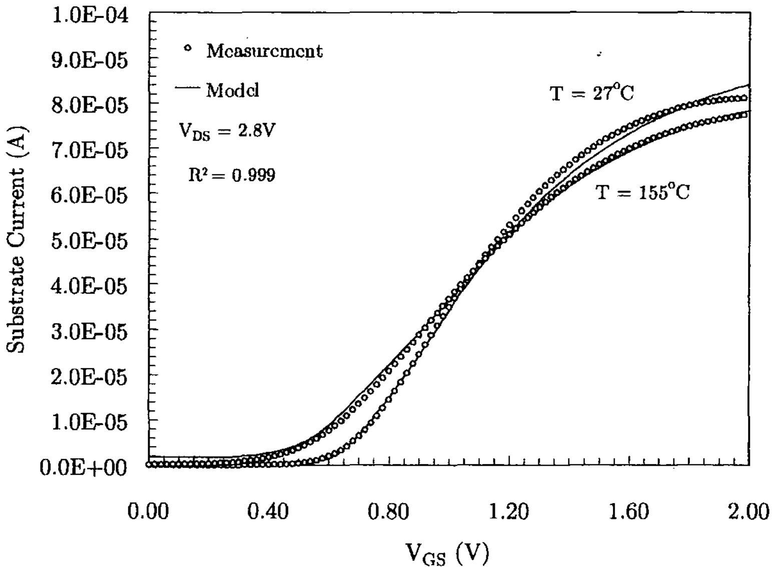
(b) p-channel

Figure 6.12: Model predictions for substrate current transition

To further illustrate the models capability to simulate the transition, figure 6.13 compares the temperature dependence of I_{Sub} below and above the transition point for an n-channel device. In figure 6.13(a) the substrate current increases at higher temperatures, whereas the opposite is true in figure 6.13(b). Model predictions agree well with measured data.



(a) Below transition point (V_{DS} = 2.4V)



(b) Above transition point (V_{DS} = 2.8V)

Figure 6.13: Substrate current characteristic above and below transition

6.6 Lifetime Prediction

Having extended the substrate current model to accurately encompass the transition in I_{Sub} temperature dependence, this section investigates whether this new finding has any effect on device lifetime. This is particularly important to clarify previous assumptions that substrate current and hence hot-carrier degradation is strictly reduced at elevated temperatures.

$$\tau = A \cdot \exp\left(\frac{B}{V_{DS}}\right) \tag{6.6}$$

Equation 6.6 is a hot-carrier lifetime model in its simplest dimensionless form. The lifetime τ is calculated based on a drain voltage acceleration law, where A and B are fitting constants. Despite providing reasonable accuracy and a very simple model, it is not based on the physical process behind hot-carrier degradation and is unsuitable for the aims of our study. A model based on the foundations of hot-carrier degradation and capable of encompassing different technologies is needed. We have already seen that the substrate current is a good indication of the level of hot-carrier stress, and hence it can be used as a measure of device degradation. Previously a number of successful lifetime models have been developed that characterise the hot-carrier effect with peak substrate current [57, 65, 66]. These semi-empirical models were shown to be valid only down to 0.25μm technology [66]. However for smaller technologies as used in this study it needs to be confirmed if they are still valid.

$$\tau = A_{HCE} \cdot \left[\left(\frac{I_{Sub}}{I_{DS}}\right)^{m}\right] \tag{6.7}$$

The above equation is a widely used hot-carrier lifetime model for sub-micron devices, where A_{HCE} and m are constants [58]. It relates device lifetime to the normalised current (I_{Sub} / I_{DS}). The model is both temperature and dimensionally dependent with the use of I_{DS} and I_{Sub}.

Using equation 6.7, figure 6.14 shows device lifetime as a function of drain voltage for an n-channel device. The main parameter degradation caused by hot-carrier effects is an increase in V_{th} and decrease in I_{DS}, hence the measurement data was obtained by stressing the device at some fixed V_{DS} and monitoring the change in saturation drain current. Device lifetime was defined at $\triangle I_{DS} = 10\%$. The gate voltage was biased at $I_{Sub(max)}$ and the complete test repeated at an elevated temperature. Measurements were obtained at the lowest V_{DS} stress possible to provide high confidence of the extrapolated lifetime, however a compromise was needed to make the test time practical.

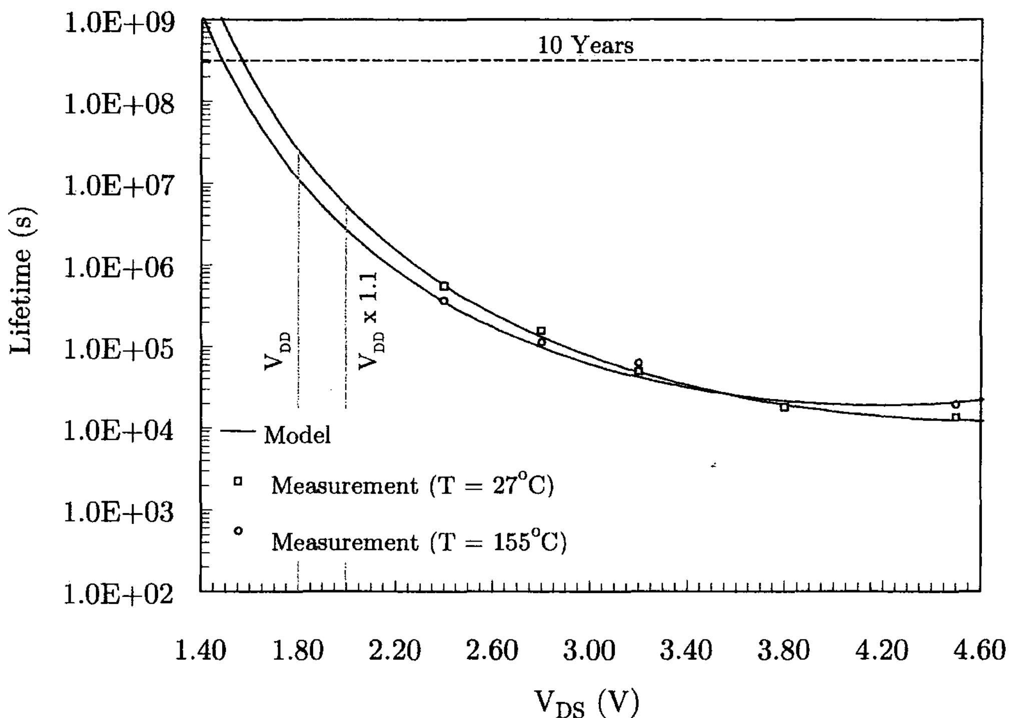


Figure 6.14: Hot-carrier lifetime (n-channel)

The measurement data captures the essential device degradation information and there is good correlation with the model predictions. The model predicts a hot-carrier lifetime at normal operation conditions (V_{DD} x 1.1 for safeguarding) of 70 days at 27°C and 34 days at 155°C. This is a DC stress worst case lifetime. According

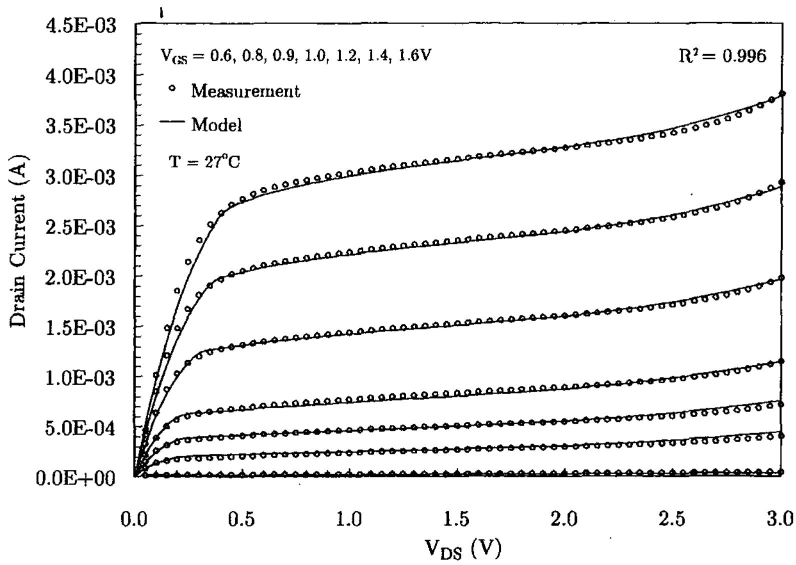
to Hu *et al* [104] the AC lifetime can be calculated using the ratio $\tau_{AC} = \tau_{DC} \cdot \frac{4}{f \cdot t_r}$ where f is the switching frequency and t_r is the gate signal rise time. Hence for the test results in figure 6.14 the AC lifetime (f = 100MHz, t_r = 1ns) at 27°C is 8.2 years and 4.1 years at 155°C. A quality and reliability report by Winbond manufacturing [105], specified hot-carrier device lifetime for their DRAM devices at DC > 0.2 year and AC > 10 year. The DC lifetime of our devices at room temperature are above the required 0.2 year period. However at elevated temperatures they are below. We propose that this is because of the transition of I_{Sub} temperature dependence at low V_{DS} presented in the previous section. At high V_{DS} the belief for long channel devices that hot-carrier effects are reduced at higher temperatures is true, but this is not the case at lower V_{DS} near real life operating conditions. This raises doubts over the validity of current hot-carrier models and their capability of predicting the influence of temperature.

6.7 Integration into the I-V Model

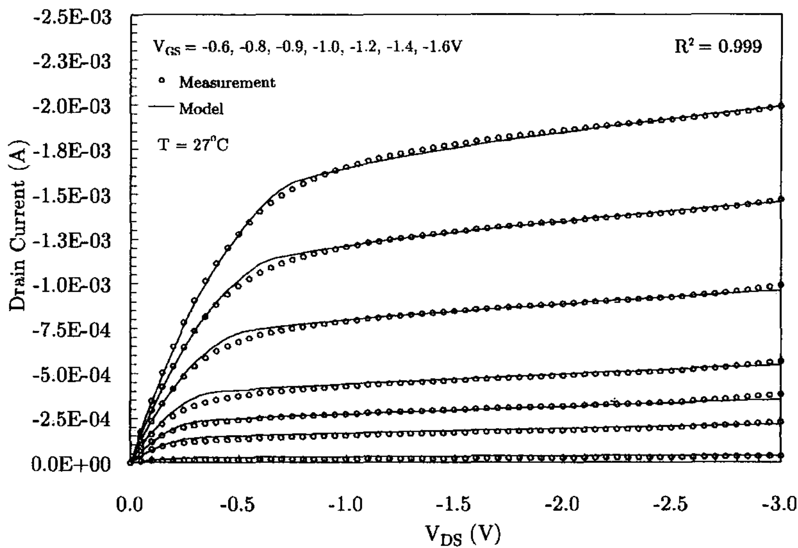
The final part of this chapter describes how the substrate current is implemented in the complete I-V model, and the resultant effect on the I_{DS} characteristic. In figure 3.8 of chapter 3 an inaccuracy was highlighted at high drain voltages. It was particularly evident in n-channel devices compared to p-channel. It was proposed that the deviation of the measurement data from the model prediction was due to the onset of impact ionisation and increase in drain current as a result.

$$I_{DSS} = I_{DS} + I_{Sub} \tag{6.8}$$

As we have seen at high drain voltages a significant substrate current is produced. Hence to complete the I-V characteristic a simple addition of the drain current and substrate current is all that is required.



(a) n-channel



(b) p-channel

Figure 6.15: Complete I-V characteristic

From figure 6.15(a) the inaccuracy in the n-channel drain current model has now been removed. The R^2 value has risen from 0.985 to 0.996. Whereas previously the model failed to follow the data at high drain voltage ($V_{DS} > 2.0$V), now the added substrate current is evident in the model prediction. The substrate current was also implemented into the p-channel model, however due to the lower level of substrate current it had no effect to the I-V characteristic in the operating range of interest.

6.8 Conclusion

A clear understanding of hot-carrier effects in a high temperature environment is essential to ensure product reliability. In this chapter we have presented results that further this understanding. It was identified that at low drain bias the substrate current for both n-channel and p-channel devices increased at elevated temperatures. This is believed to be due to the reduction of the electron relaxation length λ_e resulting in increased electric field, and hence increased substrate current. This is contrary to common belief that hot-carrier effects are strictly reduced at elevated temperatures. This helps to support a number of concerns risen in recent literature that the common belief is no longer true for deep sub-micron devices.

A new semi-empirical substrate current model was developed to account for the transition in temperature dependence of I_{Sub} at low drain voltages, and good agreement was observed between model predictions and measurement data. Despite most studies concentrating only on n-channel devices, we investigated and developed a model for both types of devices to provide a complete hot-carrier analysis.

These new findings cast some doubt on the understanding of the hot-carrier phenomena in deep sub-micron devices and are particularly disconcerting to those interested in high temperature operating conditions. If temperature effects are not considered the lifetime of devices could be overestimated, and is an obvious concern for future technologies.

Chapter 7

An Integrated Reliability Methodology

7.1 Introduction

IN THE PREVIOUS CHAPTERS separate models were developed for the operating characteristics of devices and also the prominent technological life-limiting failure mechanisms. Each operating characteristic/mechanism has been investigated for a given technology and the developed model verified experimentally. In this chapter we present a new methodology to combine these models into a complete framework.

A simulation tool is designed that can be implemented into the product design process to allow quick evaluation of device performance and reliability. The tool is designed around the current-voltage model and the failure models are provided as add-ons for device failure analysis. Simulations are undertaken to investigate how the combination of elevated temperatures and changes in device dimensions may affect the performance and reliability of future technologies.

7.2 Methodology

Behind the design of the simulation tool is a methodology: an idea or ideal of what the basis for the tool is, and how it fits into the present product development and reliability qualification process. In short our methodology is to provide a reliability tool that is based on physical models and can be implemented in the early product development stages.

Traditionally reliability assurance has relied mainly on failure detection at the end of a lengthy product development and qualification process. Due to the reliability analysis taking place towards the end of the development cycle any re-design work adds significantly to design time and cost. For today's technologies, time-to-market requirements make it no longer practical to design the product and then quantify and refine the reliability by a sequence of test and fix steps at the latter stages of development. What is required is a quick and effective way of evaluating product reliability at the earliest stages of conceptual design. In practice a designer will make a number of changes/optimisations to his or her design and it would be an oversight to neglect simulation before committing the design and waiting to discover if there are any failures or errors. Thus, the provision of a simulation tool at this stage of development would prevent the nervous wait to see if the performance of the circuit meets specification. Furthermore it would allow the user to optimise design performance and analyse the effect of any changes.

