Modelling and analysis of failures in CMOS integrated cirucuits.

Simon Johnson

Modelling and Analysis of Failures in CMOS Integrated Circuits

Simon Johnson

Abstract

The research presented in this thesis is concerned with the failure mechanisms and faults that arise in CMOS VLSI circuits. The aim of this work is to demonstrate that currently accepted methods used for the testing of CMOS integrated circuits, are inadequate for a significant proportion of defects that can occur. Failure mechanisms that occur in these circuits are investigated. A number of faults that are not covered by current testing techniques are identified. These can be characterised by their tendency to cause changes in device and circuit parameters, rather than logical faults. The parameters that are commonly affected are transistor threshold voltages, logic gate propagation delays and quiescent supply current requirements.

One of the most common faults in this category occurs when a signal track on an IC becomes open circuit. This fault, which we term the *floating gate* fault, is investigated in detail. Analysis is presented to show that the fault represents a large proportion of all possible photolithography faults. Mechanisms which can lead to this fault occurring during the operation of ICs are identified. Experimental analysis of test devices and circuits indicates that the characteristics of this fault can result in non-logical errors in circuits. Standard stuck-at and stuck-open test techniques are shown to be inadequate for a large number of these failures.

A theoretical model of the floating gate MOS transistor is developed. This provides a physical understanding of the operation of the device. It also enables prediction of the characteristics of floating gate transistors for a wide range of device parameters. The model is used to develop a circuit simulation model. Analysis of a wide range of CMOS circuits is presented to demonstrate the effect of the floating gate fault on circuit operation. The fault is shown to be undetectable by stuck-at and stuck-open testing for a wide range of device parameters. The results of this analysis enable a testing method to be identified for the floating gate fault. To ensure detection of these faults, quiescent power supply monitoring is necessary in addition to thorough stuck-open testing. Provision of this method of testing can improve the detection of faulty circuits by three to four percent.

Modelling and Analysis of Failures in CMOS Integrated Circuits

Simon Johnson

This thesis is submitted to the University of Durham in candidature for the degree of Doctor of Philosophy

School of Engineering and Computer Science

University of Durham

1993

Declaration

The material contained within this thesis is the sole work of the author and has not been submitted previously for a degree in this or any other University.

Statement of Copyright

Acknowledgements

I would like to thank my supervisors, Professor Phillip Mars and Dr. Martin Morant, for their constructive criticism and careful reading of this thesis. I would also like to thank Martin for many helpful discussions on the work over the past years.

I must also thank my wife, Eleanor, for reading every word of this thesis, and, despite not understanding a word of it (she claims!), correcting my grammar impecably. The time that Eleanor and David and Rosie have given to this work will always be appreciated.

Contents

Chapter 1

Introduction 1

1.1 Aims of the Thesis 2

1.2 An Outline of the Thesis 2

Chapter 2

Failure Mechanisms and Fault Models 4

2.1 Terminology 4

2.2 Failure Mechanisms 5

2.2.1 Production Defects 5

2.2.2 Lifetime or Field Failures 7

2.2.3 Faults and the Electrical Effects of Production Defects 10

2.2.4 The Electrical Effect of Field Failure Defects 19

2.2.5 Summary of Failure Mechanisms 24

2.3 Fault Models 25

2.3.1 Classical Fault Models 25

2.3.2 CMOS Stuck-Open Fault Model 28

2.3.3 Parametric Fault Models 30

2.4 The Correlation Between Failure Mechanisms and Fault Models 33

2.5 Chapter Summary 38

Chapter 3

The Floating Gate Fault 39

3.1 Causes of the Floating Gate Fault 39

3.2 The Significance of the Floating Gate Fault 44

3.2.1 Layout Design Errors 44

3.2.2 Photolithography Defects 44

3.2.3 Photolithography Fault Distribution Experiments 46

3.2.3.1 Generation of Defects 46

3.2.3.2 Assessment of Defects 48

3.2.3.3 Defect Simulations 52

3.2.3.4 Results of the Simulations 57

3.2.3.5 Summary of Fault Distribution Experiments 58

3.2.4 Electromigration 58

3.2.5 Contact Migration 60
3.2.6 Stress Migration 60
3.2.7 Summary of the Significance of Floating Gate Faults 61
3.3 The Electrical Effect of the Floating Gate Fault 61
3.4 Test Circuits for the Analysis of the Floating Gate Fault 65
3.5 Chapter Summary 70