On the basis that a tool is required at the early stages of product development, the methodology we develop aims to provide an element of built-in reliability to the design process. To develop some means of simulating device reliability with simple but effective models, to replace the traditional end of the line reliability testing.

Figure 7.1 illustrates how it is envisaged that the simulation tool fits into the integrated reliability methodology. It can be seen that the reliability analysis is built into the early stages on the development process before any fabrication or material cost. The core elements of the tool are the I-V and failure mechanism models that take the user specified technology and operating parameters and calculate the resultant effect on device performance or lifetime. Each model is based on a physics of failure approach built around the understanding of the mechanism.

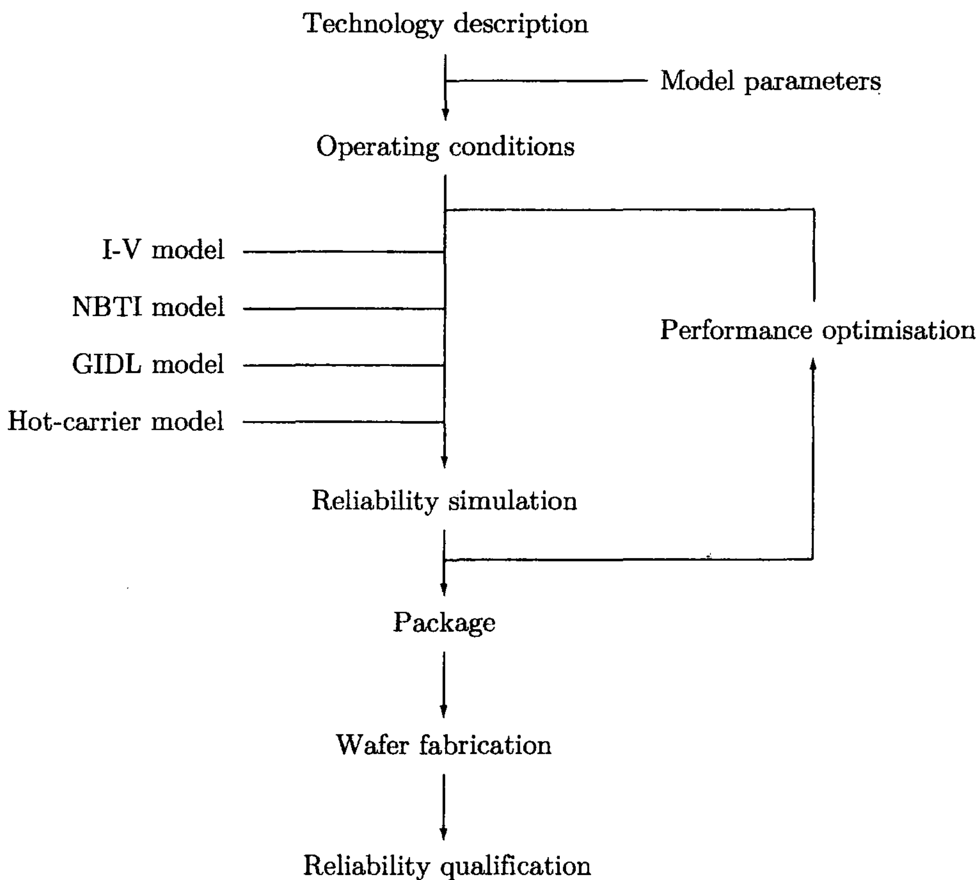


Figure 7.1: Integrated reliability methodology

With a top-down approach the operating conditions are fed into models to simulate the actual failure mechanisms occurring in the application environment. This allows the influence of outside factors, such as temperature, on device performance to be

addressed. Designers can use the simulations to aid optimisation and provide the opportunity for interactive feedback during the design process. How design choices affect reliability of the chosen design can be analysed and then any faults fixed by re-design before fabrication. The intention is to give the user the opportunity to make appropriate trade-offs between performance and reliability, and reduce product development cost and time.

7.3 Tool Design

This section describes how the integrated reliability methodology was implemented into a usable simulation tool. A means of combining the current-voltage and failure models into a complete tool was needed. Up to this point all the mathematical calculations and models had been written using MathCAD™ worksheets for analysis. If each model were to be treated separately, then this program is excellent as it allows for easy input of formulae and provides quick visual results. A way of using these files in the proposed tool was investigated. MATLAB™ was considered as an option and the complete I-V model was implemented in this program. However only very slight improvements in simulation time were noticed. Additionally one of the aims of this tool is to allow the user to easily change a parameter and see the resultant effects and it was felt the MathCAD interface was more suited to this. In addition, the suppliers of MathCAD provide an extension that allows MathCAD files to be embedded into an Excel™ spreadsheet. Several input and output links can be made to allow data to be transferred from the MathCAD file to an Excel spreadsheet and vice versa. This provided for the exact ideal of the methodology: a way of giving the user access to the workings behind the models and calculations whilst providing a visual user interface.

Figure 7.2 illustrates the MathCAD file that was developed containing the models. The complete I-V model and failure models were written into a single file. Drop-down areas were then used to section the various input parameters and models. A selection of these drop-down boxes have been expanded in the diagram and placed behind the main model file.

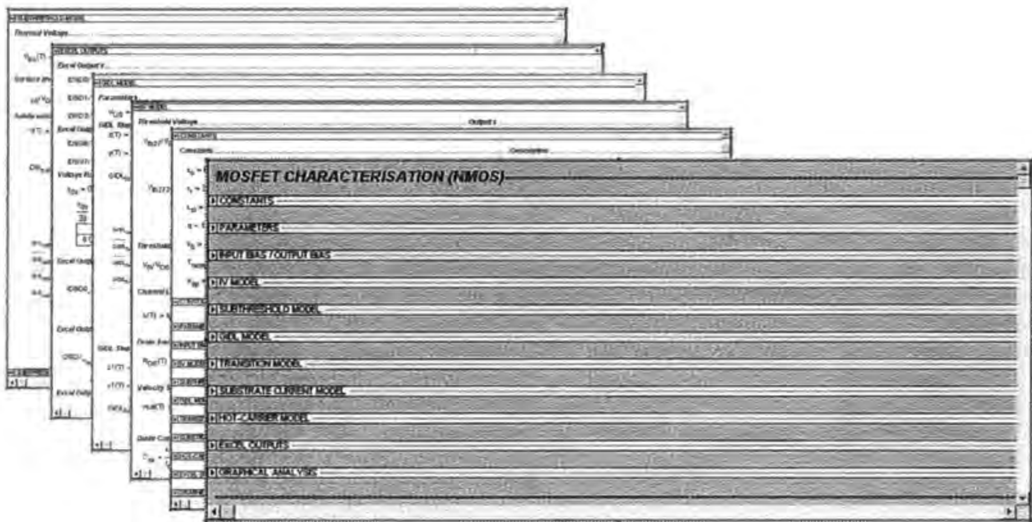


Figure 7.2: Layers of MathCAD development

The MathCAD application calculates top-down and so all constants and input parameters are placed at the top. The parameters drop box contains all the technological and dimensional parameters such as device channel width and length, gate oxide thickness, and carrier mobility values. The extracted fitting parameters obtained from the accelerated DC tests are also contained in this box. User defined parameters such as drain and gate voltages, operating temperature and stress times are contained in the input bias/output bias drop-down box; this allows the user to set the operating and environmental conditions as appropriate. Thereafter the models developed in the previous chapters including the I-V, GIDL, substrate current, and hot-carrier models are implemented into the corresponding drop box.

Visual analysis is designed into the MathCAD file in the graphical analysis box, which contains a collection of graphs illustrating the model predictions. However for successful tool implementation, real time comparison to experimental data would be a great asset. To achieve this, the MathCAD file was embedded into an Excel worksheet.

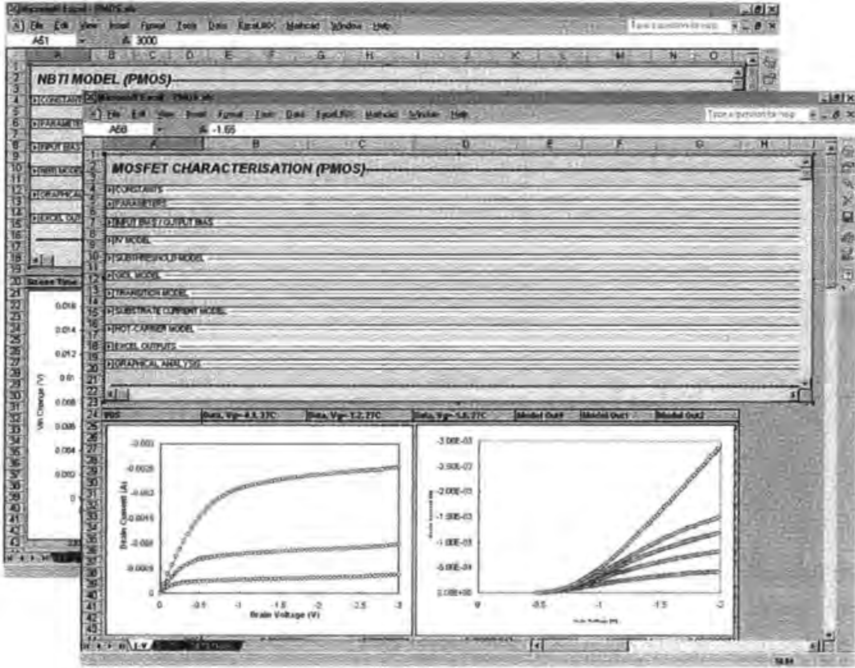


Figure 7.3: Embedded MathCAD file

Figure 7.3 shows a snapshot of the Excel worksheet. The MathCAD file is embedded into the worksheet and then different input/output variables defined. For example, in the first worksheet the MathCAD file illustrated in figure 7.2 is embedded. The predicted drain current outputs are entered into the worksheet at the appropriate point. All that is then required is to click the calculate button on the MathCAD toolbar to generate the result. Depending on the set input parameters, the MathCAD file outputs the corresponding drain current values and updates the graph. This allows for immediate comparisons between the measured data and model predictions. Up to ten inputs/ouputs are available for each embedded file and so a

range of model simulations can be output into the worksheet. For instance the second sheet compares the model predictions to measured data of a drain current characteristic at varying temperatures. An additional sheet for the NBTI model is provided for in a third sheet. This is separate from the main I-V model as it is only required for p-channel devices.

7.4 Simulation Results

With the tool implemented into the simulation framework as described in the previous section, it can be used to predict changes for a wide range of parameters. In this section the most prominent factors that influence future technologies are investigated and simulation results presented.

7.4.1 Scaling

Scaling of device dimensions has always been at the forefront of technology development. With dimensions now entering the nanometer scale, it becomes vital that any reliability or performance problems relating to reduced dimensions are addressed. Thus, the provision of a tool such as the one developed in this work is essential to predict the fast change in device scaling. In 2006 Li *et al* [15] stated that the advancement of device failure modelling had fallen behind the development of CMOS technology which has raised many new issues related to both circuit performance and reliability. Throughout the history of CMOS development a new technology generation has typically been introduced every 2 years. At present we are approaching a time when further scaling may be limited by physical factors and the capability to predict whether the present rate of scaling can continue would be of great use.

The two main reasons to decrease the dimensions of a MOSFET are circuit density and speed. Ever decreasing device dimensions means higher device density, shorter propagation delay times, and hence improved performance. However at the nanometer scale there are some notable concerns including increased leakage current, direct tunnelling in ultra thin gate oxides, and increased power dissipation [26].

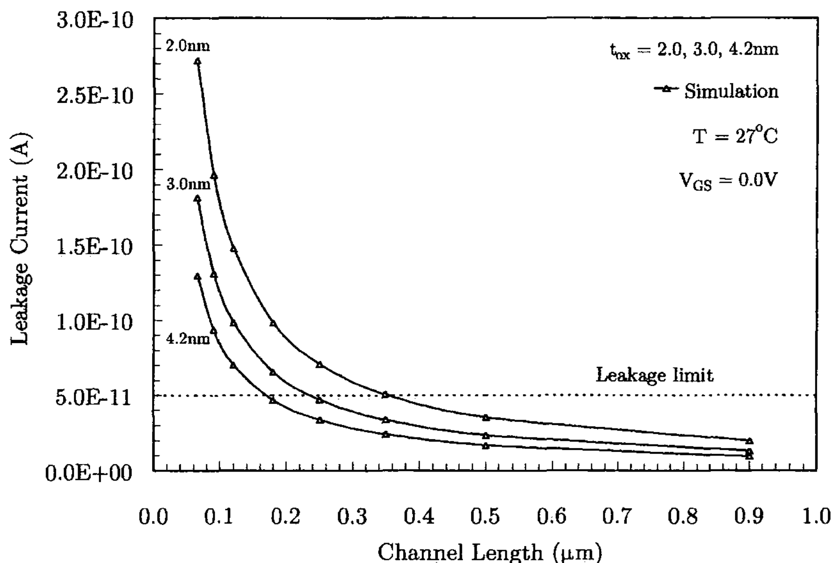
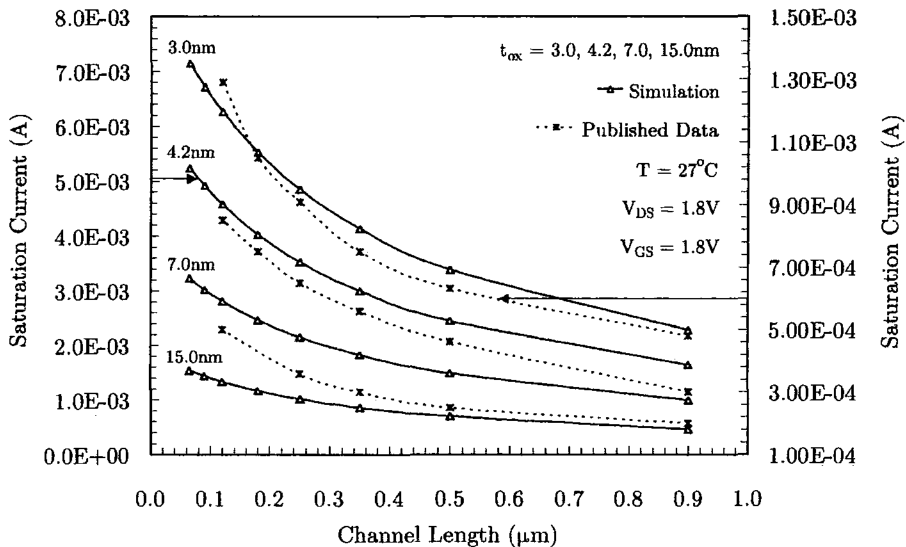


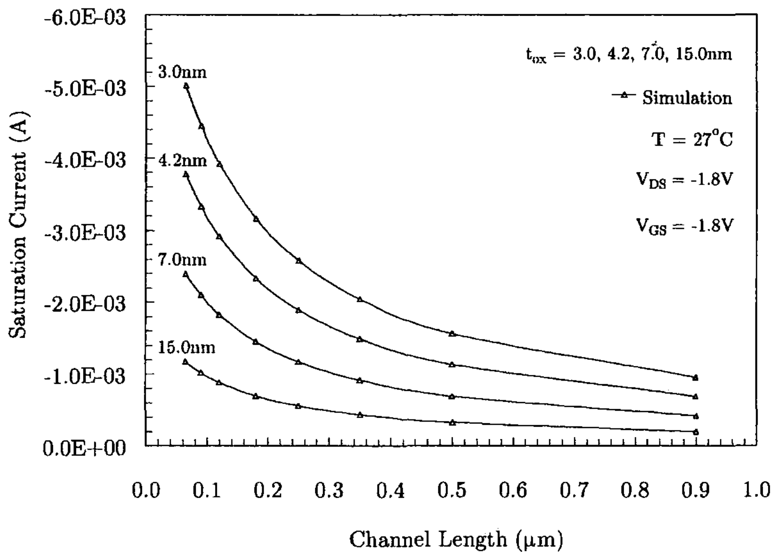
Figure 7.4: Leakage current as a function of channel length (n-channel)

Figure 7.4 illustrates an example result using the simulation tool. The leakage current of a n-channel device was simulated at varying channel lengths and oxide thicknesses. The typical leakage current limit of 50pA set by the wafer manufactures is also shown. As the channel length is reduced the leakage current rises. The leakage is also enhanced for thinner gate oxides. This is because as devices become smaller and the oxide thickness is reduced the electric field rises, resulting in increased leakage. This trend is particularly relevant to battery operated devices where an increase in leakage causes higher power dissipation when the device is in its OFF-state, and hence reduced battery life. To overcome increased leakage it is common practice to scale the operating voltage. For today's and future technologies this may no longer be an option. Traditionally the threshold voltage along with the operating

voltage have been reduced relative to the reduction in device dimensions. However the difference between the thermal voltage and threshold voltage has become smaller and smaller to the extent that it can no longer be reduced. Hence there is a physical limit to the difference between the operating voltage and threshold voltage. As well as leakage limits, another important parameter affected by scaling is saturation drain current.



(a) n-channel



(b) p-channel

Figure 7.5: Saturation current as a function of channel length

From figure 7.5 we can see that for both n-channel and p-channel devices with thick oxides the scaling of channel length has little effect on the saturation drain current. We would expect this from equation 3.19. As the channel length approaches zero, the drain current approaches a constant value. However for thinner oxides the oxide capacitance and electric field is greater, magnifying current increase. The simulation results also match the trend of previously published data [106] for similar devices.

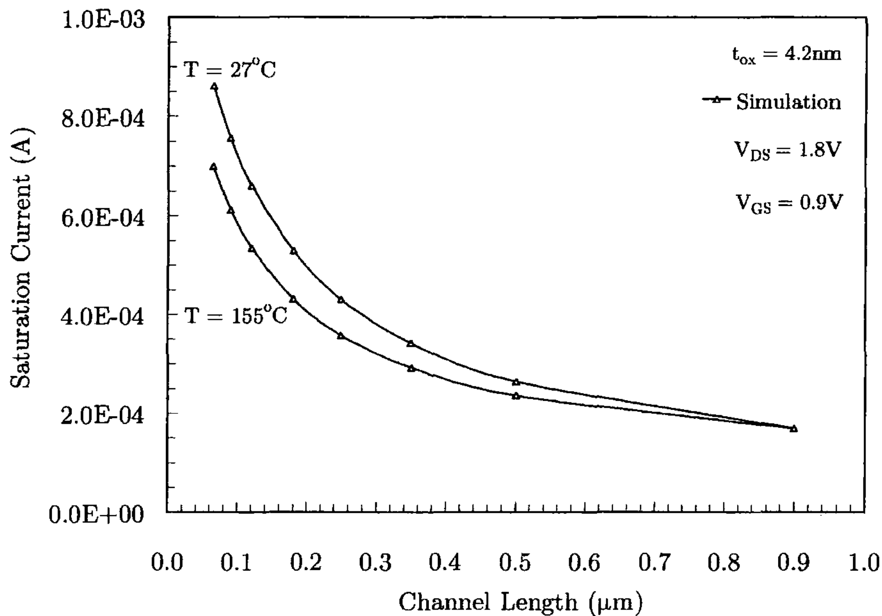


Figure 7.6: Influence of temperature on scaling channel length (n-channel)

Further analysis of scaling and its effect on the saturation drain current was undertaken, simulating the influence, if any, of temperature on the device. In figure 7.6 the saturation current is predicted as a function of channel length at 27°C and also 155°C. For large channel devices ($l > 0.5\mu m$) the reduction in drain current is minimal. As the the channel length is reduced the influence of temperature becomes more prominent. In section 3.4, we showed that when a device is in saturation the dominant cause of drain current reduction is a decrease in carrier mobility. Hence it is proposed that as channel length is reduced the influence of temperature on mobility reduction is increased and subsequently we see a greater decrease in current.

Being able to predict scaling trends of the saturation drain current is vital for foreseeing the performance of future technologies. The reason being that this determines the time needed to charge and discharge capactive loads in critical paths. This therefore impacts speed more so than any other parameter. Equation 7.1 [107] shows a simple expression for the propagation delay of an inverter.

$$\tau_d = \frac{C_L \cdot V_{DD}}{4} \cdot \left(\frac{1}{I_{DS_{(n)}}} + \frac{1}{I_{DS_{(p)}}} \right) \tag{7.1}$$

Using the simulated drain current values, the average time τ_d to discharge a load capacitor C_L from V_{DD} to V_{DD} / 2 and to charge from 0 to V_{DD} / 2 can be calculated. The simulated time delay for a single inverter where C_L is assumed to be constant and the p-channel device is twice the width of the n-channel is shown in figure 7.7.

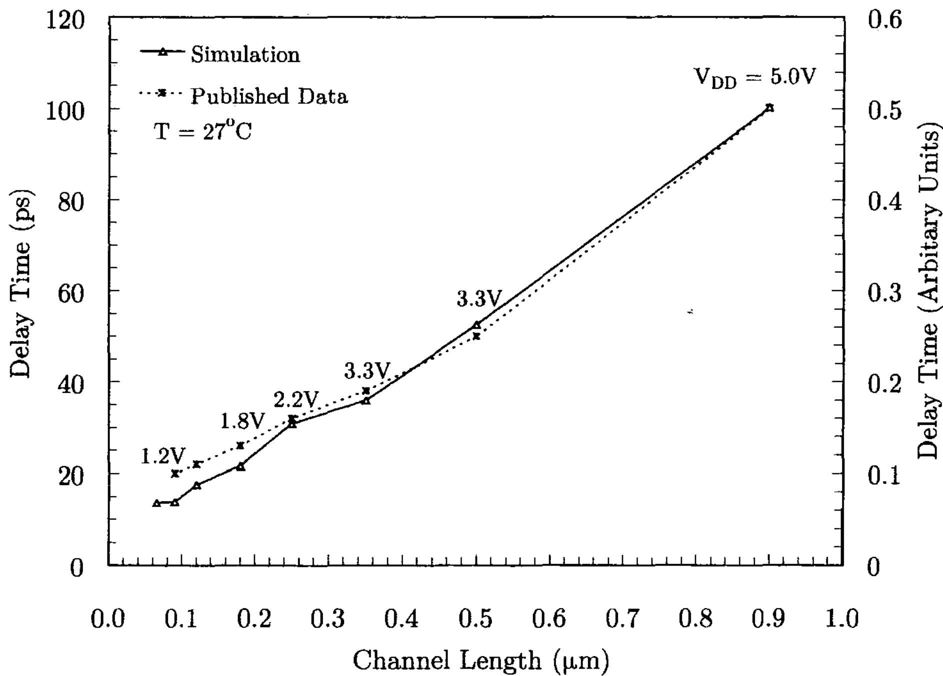


Figure 7.7: Delay time as a function of channel length

It can be seen that the simulation follows previously published trends [108], with the switching speed roughly doubling every two generations. The prediction for the 0.18μm agrees well with manufacturers data that specify an approximate delay of

27ps per stage. The scaling method used in this prediction was both a voltage and gate oxide thickness reduction. The corresponding values used in the simulation for past, present and future technologies are shown in table 7.1.

Parameters	0.9μm	0.5	0.35	0.25	0.18	0.12	0.09	0.065
V_{DD} (V)	5.0	3.3	3.3	2.2	1.8	1.8	1.8	1.2
t_{ox} (nm)	19.0	11.0	9.0	6.5	4.2	4.0	3.4	2.2

Table 7.1: Scaling parameters

This constant scaling method attempts to keep the electric field approximately constant by reducing the oxide thickness and operating voltage appropriately for each technology. As shown in figure 7.7 this approach allows the required increase of device speed. However the desire for increased speed and subsequent reduction in dimensions can enhance stress and have a detrimental effect on device reliability.

7.4.2 Failure Mechanisms

In addition to using the simulation tool to predict the effect of scaling on device performance, it can also be used to analyse device degradation. Within the tool, each failure model was incorporated into a drop-down box in the MathCAD file as described in section 7.3. Each failure model can then use data from the MathCAD calculations. It was assumed that each of the mechanisms (NBTI, GIDL, and HCE) were independent from each other, and hence implemented in separate models. By using different parameters as the degradation monitor for each model, i.e, threshold voltage in NBTI, offset current in GIDL, and substrate current in HCE, each mechanism can be separately analysed and provide comparisons of resulting degradation. It was believed that developing a tool capable of simultaneously simulating the interaction of two or more mechanisms would significantly increase computation time and the results would be difficult to verify experimentally.

7.4.2.1 NBTI

With the increase of NBTI effects being seen in recent deep sub-micron technologies, it is essential that any simulation tool should include this effect. Previously, simulation tools such as HOTRON [11], RELY [12], BERT [13], CAS [14], and SCALE [65] have only concentrated on hot-carrier degradation. In chapter 4 we presented a new model for NBTI effects in p-channel devices. To implement this model into the simulation tool an extra worksheet was added to embed the NBTI model within the Excel framework. This was seen previously in figure 7.3. Within the NBTI model a drop-down box was used to define the stress conditions allowing the user to specify stress voltages, stress times, stress temperatures and signal frequencies. A snapshot of the result is shown in figure 7.8.

Figure 7.8: Snapshot of NBTI model in simulation tool

Having the ability to transfer data between embedded MathCAD files means that the predicted change in threshold voltage as a result of the NBTI stress can be fed back into the I-V model. Hence the user can specify the stress conditions and then simulate the resultant effect on device characteristics. Typical calculation time to calculate the change in V_{th} and update the current-voltage characteristics was around 11 seconds. The only modification of the I-V model needed was the threshold voltage calculation which is now expressed as:

$$V_{th} = V_{th_{(\mathrm{Fresh})}} + \triangle V_{th_{(\mathrm{NBTI})}} \tag{7.2}$$

where $V_{th_{(\mathrm{Fresh})}}$ and $V_{th_{(\mathrm{NBTI})}}$ are the original threshold voltage and voltage change after NBTI stress. Figure 7.9 shows an example NBTI stress simulation. The device was stressed with a gate voltage of -2.8V for 72 hours at 155°C. The model simulates the increase in threshold voltage due to the creation of interface traps. Subsequently the saturation drain current, and device operating speed, is reduced after stress.

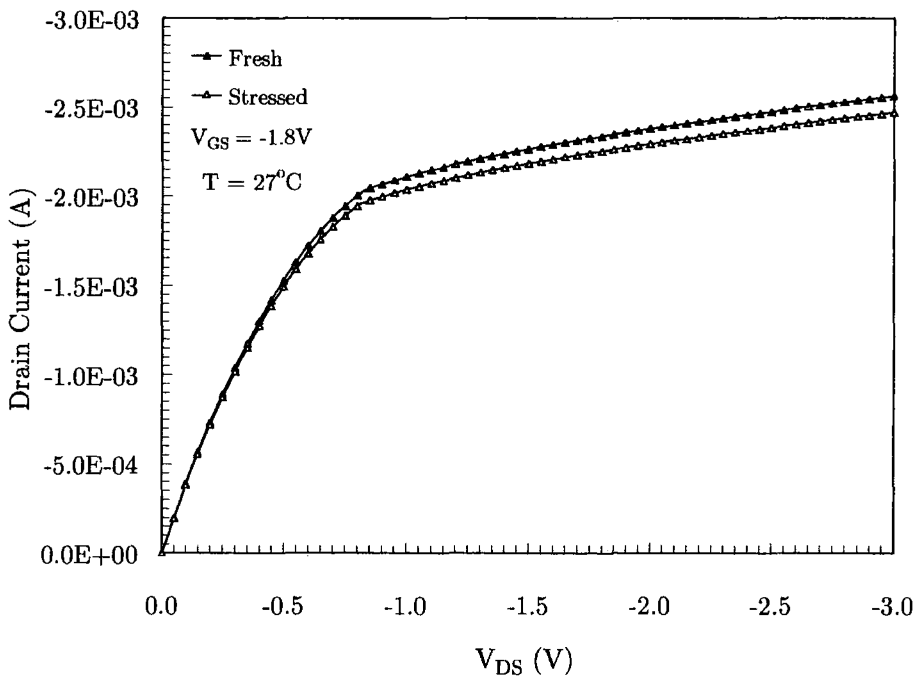


Figure 7.9: NBTI stress simulation

In chapter 4 we highlighted the enhancement of NBTI stress at elevated temperatures. This can also be simulated by the tool. Figure 7.10 illustrates an NBTI simulation at three stress temperatures: 27°C, 155°C and 250°C, with a negative stress voltage of -2.8V. The resultant increase in threshold voltage is depicted in an $I_{DS} - V_{GS}$ characteristic biased in the saturation region at room temperature.