Chapter 4

Voltage Contrast Microscopy of Integrated Circuits 72
4.1 Scanning Electron Microscopy Analysis 72
4.1.1 The Principles of Voltage Contrast in the Scanning Electron Microscope 73
4.1.2 Quantitative Voltage Measurements using Image Contrast 76
4.1.3 Voltage Measurements using Secondary Electron Energy Analysis 77
4.1.4 Voltage Measurements on Buried Nodes 78
4.2 The Measurement System 81
4.3 Intensity Analysis for Quantitative Voltage Contrast 81
4.3.1 Voltage Contrast Intensity Measurements for the Contact Chain 83
4.3.2 Obtaining Track Voltages from Intensity Measurements 89
4.3.3 Conclusions on the use of Intensity Measurements for Qualitative Voltage Contrast 90
4.4 Voltage Measurements on Floating Gate Devices 90
4.5 Chapter Summary 91

Chapter 5

Experimental Analysis of Floating Gate Transistors 93
5.1 Analysis of Floating Gate Transistors 93
5.1.1 I_{DS} - V_{DS} Characteristics for Floating Gate Transistors 94
5.1.2 The Floating Gate Potential and the V_{FG} - V_{DS} Relationship 98
5.1.3 Measurements on p-channel Floating Gate Transistors 104
5.1.4 The Measured V_{FG0} Values and the Source of Residual Charge 105
5.2 Operation of Inverters with Floating Gate Faults 107
5.2.1 Output Characteristics for the Inverter Circuits 110
5.2.2 Floating Gate Voltages in the Inverter Circuits 112
5.3 Operation of Logic Gates with Floating Gate Faults 113
5.3.1 Problems with the Electrical Testing of the Logic Gates 113
5.3.2 Logical Operation of Gates with Floating Gate Faults 116
5.3.3 Floating Gate Voltage Measurements 117
5.3.4 Dynamic Testing of Logic Gates with Floating Gate Faults 119

5.4 Chapter Summary....120

Chapter 6

A Model Of The Floating Gate MOS Transistor....122

6.1 Existing Models....122
6.1.1 The Floating Gate Avalanche MOS Transistor....123
6.1.2 The Experimental Analysis of Circuits Containing Floating Gate Faults....125
6.2 Operation of the Faulty Floating Gate MOS Transistor....128
6.2.1 The Floating Gate Potential with Zero Drain Bias....129
6.2.2 The Floating Gate Potential with Applied Drain Bias....131
6.3 Refinements of the Model....134
6.4 Analysis of the Floating Gate MOS Transistor Operation....136
6.4.1 Floating Gate Transistor Charge Distributions....137
6.4.2 Channel Potential Distribution....139
6.5 An Algorithm for the Calculation of the Floating Gate Voltage....141
6.6 The Floating Gate MOS Transistor Model Program, FLGMOD....146
6.7 Some Basic Results of the Model....147
6.8 Chapter Summary....149

Chapter 7

CMOS Circuit Performance and Testing for Floating Gate Faults....150

7.1 Circuit Simulation Models For The Floating Gate Transistor....150
7.2 The Operation of Logic Gates with Floating Gate Faults....155
7.2.1 Static Operation of a CMOS Inverter....156
7.2.2 Dynamic Operation of a CMOS Inverter....160
7.2.3 Multiple Input Logic Gates....164
7.2.4 Summary of Logic Gate Operation with Floating Gate Faults....170
7.3 Transmission Gates and Other Circuits....172
7.3.1 Dynamic Circuits and Floating Gate Faults....175
7.4 Testing for Floating Gate Faults....176
7.4.1 The Stuck-At Fault Model....176
7.4.2 The Stuck-Open Fault Model....177
7.4.3 The Stuck-Closed Fault Model....178
7.5 The Implications of Inadequate Testing....180
7.6 Chapter Summary....183