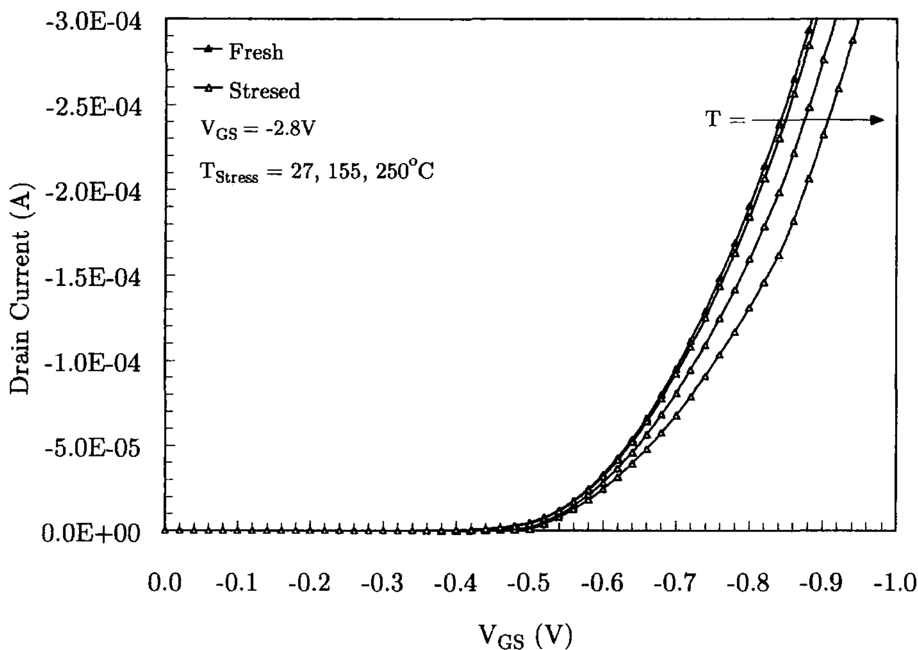


Figure 7.10: NBTI stress simulation at varied temperatures

At higher stress temperatures an increase in threshold voltage shift can be seen. This simulation result supports the belief that it is the combination of both elevated temperature and a negative voltage that causes the most significant degradation as only a small V_{th} shift is seen at 27°C. In section 4.4 the increased shift in threshold voltage was attributed to the increased breaking of silicon-hydrogen bonds at the silicon-oxide interface; elevated temperatures and negative voltage stress causes the dissociation of hydrogen species away from the $Si/Si0_2$ interface leaving behind an interface trap. This changes the charge distribution and hence causes an increase in threshold voltage.

As well as temperature causing an increase in NBTI degradation, it is expected that an increase in oxide field would also affect the mechanism. Device scaling inevitably reduces the gate oxide thickness and results in an increase in oxide electric field. The NBTI model developed in chapter 4 had an oxide thickness dependency built into equation 4.2. Thus a simulation could be undertaken to investigate how the NBTI mechanism may be affected by future scaling and the progression towards ultra deep sub-micron technologies.

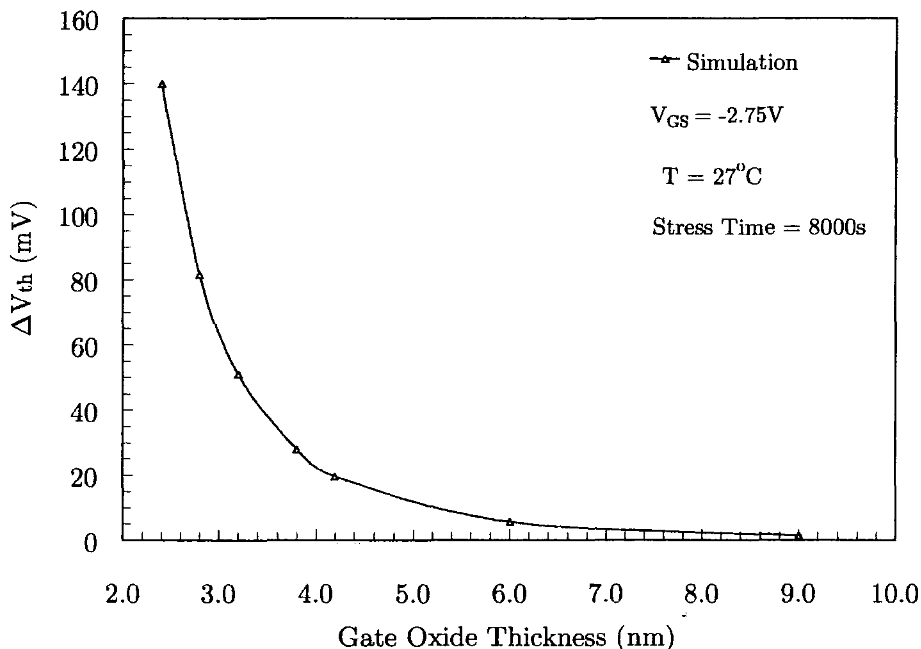
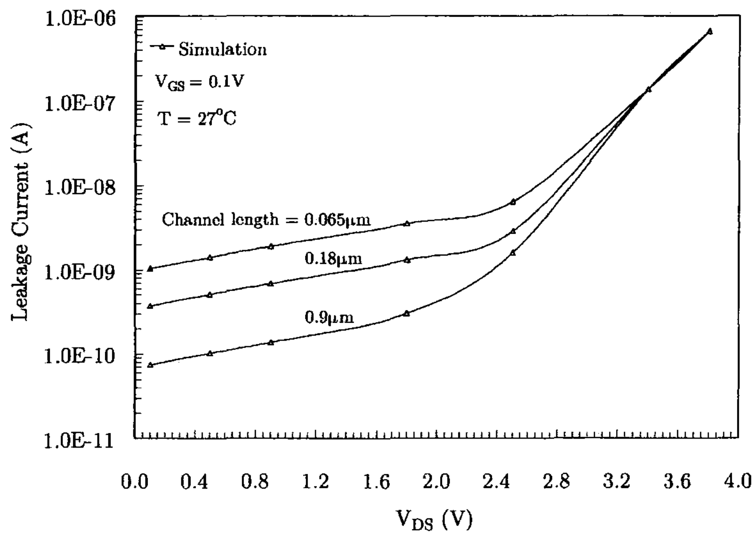


Figure 7.11: NBTI stress simulation as a function of oxide thickness

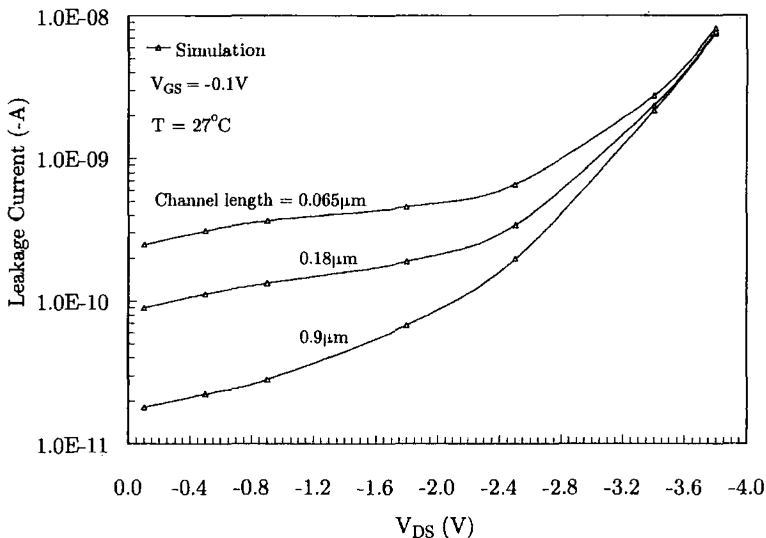
Oxide thickness values were chosen to be representative of recent, past and future technologies. For past technologies (t_{ox} > 6nm), the shift in V_{th} after an 8000sec stress is seen to be negligible. However as the oxide thickness is reduced (t_{ox} < 4nm), we can see that the V_{th} shift is significantly increased. This is due to an increased oxide electric field and hence increased probability of interface trap generation. Simulations undertaken to see if any reductions in channel length had an effect found that ΔV_{th} was minimal, agreeing with results presented by Chaparala *et al* [109]. This is to due to NBTI stress not involving any lateral electric field stress.

7.4.2.2 GIDL

Analysis of the GIDL mechanism in chapter 5 highlighted that the increased leakage current was attributed to band-to-band tunnelling taking place in the gate-to-drain overlap region. This would suggest that GIDL would not be affected by a reduction in channel length. With the GIDL model implemented into the tool, figure 7.12 shows the resulting simulated leakage current as a function of drain voltage.



(a) n-channel



(b) p-channel

Figure 7.12: Voltage dependency of GIDL for varied channel lengths

It can be seen from figure 7.12 that for low V_{DS} the leakage current increases with smaller channel lengths. This agrees with figure 7.4 and is attributed to increased electric field. However at higher drain voltages, where GIDL is the dominant mechanism, the leakage is insensitive to channel length variation. This is because the tunnelling depends only on conditions in the immediate gate-to-drain overlap region and any reduction in channel length has little effect on the mechanism. These findings agree with a study by Chung *et al* [40] which found that GIDL was virtually independent of channel length and support the results presented in chapter 5.

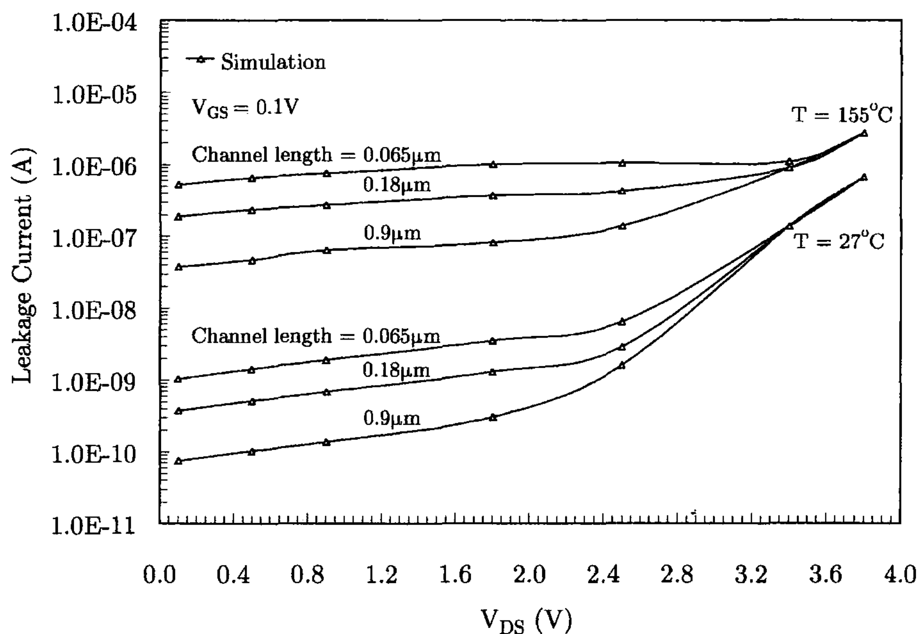


Figure 7.13: Temperature dependency of GIDL at varied channel lengths

Further analysis of GIDL is shown in figure 7.13. The simulation shows that increased temperature has no effect on the channel length dependency of the GIDL characteristic. This simulation supports findings in chapter 5, that in the low V_{DS} region the increased leakage at high temperatures can be attributed to thermal emission and in the high-field region the leakage current is insensitive to temperature since the tunnelling probability and electric field across the oxide is not strongly dependent on temperature.

7.4.2.3 Hot-Carrier Effects

In a similar fashion to the NBTI and GIDL models, a hot-carrier model was written into a drop-down box in the simulation tool. A snapshot of the lifetime model is shown in figure 7.14. Drain and substrate current values are fed from above drop-down boxes to allow for lifetime calculations. Also shown in figure 7.14 is the addition of substrate current with the current-voltage calculation.

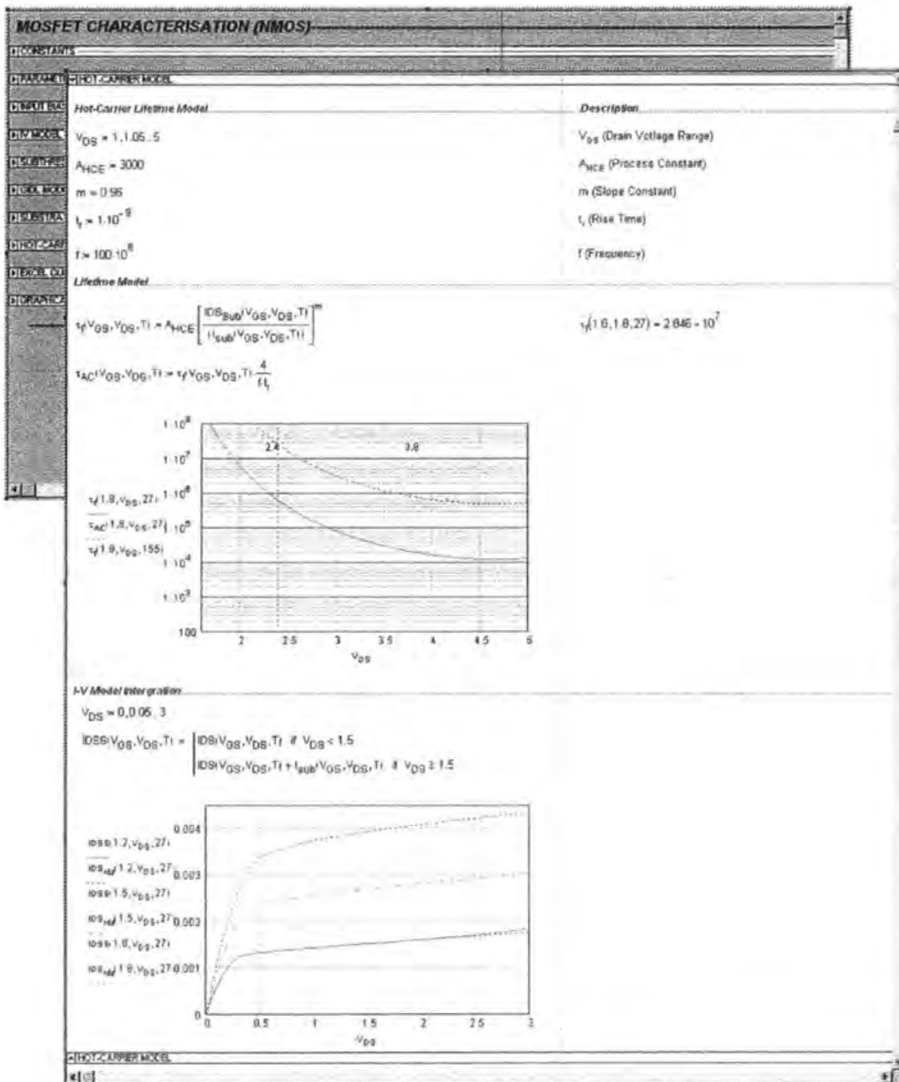


Figure 7.14: Snapshot of HCE model in simulation tool

The hot-carrier study covered in chapter 6 was based around a substrate current model. This was because the substrate current and hot-carrier mechanism are physically linked via impact ionisation. Therefore the substrate current is an excellent measure of hot-carrier stress. However, during hot-carrier stress the main degradation that affects device performance is a threshold voltage shift. This is because electrons may gain enough energy to be injected into the oxide and subsequently change the oxide charge distribution. Hence for the complete simulation tool changes in V_{th} were linked to the substrate current with the following expression.

$$\Delta V_{th} = C_{HCE} \cdot \left(\frac{I_{Sub}}{W}\right)^{m} \cdot t \tag{7.3}$$

where C_{HCE} is a process constant, t stress time, and m typically takes values between 0.2-0.5. Figure 7.15 shows a hot-carrier stress simulation. The n-channel device was biased at maximum substrate current with a drain voltage of 3.4V. The simulated stress time was 72 hours and the temperature was set at 27°C.

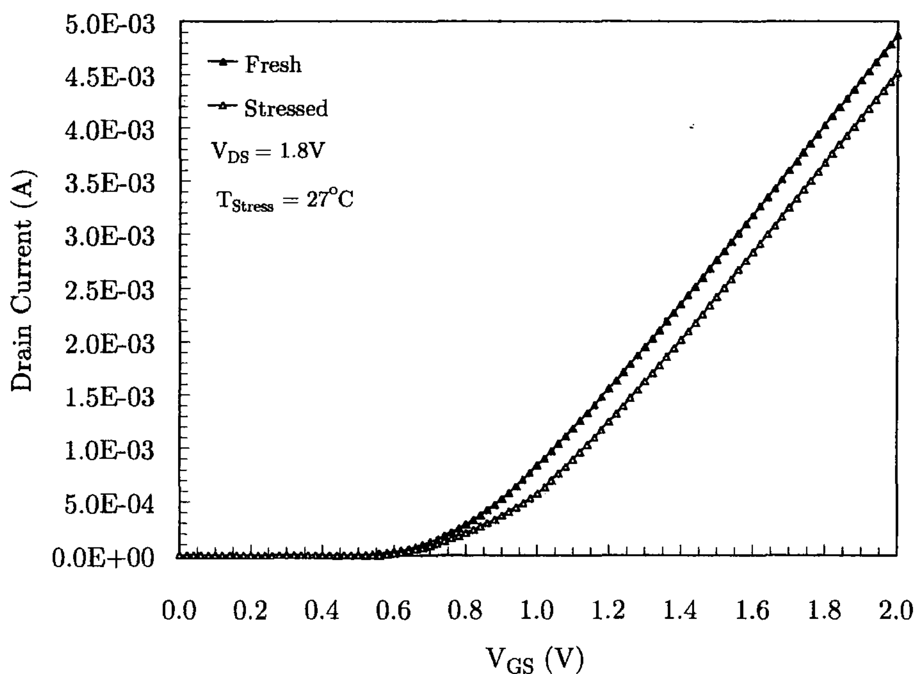


Figure 7.15: Hot-carrier stress simulation

From figure 7.15 we can see that after hot-carrier stress the introduction of interface traps has increased the threshold voltage making it harder to turn the device on. Similar results were seen for p-channel simulations. With the hot-carrier model integrated into the simulation tool, further investigations were undertaken for a range of channel lengths. In this simulation we investigated whether the substrate current has any channel length dependence. The results are presented in figure 7.16.

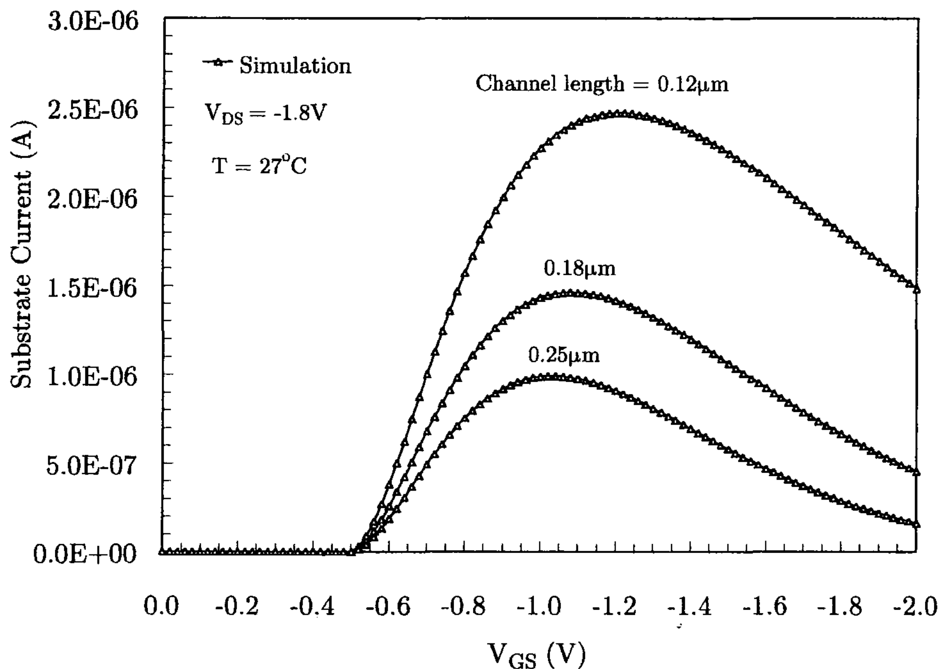


Figure 7.16: Substrate current as a function of channel length (p-channel)

As the channel length is reduced we see an increase in substrate current. This occurs because as the channel length is scaled, there is an increase in electric field which enhances carrier velocity in the channel. Hence carriers with more energy result in increased impact ionisation and substrate current generation for shorter channel lengths. To further understand the relationship between electric field and channel length figure 7.17 illustrates the relationship between the two. As expected, the electric field for both the n-channel and p-channel device increase with reduced channel length under constant voltage scaling.

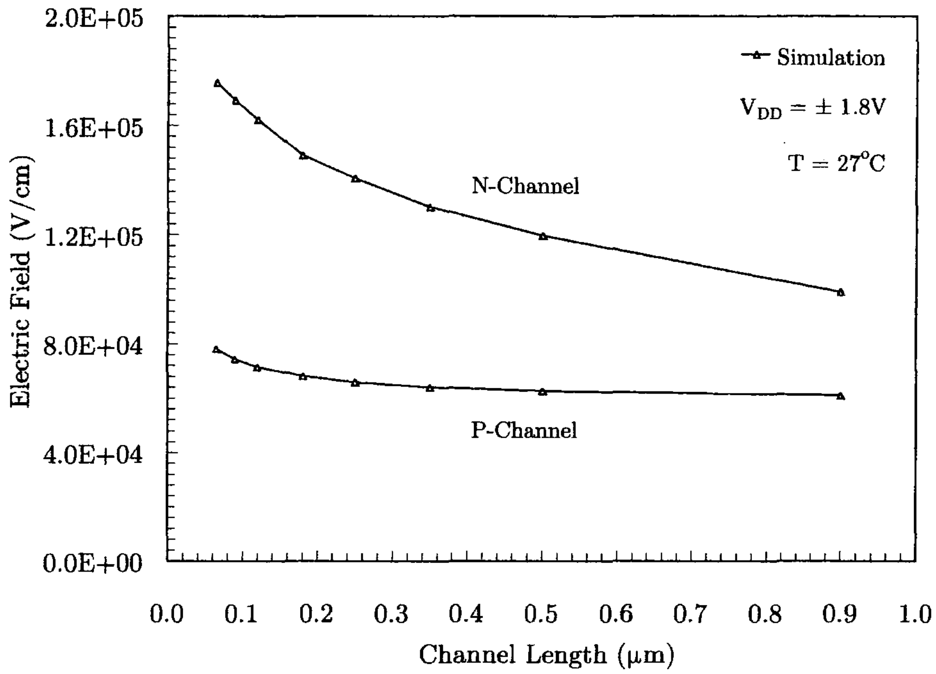


Figure 7.17: Maximum electric field as a function of channel length

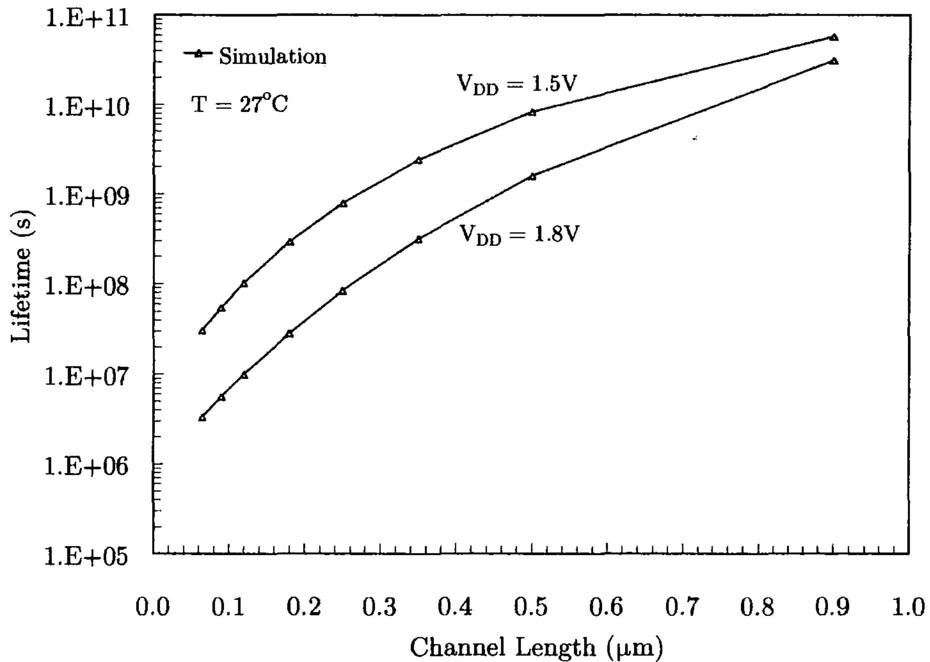


Figure 7.18: Device lifetime as a function of channel length (n-channel)

Inevitably higher electric fields increase electrical stress and reduce device lifetime. In figure 7.18, the DC worst case lifetime is simulated for an n-channel device. It can be seen that as the channel length is reduced the lifetime is shortened. The decrease in lifetime for shorter channel devices occurs due to the increased electric field and hot-carrier stress. Also shown in figure 7.18 is the projected lifetime for a lower operating voltage. As expected, with reduced voltage and electric field, hot-carrier degradation is reduced and hence the lifetime of the device is lengthened.

7.5 Conclusion

In this chapter a simulation tool developed from a new integrated reliability methodology has been presented. Failure models have been integrated into the tool which can be used by designers to compare simulations and measured data and to evaluate the effect of any design change on device performance and reliability. The idea behind this methodology is to allow any design changes to be undertaken at the early stages of product development, to ensure robust design and reduce testing and re-design time.

In section 7.3 the integration of the failure models into a simulation tool was covered. With the use of MathCAD and Excel a tool was developed that allowed the user to delve into its inner workings and become familiar with their operation and function. Each failure model was integrated into separate drop-down boxes and the relevant information fed between the MathCAD file and excel for analysis. The tool allowed the user to change any parameter and visualise the resultant calculation immediately. This enabled the user to simulate many changes such as the impact of scaling on the reliability and performance of future ultra deep sub-micron technologies.

Simulations predicted that as devices become smaller the saturation drain current increases and hence delay time decreases. A simulation of a single inverter showed that as the channel length was reduced, it becomes possible to approximately maintain the doubling of speed every two device generations. Additionally it was shown that the increase in saturation drain current is greater in shorter channel devices for thinner gate oxides. This is due to the increased electric field. However the reduction of device dimensions was not without its disadvantages and leakage current simulations showed a significant increase for short channel devices. This increase was enhanced for thinner gate oxides and could be the inevitable physical limit to future scaling. The influence of temperature on device performance for small devices was also investigated revealing that the reduction in drain current at elevated temperatures was greater for smaller devices. It was proposed that this was because the main cause of drain current reduction, mobility degradation, was enhanced in short channel devices.

The NBTI model was integrated into the complete tool in section 7.4.2.1. It allowed the user to specify any number of stress configurations and see the resultant effects. Simulations showed the expected increase in the threshold voltage after a NBTI stress caused a decrease in drain current values reducing device speed. The ability to predict this degradation at varied temperatures was also shown. Furthermore simulations showed that the change in threshold voltage associated with NBTI stress significantly increased as gate oxide thickness was reduced. This is because NBTI is linked to only vertical electric fields and it is seen that reduction in channel length plays no part in the level of degradation.

GIDL simulations identified an increase in leakage at low drain voltages for shorter channel devices. However at higher drain voltages the GIDL current was found not

to be influenced by variations in channel length. The reason for this is that GIDL current is caused by tunnelling only in the immediate gate-to-drain overlap region. Hence a reduction in channel length has little affect on the mechanism. Simulations found also that increased temperature had no effect on the channel length dependency of GIDL.