Chapter 8

Summary and Conclusions.....184

8.1 A Summary of the Thesis.....184

8.2 The Conclusions of the Thesis.....188

8.3 Suggestions for Further Work.....189

8.4 Some General Conclusions.....190

Appendix A

Listing of the FLGMOD Program.....191

References.....203

Chapter 1
Introduction

The testing of integrated circuits is a problem that does not diminish as the complexity of devices continues to increase. Higher levels of integration result in an increase in the ratio of the number of transistors to input and output pins. This makes access to the core of the chip, and therefore testing, more difficult. The problem is increased by a raised level of concern over the reliability of VLSI circuits. Decreasing feature size and increasing circuit density, increase the probability of failure due to wearout mechanisms in such circuits. Thorough testing is important for improved reliability, as any undetected faults that remain on a chip after testing can result in early failure of a circuit in the field.

The ability of production tests to detect defects in CMOS integrated circuits (ICs) has improved as understanding of the failure mechanisms in these has increased. This has been one of the factors that has resulted in improved reliability of ICs [1.1], despite many other factors which should reduce the mean-time-to-failure for these circuits [1.2]. It would, however, be short sighted to assume that the current state of knowledge of failure mechanisms and testing techniques are sufficient to guarantee reliable circuits as the level of integration increases. Indeed, it may be insufficient to assume that current testing techniques are adequate even for today's circuits.

Improved reliability is one incentive for thorough testing of ICs. A second is the use of built-in-self-test on VLSI circuits. We have already seen that the testing problem increases as the level of integration rises. One solution to this problem is to build testing circuits onto the chip itself, thus overcoming the input / output bottleneck that VLSI circuits present in testing [1.3]. One of the requirements for such circuits is fully automated testing which should cover as many types of defect as possible and so we see a further need for thorough, and simply automated, testing techniques. Built-in-self-test circuits also provide a route to fault-tolerant ICs. Incorporation of redundant units or components and a comprehensive test system can allow a circuit to continue to operate correctly, even with many faults present [1.4].

In extending the ability of tests to detect physical defects, it is becoming necessary to think beyond simple logical faults. Timing faults [1.5], which only cause a circuit to fail above (or below) a certain operating speed, are an example of this extension from logical to *parametric* faults. It is no longer sufficient to use the logical function of a circuit as proof that it is free from faults. The use of other parameters, such as timing or supply current [1.6] is seen as a necessity to cover the range of subtle faults that exist in complex VLSI circuits.

1.1 The Aims of this Thesis

The aim of this thesis is to demonstrate that current testing techniques for CMOS VLSI circuits are inadequate for a significant proportion of the physical defects that can occur. We intend to demonstrate that a particular class of fault, the *floating gate fault*, is not detected in a large number of cases by the tests commonly used for CMOS circuits. Furthermore, it will be shown that such faults represent a relatively large fraction of the range of possible faults that can occur.

In order to understand the effects of the fault, test structures have been designed and fabricated. A theoretical model of the faulty device will be developed in this thesis to provide an understanding of the floating gate transistor. This will enable the effects of the fault on circuit operation to be established.

Finally we will demonstrate that current logical testing techniques do not reveal the fault in a large number of cases. Consequently, a testing strategy to improve the coverage of the floating gate fault must be identified. It is the aim of this work to improve CMOS fault coverage for both production test and for future built-in-self-test systems. In this way it is hoped that a contribution can be made to improving integrated circuit reliability and the goal of the completely fault tolerant IC.

1.2 An Outline of the Thesis

The thesis starts with a survey and investigation of failure mechanisms in CMOS circuits. The relationship between the physical defects that can occur and the resulting faults is established. The effect of such faults on device and circuit performance is analysed and the failures that may result are identified. A distinction is made between circuit failures resulting from production defects and those that occur through intrinsic mechanisms over the lifetime of a component. Although different sources of physical defects exist, many of the resulting faults and failures are seen to be similar. The details of failure mechanisms are generally abstracted into *fault models* which can be used to simplify the generation of tests. The adequacy of such models for describing defects is investigated. We demonstrate that some defects are poorly represented by these models and that such defects are not uncommon. It is clear that certain forms of failure require more detailed investigation.