Simulations using a hot-carrier model in section 7.4.2.3 demonstrated the effect of hot-carrier stress on operating characteristics. A threshold voltage expression linked to the substrate current calculations was presented. The increase of V_{th} after hot-carrier stress and the resultant effect on the $I_{DS} - V_{GS}$ curve was demonstrated. The substrate current was found to increase for shorter channel lengths and simulations showed that this was because of increased electric field in the pinch-off region. Hence device lifetime decreased for smaller devices. However if the operating voltage is reduced at the same time as dimensions the electric field can be reduced and subsequently lifetime lengthened.

In summary, a new simulation tool has been developed that for the first time includes NBTI effects as well as GIDL and hot-carrier degradation. Its design within MathCAD and Excel means that new models can be easily added in the future. It provides for a quick and easy to use tool that allows for immediate comparison of measured and simulated data. Any parameter can be easily modified by the user and all workings are completely accessible. The tool can provide the designer with information to understand how devices degrade over time and can reinforce any design changes in the early stages of product development.

Chapter 8

Future Work and Conclusions

THE MAIN AIM of this research was to address the impact of temperature on device performance and failure mechanisms utilising a physics-of-failure approach. The thesis has focused on individual aspects of device operational characteristics and device degradation accumulating in the development of a new methodology and simulation tool. This chapter covers areas of future work that it is felt could further the foundations presented in this thesis. Conclusions of the thesis are presented in section 8.2.

8.1 Future Work

In addition to the novel results and analysis presented in this thesis the research provides a framework and foundation for future development. The development of a simulation tool together with several new failure models focused on the influence of temperature, have provided many new results. However analogous to the ever increasing desire for higher performance and the requirement of smaller devices, there is scope for further development of the models and simulation tool.

8.1.1 Layout Diagnostics

A simulation tool was developed in chapter 7 to predict performance and degradation at device level. Potentially this tool could be developed further to provide circuit anaylsis. Throughout the thesis deivce lifeitime has been defined as a 10% shift in a certain parameter. Whilst this is an arbitrary value that reflects the degree of wear-out, this level of drift may not necessarily lead to a circuit failure. If the simulation tool could include circuit layout information, signals levels could be analysed and allow analysis of circuit functionally and identification of high risk areas.

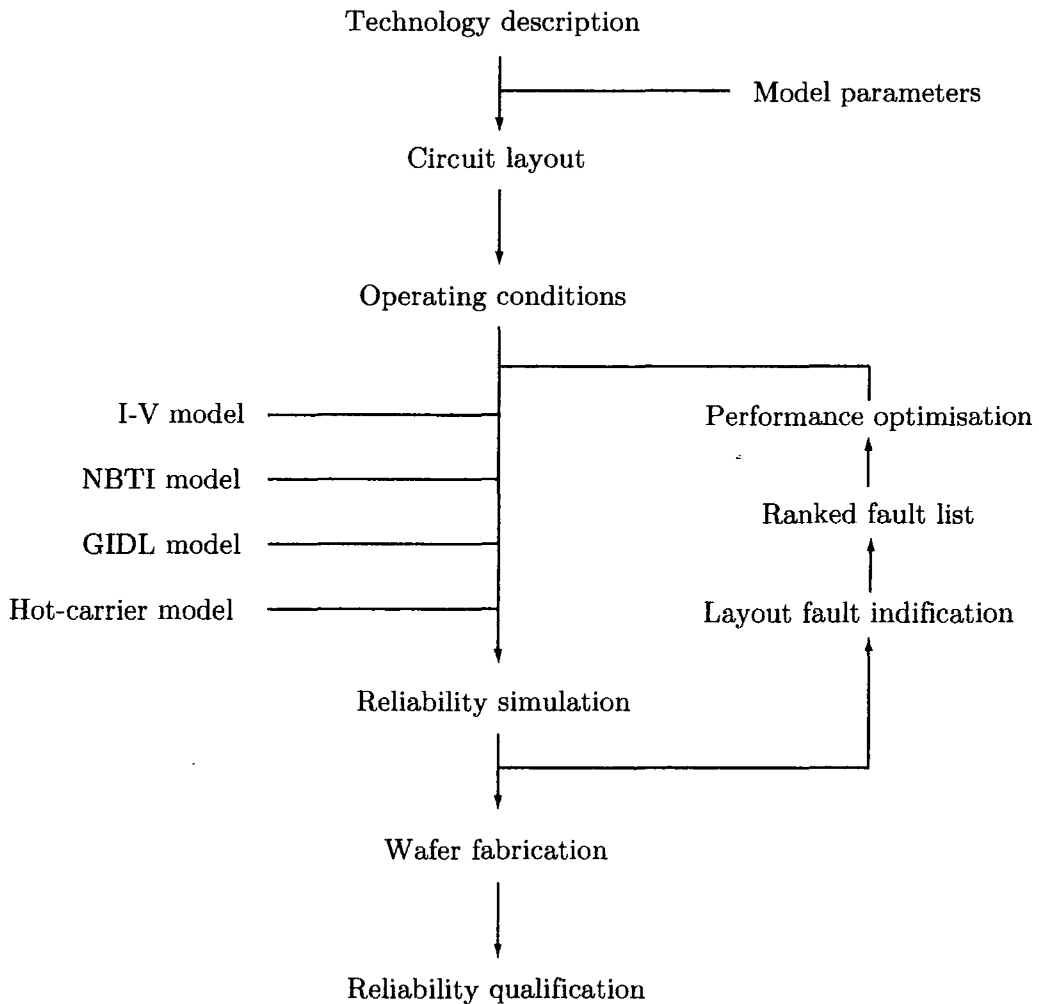


Figure 8.1: Development of methodology for layout diagnostics

Figure 8.1 illustrates how the methodology presented in section 7.2 may be adapted to provide for layout diagnostics. Ideally the circuit simulation would be able to use layout information to translate degradation information into a ranked fault list and allow identification of weak-spots in the circuit. For example, analysing the signals each device is exposed too would allow those devices which experience high DC bias levels to be highlighted for potential NBTI damage. Threshold voltage shifts could be calculated and the resultant effect on circuit functionality simulated. However this is not a trivial exercise since some of the NBTI stress effects are recovered in dynamic circuits and extra thought would be needed as to how such a calculation is performed. Furthermore circuit simulations would have to take into account process variability, potentially over millions of devices.

The current tool provides accurate simulation at device level. Development of the tool to render it capable of layout analysis would be an excellent enhancment. However the added accuracy provided by cirucit analyisis would inevitably increase complexity, and experimental work, whilst compromising computational efficiency.

8.1.2 Competing Failures

The simulation tool developed in chapter 7 included three different degradation models in the same environment. Although all three models work in conjunction with the I-V model, each mechanism was treated by an individual model such that they are independent of each other. Using different parameters as the degradation monitor for each model, i.e, threshold voltage in NBTI, offset current in GIDL, and substrate current in HCE, each mechanism was separately analysed to provide comparisons of resulting degradation. A disadvantage of this method is that it does not account for any interaction between mechanisms.

A device will experience varying stress conditions throughout its operating life, and so will be exposed to different degradation mechanisms at different times. This means that the generation of interface traps, for example, may be contributed too by both hot-carrier effects and NBTI during a particular device lifetime. Hence it would be advantageous to have the capability of simulating the interrelationship of two or more mechanisms. This could be potentially undertaken by analysing signals patterns over time to determine the relative occurrence of each stress condition associated with each failure mechanism. However time constraints may make this impractical and it would be difficult to simulate actual real-life signal characteristics. Hence it seems at present the most accurate way of predicting lifetime is to treat each mechanism separately for direct comparison.

8.1.3 Tool Implementation

In section 7.3 the design of the simulation tool using a combination of MathCAD and Excel was presented. These packages were chosen because I wanted to design a tool that not only provided a front-end interface, but also allowed users to delve into its workings. This however does not restrict the implementation of the models developed in this work to these packages and any number of programs could be used. For instance some time was spent implementing the models in MATLAB, the popular industry standard mathematical analysis program. However I found the use of MathCAD and Excel provided for an excellent compromise between real-time visual analysis and computational efficeincy. Furthermore if a tool was required with only front-end interface, such as SPICE, then there is no reason why the models could not be written into such a tool.

8.2 Conclusions

The design of a novel simulation tool that encompasses prominent and new failure mechanisms in modern CMOS devices has hopefully gone some way to reduce the gap between failure modelling and the continued development of CMOS technology. Issues relating to the effect of temperature on deep sub-micron devices have been addressed and novel findings presented to further the understanding of CMOS reliability.

A new high temperature MOSFET model was developed to form the foundation of the simulation tool. A computationally efficient and accurate I-V model gave excellent agreement with measured data and any discontinuities at boundaries between regions were overcome. Using a small set of parameters the model accounted for the major physical effects in modern devices such as velocity saturation, CLM, DIBL, and carrier mobility reduction due to a vertical field. Furthermore the influence of temperature, notably a reduction in threshold voltage, carrier mobility and drain current at elevated temperatures were included in the model with the predictions showing close resemblance to experimental results.

It was shown that with an understanding of the physical process behind many of the failure mechanisms, relatively simple but accurate models can be developed. NBTI damage was demonstrated to be greater in p-channel devices compared to n-channel devices, and a number of distinct NBTI characteristics investigated. A new NBTI model capable of predicting the influence of temperature, recovery and lock-in effects and dynamic degradation was presented. Experimental results demonstrated that in devices that experience dynamic signals, lifetime is prolonged due to hydrogen diffusing back to the $Si/Si0_2$ interface during relaxation periods.

A temperature dependent model for GIDL was developed and validated with good agreement to experimental data. Experimental results demonstrated that GIDL current was almost independent of temperature since the tunnelling across the oxide does not strongly depend on temperature. This supported the choice to accurately model GIDL in both low and high field regions with the use of two separate exponential functions.

Contrary to common belief hot-carrier effects were found to increase at elevated temperatures and low drain bias. A new semi-empirical substrate current model was developed to account for the temperature dependance of the substrate current with good agreement to measurement data. It was highlighted that these findings cast some doubt on the understanding of the hot-carrier phenomena in deep sub-micron devices and is particularly disconcerting to those interested in high temperature operating conditions.

A simulation tool was realised to allow analysis in the early stages of product development. The effects of device scaling on performance were simulated showing an increase in drive current and speed with decreased dimensions. A simulation of a single inverter demonstrated that it was possible to roughly maintain the doubling of speed every two device generations. However an increase in leakage current for thinner oxides highlighted how increased leakage could pose as the inevitable physical limit to future scaling of conventional CMOS devices. Simulations showed that both NBTI and hot-carrier effects were enhanced with decreasing dimensions. It was demonstrated that GIDL leakage at high drain voltages was not affected by variations in channel length as it is caused by tunnelling in the gate-to-drian overlap region only.

The aim of this work was to provide a new methodology with the potential to allow any design changes to be made in the early stages of product development. This aim has been achieved with the development of a new simulation tool and several novel models. Currently a number of papers are being prepared for journal submission which cover individual failure models and also the simulation tool as a whole. It is hoped that the thesis has provided a foundation for further development and the possibility of using its findings in a practical environment.

Gethin Lloyd Owens

Bibliography

[1] A. J. Kennedy, "Modern views of the fatigue processes in metals," *British Journal of Applied Physics*, vol. 15, no. 3, pp. 229–234, March 1964.

[2] A. Coppola, "Pioneering reliability laboratory addresses information technology," Reliability Analysis Centre, Tech. Rep., 1998.

[3] C. Smith, *Introduction to Reliability in Design.* McGraw-Hill Book Co., Singapore: McGraw-Hill, 1976.

[4] K. C. Kapur and L. R. Lamberson, *Reliability in Engineering Design.* John Wiley and Sons, 1977.

[5] P. Banerjee and J. A. Abraham, "Characterization and testing of physical failures in MOS logic circuits," *Design and Test, IEEE Journal on*, vol. 1, no. 3, pp. 76–86, August 1984.

[6] F. S. Shoucair, "Potential and problems of high-temperature electronics and CMOS integrated circuits (25-250°C) - an overview," *Microelectronics Journal*, vol. 22, no. 2, pp. 39–54, April 1991.

[7] F. J. Ferguson and J. P. Shen, "Extraction and simulation of realistic CMOS faults using inductive fault analysis," *International Test Conference*, pp. 475–484, 1988.

[8] X. Li, B. Huang, J. Qin, Y. Liu, X. H. Zhang, M. Talmor, Z. Gur, and J. B. Bernstein, "Analog-to-digital convertors reliability modeling and simulation with SPICE," Centre for Reliability Engineering University of Maryland, Tech. Rep., May 2004.

[9] F. P. McClusky and M. Pecht, "Rapid reliability assessment using CADMP-II," *International Conference on Modeling and Simulation of Microsystems Proceedings*, pp. 495–497, April 1999.

[10] P. Charpenel, F. Davenel, R. Digout, M. Giraudeau, M. Glade, J. Guerveno, N. Guillet, A. Lauriac, S. Male, D. Manteigas, R. Meister, E. Moreau, D. Perie, F. R. Madinska, and P. Retailleau, "The right way to assess electronic system reliability : FIDES," *Microelectronics Reliability*, vol. 43, no. 9-11, pp. 1401–1404, September-November 2003.

[11] S. Aur, D. E. Hocevar, and P. Yang, "HOTRON - A circuit hot electron effect simulator," *IEEE Conference on Computer-Aided Design*, pp. 256–259, November 1987.

[12] B. J. Sheu, W. J. Hsu, and B. W. Lee, "An integrated-circuit reliability simulator - RELY," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 2, pp. 473–477, April 1989.

[13] R. H. Tu, "Berkley reliability tools - BERT," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transaction on*, vol. 12, no. 10, pp. 1524–1534, October 1993.

[14] P. M. Lee, M. Kuo, K. Seki, P. K. Ko, and C. Hu, "Circuit aging simulator," *IEDM Conference Proceedings*, pp. 134–137, 1988.

[15] X. Li and J. B. Bernstein, "Advanced semiconductor wearout mechanisms lifetime and SPICE equivalent circuit modelling," *Device and Materials Reliability, IEEE Transactions on*, pp. 1–19, January 2006.

[16] J. R. Black, "Mass transport of aluminum by moment exchange with conducting electrons," *6th Annual International Reliability Physics Symposium*, pp. 148–159, 1967.

[17] K. Hinode, T. Furusawa, and Y. Homma, "Dependence of electromigration lifetime on the square of current density," *32nd Annual International Reliability Physics Symposium*, pp. 317–326, 1993.

[18] K. Tu, "Recent advances on electromigration in very-large-scale-integration of interconnects," *Journal of Applied Physics*, vol. 94, no. 9, pp. 5451–5473, November 2003.

[19] J. R. Lloyd, "Electromigration in thin film conductors," *Semiconductor Science and Technology*, vol. 12, no. 10, pp. 1177–1185, October 1997.

[20] J. Tao, N. W. Cheung, and C. Hu, "Electromigration characteristics of copper interconnects," *IEEE Electron Device Letters*, vol. 14, no. 5, pp. 249–251, May 1993.

[21] Cadence, "Electromigration for designers," Cadence Design Systems, Tech. Rep., 1999.

[22] R. Degraeve, B. Kaczer, and G. Groeseneken, "Reliability: a possible showstopper for oxide thickness scaling?" *Semiconductor Science and Technology*, vol. 15, no. 5, pp. 436–444, May 2000.

[23] S. Bruyere, E. Vincent, and G. Ghibaudo, "Limiting oxide failure mode versus oxide thickness. Some insights for deep-submicron technologies," *Integrated Reliability Workshop Final Report*, pp. 78–84, 1999.

[24] M. Depas, B.Vermeire, P. W. Mertens, M. Meuris, and M. M. Heyns, "Wear-out of ultra-thin gate oxides during high-field electron tunnelling," *Semiconductor Science and Technology*, vol. 10, no. 6, pp. 753–758, June 1995.

[25] S. Thompson, P. Packan, and M. Bohr, "MOS scaling: Transistor challenges for the 21st century," *Intel Technology Journal*, no. Q3, p. 19, 1998.

[26] Y. Taur, "CMOS design near the limit of scaling," *IBM Journal of Research and Development*, vol. 46, no. 2/3, pp. 213–222, March/May 2002.

[27] Y. Leblebici, "Design considerations for CMOS digital circuits with improved hot-carrier reliability," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 1014–1023, July 1996.

[28] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu, "The impact of gate-induced drain leakage current on MOSFET scaling," *International Electronic Devices Meeting*, pp. 718–721, 1987.

[29] C. Yang, Z. Wang, C. Tan, and M. Xu, "The degradation of p-MOSFETs under off-state stress," *Microelectronics Journal*, vol. 32, no. 7, pp. 587–591, July 2001.

[30] V. N. andN. C.Das, "Gate-induced drain leakage current in MOS devices," *Electron Devices, IEEE Transactions on*, vol. 40, no. 10, pp. 1888–1890, October 1993.

[31] A. Touhami and A. Bouhdada, "The n-type metal-oxide semiconductor field-effect transistor bias impact on the modeling of the gate-induced drain leakage current," *Semiconductor Science and Technology*, vol. 17, no. 12, pp. 1272–1277, December 2002.

[32] L. Huang, P. T. Lai, J. P. Xua, and Y. C. Cheng, "Mechanism analysis of gate-induced drain leakage in off-state n-MOSFET," *Microelectronics Journal*, vol. 38, no. 9, pp. 1425–1431, September 1998.

[33] T. Endoh, R. Shirota, M. Momodomi, and F. Masuoka, "An accurate model of subbreakdown due to band-to-band tunneling and some applications," *Elec-*

tron Devices, IEEE Transactions on, vol. 37, no. 1, pp. 290–296, January 1990.

[34] J. C. Guo, Y. C. Liu, M. H. Chou, M. T. Wang, and F. Shone, "A three-terminal band-trap-band tunneling model for drain engineering and substrate bias effect on GIDL in MOSFET," *Electron Devices, IEEE Transactions on*, vol. 45, no. 7, pp. 1518–1523, July 1998.

[35] J. H. Chen, S. C. Wong, and Y. H. Wang, "An analytic three-terminal band-band tunneling model on GIDL in MOSFET," *Electron Devices, IEEE Transactions on*, vol. 48, no. 7, pp. 1400–1405, July 2001.

[36] L. Lopez, P. Masson, D. Nee, and R. Bouchakour, "Temperature and drain voltage dependence of gate-induced drain leakage," *Microelectronic Engineering*, vol. 72, no. 1-4, pp. 101–105, April 2004.

[37] K. Rais, F. Balestra, and G. Ghibaudo, "Temperature dependence of gate induced drain leakage current in silicon CMOS devices," *Electronics Letters*, vol. 30, no. 1, pp. 32–34, January 1994.

[38] D. K. Slisher, R. G. Filippi, D. W. Storaska, and A. H. Gay, "Scaling Of Si MOSFET's For Digital Applications," Master's thesis, Rensselaer Polytechnic Institute, 1999.

[39] A. Bouhdada, S. Bakkali, and A. Touhami, "Modeling of gate-induced drain leakage in relation to technological parameters and temperature," *Microelectronics Reliability*, vol. 37, no. 4, pp. 649–652, April 1997.

[40] J. E. Chung, M. C. Jeng, J. E. Moon, P. K. Ko, and C. Hu, "Performance and reliability design issues for deep-submicrometer MOSFET's," *Electron Devices, IEEE Transactions on*, vol. 38, no. 3, pp. 545–554, March 1991.

[41] A. Bouhdada, A. Touhami, and S. Bakkali, "New model of gate-induced drain current density in an NMOS transistor," *Microelectronics Journal*, vol. 29, no. 11, pp. 813–816, November 1998.

[42] K. K. W. Liu, X. Jin and C. Hu, "BSIM 4.1.0 MOSFET model-user's manual," EECS Department, University of California, Berkeley, Tech. Rep., 2000.

[43] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *Journal of Applied Physics*, vol. 94, no. 1, pp. 1–18, July 2003.

[44] B. E. Deal, M. Sklar, A. S. Grove, and E. S. Snow, "Characteristics of surface-state charge (qss) of thermally oxidised silicon," *Electrochemical Society Journal*, vol. 114, no. 3, p. 266, 1967.

[45] A. Goetzberger, A. D. Lopez, and R. J. Strain, "Formation of surface states during stress aging of thermal si-si0_2 interfaces," *Electrochemical Society Journal*, vol. 120, no. 1, p. 90, 1973.

[46] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanisms to modelling," *Microelectronics Reliability*, vol. 46, no. 1, pp. 1–23, January 2006.

[47] S. S. Tan, T. P. Chen, H. A. Chew, and L. Chan, "Mechanism of nitrogen-enhanced negative bias temperature instability in pMOSFET," *Microelectronics Reliability*, vol. 45, no. 1, pp. 19–30, January 2005.

[48] H. Aono, E. Murakami, K. Okuyama, A. Nishida, M. Minami, Y. Ooji, and K. Kubota, "Modeling of NBTI saturation effect and its impact on electric field dependence of the lifetime," *Microelectronic Reliability*, vol. 45, no. 7-8, pp. 1109–1114, July-August 2005.

[49] S. Mahapatra, M. A. Alam, P. B. Kumar, T. R. Dalei, D. Varghese, and D. Saha, "Negative bias instability in CMOS devices," *Microelectronic Engineering*, vol. 80, no. 17, pp. 114–121, June 2005.

[50] K. O. Jeppson and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of NMOS devices," *Journal of Applied Physics*, vol. 48, no. 5, pp. 2004–2014, 1977.

[51] M. Ershov, S. Sexena, S. Minehane, P. Clifton, M. Redford, R. Lindley, H. Karbasi, S. Graves, and S. Winters, "Degradation dynamics, recovery, and characterization of negative bias temperature instability," *Microelectronics Reliability*, vol. 45, no. 1, pp. 99–105, January 2005.

[52] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectronic Reliability*, vol. 45, no. 1, pp. 71–81, January 2005.

[53] C. Schlunder, R. Brederlow, P. Wieczorek, C. Dahl, J. Holz, M. Rohner, S. Kessel, V. Herold, K. Goser, W. Weber, and R. Thewes, "Trapping mechanisms in negative bias temperature stressed p-MOSFET's," *Microelectronic Reliability*, vol. 39, no. 6-7, pp. 821–826, June-July 1999.

[54] S. S. Tan, T. P. Chen, and L. Chan, "Dynamic NBTI lifetime model for inverter-like waveform," *Microelectronics Reliability*, vol. 45, no. 7-8, pp. 1115–1118, July-August 2005.

[55] G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng, and D. L. Kwong, "Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling," *IEEE Electron Device Letters*, vol. 23, no. 12, pp. 734–736, December 2002.

[56] M. Agostinelli, S. Lau, S. Pae, P. Marzolf, H. Muthali, and S. Jacobs, "PMOS NBTI-indcued circuit mismatch in advanced technologies," *Microelectronics Reliability*, vol. 46, no. 1, pp. 63–68, January 2006.

[57] C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-electron-induced MOSFET degradation-model, monitor, and improvement," *Electron Devices, IEEE Transactions on*, vol. 32, no. 2, pp. 375–384, February 1985.

[58] G. Groeseneken, R. Bellens, G. V. den Bosch, and H. E. Maes, "Hot-carrier degradation in submicrometre MOSFETs: from uniform injection towards the real operating conditions," *Semiconductor Science and Technology*, vol. 10, no. 9, pp. 1208–1220, September 1995.

[59] K. R. Hofmann, C. Werner, W. Weber, and G. Dorda, "Hot-electron and hole-emission effects in short n-channel MOSFET's," *Electron Devices, IEEE Transactions on.*, vol. 32, no. 3, pp. 691–699, March 1985.

[60] T. H. Ning, P. W. Cook, R. H. Dennard, C. M. Osburn, S. E. Schuster, and H. N. Yu, "1μm MOSFET VLSI technology : Part IV - hot-electron design considerations," *Electron Devices, IEEE Transactions on*, vol. 26, no. 4, pp. 346–353, 1979.

[61] Y. Z. Chen and T. W. Tang, "Numerical simulation of avalanche hot-carrier injection in short-channel MOSFET's," *Electron Devices, IEEE Transactions on*, vol. 35, no. 12, pp. 2180–2188, December 1988.

[62] W. Li, S. Yuan, S. Chetlur, J. Zhou, and A. S. Oates, "An improved substrate current model for deep submicron MOSFET's," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 11, pp. 1985–1988, November 2000.

[63] X. Gao, J. J. Liou, J. Bernier, and G. Croft, "An improved model for substrate current of sub-micron MOSFET's," *Solid-State Electronics*, vol. 46, no. 9, pp. 1395–1398, September 2002.

[64] J. S. Kolhatkar and A. K. Dutta, "A new substrate current model for submicron MOSFET's," *Electron Devices, IEEE Transactions on*, vol. 47, no. 4, pp. 861–863, April 2000.

[65] M. M. Kuo, K. Seki, P. M. Lee, J. Y. Choi, P. K. Ko, and C. Hu, "Simulation of MOSFET lifetime under AC hot-electron stress," *Electron Devices, IEEE Transactions on*, vol. 35, no. 7, pp. 1004–1011, July 1988.

[66] X. Li and J. B. Bernstein, "CMOS failure modelling and reliability simulation part II : Hot carrier injection," *Device and Materials Reliability, IEEE Transactions on*, vol. 94, no. 8, pp. 1–6, August 2005.

[67] A. Bravaix, D. Goguenheim, N. Revil, E. Vincent, M. Varrot, and P. Mortini, "Analysis of high temperature effects on performances and hot-carrier degradation in DC/AC stresses 0.35μm n-MOSFET's," *Microelectronics Reliability*, vol. 39, no. 1, pp. 35–44, January 1999.

[68] S. H. Hong, S. M. Nam, B. O. Yun, B. J. Lee, C. G. Yu, and J. T. Park, "Temperature dependence of hot carrier induced MOSFET degradation at low gate bias," *Microelectronics Reliability*, vol. 39, no. 6-7, pp. 809–814, June-July 1999.

[69] T. C. Ong, P. K. Ko, and C. Hu, "Hot-carrier current modeling and device degradation in surface-channel p-MOSFET's," *Electron Devices, IEEE Transactions on*, vol. 37, no. 7, pp. 1658–1666, July 1990.

[70] B. L. Draper and D. W. Palmer, "Extension of high-temperature electronics," *Components, Hybrids, and Manufacturing Technology, IEEE Transactions on*, vol. 2, no. 4, pp. 399–404, December 1979.

[71] W. C. Nieberding and J. A. Powell, "High-temperature electronic requirements in aeropropulsion systems," *Industrial Electronics, IEEE Transactions on*, vol. 29, no. 2, pp. 103–106, May 1982.

[72] W. Kanert, H. Dettmer, B. Plikat, and N. Seliger, "Reliability aspects of semiconductor devices in high temperature applications," *Microelectronics Reliability*, vol. 43, no. 9-11, pp. 1839–1846, September-November 2003.

[73] F. V. Thome and D. B. King, "A summary of high-temperature electronics research and development," *AIP Conference Proceedings*, no. 9, pp. 254–259, October 1991.

[74] R. F. Jurgens, "High-temperature electronics applications in space exploration," *Industrial Electronics, IEEE Transactions on*, vol. 29, no. 2, pp. 107–110, May 1982.

[75] J. L. Prince, B. L. Draper, E. A. Rapp, J. N. Kronberg, and L. T. Fitch, "Performance of digital integrated-circuit technologies at very high-temperatures," *Components, Hybrids, and Manufacturing Technology, IEEE Transactions on*, vol. 3, no. 4, pp. 571–579, 1980.

[76] L. J. Palkuti, J. L. Prince, and A. S. Glista, "Integrated circuit characteristics at 260°C for aircraft engine-control applications," *Components, Hybrids, and Manufacturing Technology, IEEE Transactions on*, vol. 2, no. 4, pp. 405–412, 1979.

[77] F. S. Shoucair, W. Hwang, and P. Jain, "Electrical characteristics of LSI silicon MOSFET's at very high temperatures Part I and Part II," *Microelectronics Reliability*, vol. 24, no. 3, pp. 465–485 and 487–510, 1984.

[78] F. S. Shoucair and J. M. Early, "High temperature diffusion leakage-current-dependent MOSFET small signal conductance," *Electron Devices, IEEE Transactions on*, vol. 31, no. 12, pp. 1866–1872, 1984.

[79] F. S. Shoucair, "Scaling, subthreshold, and leakage current matching characteristics in high-temperature VLSI CMOS devices," *Components, Hybrids,*

and Manufacturing Technology, IEEE Transactions on, vol. 12, no. 4, pp. 780–788, December 1989.

[80] P. L. Dreike, D. M. Fleetwood, D. B. King, D. C. Sprauer, and T. E. Zipperian, "An overview of high-temperature electronics device technologies and potential applications," *Components, Packaging, and Manufacturing Technology, IEEE Transactions on*, vol. 17, no. 4, pp. 594–604, December 1994.

[81] S. L. Delage and C. Dua, "Wide band gap semiconductor reliability : Status and trends," *Microelectronics Reliability*, vol. 43, no. 9-11, pp. 1705–1712, September-November 2003.

[82] J. W. Klein, "Silicon and gallium arsenide in high temperature electronics applications," *International Symposium on Signals, Systems, and Electronics (ISSSE)*, pp. 63–68, October 1995.

[83] F. S. Shoucair and P. K. Ojala, "High-temperature electrical characteristics of GaAs MESFETS (25-400°C)," *Electron Devices, IEEE Transactions on*, vol. 39, no. 7, pp. 1551–1557, July 1992.

[84] C. Johnson, "Electronics and electronic materials for harsh environments, global watch mission report," Department of Trade and Industry, Tech. Rep., October 2003.

[85] K. Fricke, H. L. Hartnagel, R. Schutz, G. Schweeger, and J. Wurfl, "A new GaAs technology for stable FETS at 300°C," *Electron Devices, IEEE Transactions on*, vol. 10, no. 12, pp. 577–579, December 1989.

[86] R. Campbell, "Whatever happened to silicon carbide," *Industrial Electronics, IEEE Transactions on*, vol. 29, no. 2, pp. 124–128, May 1982.

[87] T. E. Zipperian, R. J. Chaffin, and L. R. Dawson, "Recent advances in gallium phosphide junction devices for high-temperature electronic applications," *In-*

dustrial Electronics, IEEE Transactions on, vol. 29, no. 2, pp. 129–136, May 1982.

[88] J. W. Kronberg, "High temperature behavior of MOS devices," *South East Con'81 IEEE Proceedings*, pp. 735–297, April 1981.

[89] D. Dufourt and J. L. Pelloie, "SOI design challenges," *Microelectronics Reliability*, vol. 43, no. 9-11, pp. 1361–1367, September-November 2003.

[90] T. Ouisse, G. Reichert, S. Cristoloveanu, O. Faynot, and B. Giffard, "Analysis of SIMOX metal-oxide-semiconductor transistors operated in the high temperature range," *Materials Science and Engineering*, vol. 29, no. 1-3, pp. 21–23, January 1995.

[91] D. Flandre, "Silicon-on-insulator technology for high-temperature metal-oxide-semiconductor devices and circuits," *Materials and Science Engineering*, vol. 29, no. 1-3, pp. 7–12, January 1995.

[92] J. Korec, "Silicon-on-insulator technology for high-temperature, smart power applications," *Materials Science and Engineering*, vol. 29, no. 1-3, pp. 1–6, January 1995.

[93] B. Gentinne, J. P. Eggermont, D. Flandre, and J. P. Colinge, "Fully depleted SOI CMOS technology for high temperature IC applications," *Materials Science and Engineering*, vol. 46, no. 1-3, pp. 1–7, April 1997.

[94] G. Reichert, C. Raynaud, O. Faynot, F. Balestra, and S. Cristoloveanu, "Submicron SOI-MOSFET's for high temperature operation (300-600K)," *Microelectronics Engineering*, vol. 36, no. 1-4, pp. 359–362, June 1997.

[95] A. Kloes, "Unified current equation for predictive modeling of submicron MOSFET's," *Solid-State Electronics*, vol. 49, no. 1, pp. 85–95, January 2005.

[96] *JEDEC Standard - JESD90 - A Procedure for Measuring P-Channel MOSFET Negative Bias Temperature Instability*, JEDEC Solid State Technology Association, 2004.

[97] Y. Cheng, M. C. Jeng, Z. Liu, J. Huang, M. Chan, K. Chen, P. K. Ko, and C. Hu, "A physical and scalable I-V model in BSIM3v3 for analog/digital circuit simulation," *Electron Devices, IEEE Transactions on.*, vol. 44, no. 2, pp. 277–287, February 1997.

[98] D. M. Kim, H. C. Kim, and H. T. Kim, "Modeling and extraction of gate bias-dependent parasitic source and drain resistances in MOSFET's," *Solid-State Electronics*, vol. 47, no. 10, pp. 1707–1712, October 2003.

[99] J. H. Huang, Z. H. Liu, M. C. Jeng, P. K. Ko, and C. Hu, "A Robust physical and predictive model For deep-submicrometer MOS circuit simulation," *IEEE Custom Integrated Circuits Conference*, pp. 14.2.1–14.2.4, 1993.

[100] Z. D. Prijic, "Analysis of temperature dependence of CMOS transistors' threshold voltage," *Microelectronics Reliability*, vol. 31, no. 1, pp. 33–37, 1991.

[101] R. Wang, "Threshold voltage variations with temperature in MOS transistors," *Electron Devices, IEEE Transactions on*, vol. 18, no. 6, pp. 386–388, June 1971.

[102] Y. Cheng, K. Imai, M. C. Jeng, Z. Liu, K. Chen, and C. Hu, "Modeling temperature effects of quarter micrometer MOSFET's in BSIM3v3 for circuit simulation," *Semiconductor Science and Technology*, vol. 12, no. 11, pp. 1349–1354, November 1997.

[103] *JEDEC Standard - JESD28-A - A Procedure for Measuring N-Channel MOSFET Hot-Carrier-Induced Degradation Under DC Stress*, JEDEC Solid State Technology Association, 2001.

[104] C. Hu, "Reliability phenomena under AC stress," *Microelectronics Reliability*, vol. 38, no. 1, pp. 1–5, January 1998.

[105] Winbond, "Quality and reliability report - reliability testing," Winbond Electronics Corp., Tech. Rep., 2006.

[106] M. C. Jeng, J. E. Moon, G. May, P. K. Ko, and C. Hu, "Design guidelines for deep-submicrometer MOSFETS," *IEDM Conference Proceedings*, pp. 386–389, 1988.

[107] V. Reddy, A. T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost, and S. Krishnan, "Impact of negative bias temperature instability on digital circuit reliability," *Microelectronics Reliability*, vol. 45, no. 1, pp. 31–38, January 2005.

[108] C. Hu, "Future CMOS scaling and reliability," *IEEE Proceedings*, vol. 81, no. 5, pp. 682–689, May 1993.

[109] P. Chaparala and D. Brisbin, "Impact of NBTI and HCI on PMOSFET threshold voltage drift," *Microelectronics Reliability*, vol. 45, no. 1, pp. 13–18, January 2005.

[110] E. Arkut and A. Ingolfsson, "Let's put the squares in least squares," *INFORMS, Transactions on Education*, vol. 1, no. 1, pp. 47–50, September 2000.

Appendix A

Experimental Setup

The experimental test rig used for device measurements is illustrated in figure A.1. It was designed and manufactured in-house to suit the specific testing requirements. A grounded metal base (I + H) was used to support the board (F) upon which the device package (G) was mounted. Dual in-line sockets were used to allow replacement of test packages. An aluminium heating block (E) was placed in contact with the underside of the package. The block was held in place with an adjustable screw (B) through a fixed bar (C + D). The heating block was milled to allow insertion of a heating element to provide a temperature range up to 155°C. Fixing screws (A) secure the board and heating components to the base.

Silicone heatsink compound was placed between the heating block and package underside surface to aid heat conduction. The package was not hermetically sealed to allow direct device temperature monitoring and measurement. Stable temperatures were ensured with regular infrared measurements. Additionally the package was held upside down to prevent the bond wires sagging at high temperatures and possibly causing short circuits. Shielded wires were used for all electrical connections.

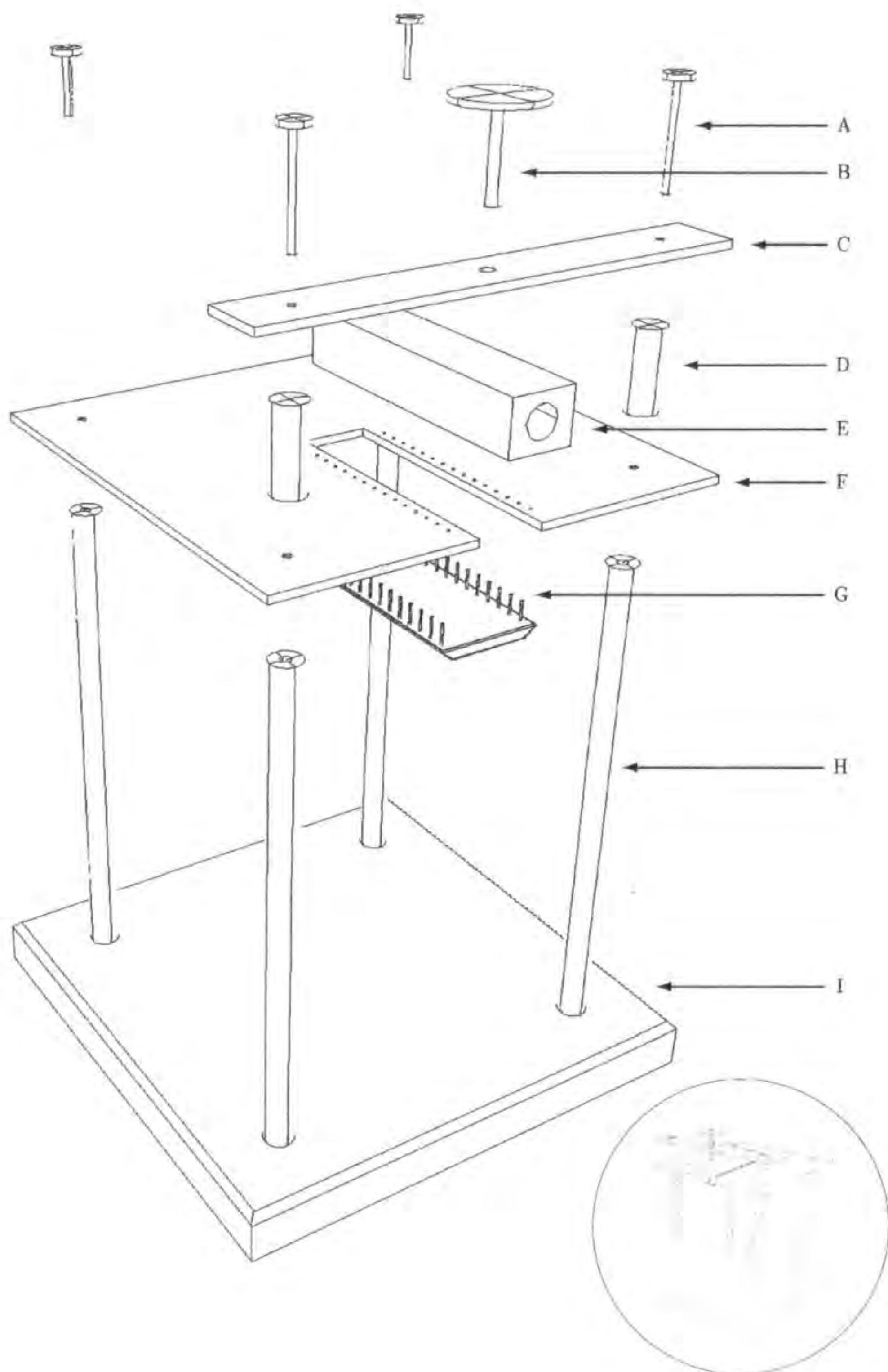


Figure A.1: Schematic diagram of test rig

Appendix B

Parameter Optimization

Contained within this Appendix is a example of parameter optimisation using the least squares technique. Determining model parameters is an essential part of the simulation process. Some tools such as those contained in MATLAB™ have been developed to help with parameter optimisation. One technique adopted within such tools and commonly used in curve fitting is least squares optimisation.

The least squares method seeks to minimise the sum of the square of the differences between a predefined functional model and a data set. In the case of a straight line the functional model is given by $y = mx + b$, the task is then to minimise the cost function $J = \sum((mx + b - data)^2)$ by selecting optimal values of m and b. These optimal parameter values are found by differentiating the function J with respect to m and b and setting these differentiated functions equal to zero, thus identifying the minima of the function J. MATLAB™ code to implement this method is shown, using the subthreshold function and parameters n and v_{off} as an example. Further details on least squares can be found in [110].

```
% MATLAB Parameter Optimisation                                      %
% Subthreshold Parameter Extraction                                  %
%                                                                    %
% ISub = Is0 * (exp((Vgs-Vth-Voff) \ (n*vtm)) * (1 - (exp(-Vds / Vtm)) %
%                                                                    %
% Where  Is0 = u * (W/L) * Cox * Vtm                                 %
%--------------------------------------------------------------------%

%-------Define constants
Vth = 0.66;
Vds = 0.1;
Vtm = 0.026;
Is0 = 1e-5;

%-------Define data
x = Data(:,1);
y = Data(:,2);
y = log(y);

%-------Plot measurement data
hold off
semilogy(x,exp(y),'o')
hold

%-------Define the number of data  points (n)
n = 36;
```

```
%-------Calculate data sums for least squares calculation
sumy = sum(y);
sumx = sum(x);
sumx2 = sum(x.^2);
sumxall = sumx^2;
sumxy = sum(x.*y);

%-------Least squares calculation
m = (n*sumxy - (sumx*sumy)) / (n*sumx2 - sumxall)
b = (sumx2*sumy - (sumx*sumxy)) / (n*sumx2 - sumxall)
f = m.*x + b;

%-------Plot optimised model
semilogy(x,exp(f),'r')

%-------Subthreshold swing (n) calculation
n = 1 / (m*Vtm)

%-------Subthreshold offset (Voff) calculation
c = 1 - exp(-Vds / Vtm);
Voff = (b-log(Is0*c) + (Vth / (n*Vtm))) * n * Vtm

%-------Subthreshold zero current (Ioff) calculation
Ioff = exp(b)
```