In Chapter 3, the most significant fault that is not adequately represented by existing models is investigated. This fault arises when an open circuit occurs in a signal line on a CMOS chip. This results in the gates of any transistors driven by this signal becoming isolated. We refer to this as a *floating gate fault*. The processes which give rise to this form of failure are identified and the significance of the fault is quantified by analysis of defect distributions in typical cell layouts. The effect of this type of fault is outlined and the need for experimental analysis is established. Test structures for the investigation of the fault are then described.

Chapters 4 and 5 describe the analysis of the test devices that have been fabricated for this work. In Chapter 4, the use of the voltage contrast technique in the scanning electron microscope is investigated as a method of obtaining floating gate device characteristics. This is found to be inadequate to provide the information required. Electrical measurements on floating gate devices and circuits are described in Chapter 5. The characteristics of floating gate transistors are established and various device parameters investigated. Two significant parameters associated with the device are identified. The effect of floating gate faults on simple CMOS circuits is analysed. Wide variations in the significance of the effect of the fault on circuit operation is observed.

Chapter 6 describes the development of a theoretical model for the floating gate MOS transistor. This is done to provide a physical understanding of the operation of the device, which is essentially a normal MOS transistor with a fixed gate charge. A computer program which has been used to calculate various device characteristics is described, and the predictions of the model are compared with the measurements reported in Chapter 4.

In Chapter 7 the results of the experiments and theoretical model are used to analyse the effect of floating gate faults on CMOS circuits. A simulation model is developed to assist in the analysis. The static and dynamic operation of a range of logic gates is investigated. The conclusions that are drawn are used to identify testing techniques for floating gate faults and a recommendation is made for satisfactory testing of these faults. Finally, the implications of inadequate testing for floating gate and other faults is discussed.

We conclude the thesis in Chapter 8 with a summary of the work and consideration of the significance of the results. Some suggestions for further research in this area are also made.

Chapter 2
Failure Mechanisms and Fault Models

Failure mechanisms in integrated circuits are generally well understood and electrical models of the effect of failures have been established for many years. The dominance of CMOS as the technology for VLSI circuits has led to a refinement in the models used for fault analysis. This has occurred because the electrical effect of some faults is not accurately represented by the more traditional fault models. Some failure mechanisms, however, are still incorrectly modelled in CMOS resulting in inadequate test generation and fault coverage. Consequently, circuits may pass stringent testing procedures and still contain significant faults which can lead to failures in the operation of devices.

In this chapter we will review the current understanding of defects as it is described in the literature. From this we will establish the most common forms of defect that arise in CMOS IC's, both from production and during operation of the circuits. We will then analyse the effects of these defects on circuit performance and investigate the effectiveness of existing fault models for detecting the resulting faults. It will be shown that some failure mechanisms produce faults that may not be detected by existing fault models. The aim of later chapters will be to show how this situation can be improved by the addition of fault models for a particular class of defect - the open circuit gate.

2.1 Terminology

Various terms are used to describe faults in electronic circuits and these are not always used consistently by different authors. It is therefore worthwhile defining the terms that will be used in this thesis.

Defect: the physical deviation of a feature from the required form. For example, a gap in a feature or a deviation from the specified width of a feature.

Fault: the effect of a defect on a circuit or component. For example, a defect such as *a gap in a feature* may lead to the fault of *an open circuit track*; a defect such as *an oxide pinhole* may lead to the fault of a *conductive path from gate to channel in an MOS transistor.*

Fault effect: the effect of a fault on a component or circuit. For example, a fault which creates *a conductive path between the gate and channel of an MOS transistor* will have the effect of *decreasing the charge storage time on the gate.*

Failure: incorrect operation of a component or circuit due to the presence of a fault. For example an *open circuit track* may lead to *the failure to discharge a circuit node when required within the specified time.*

Failure mechanism: the process by which a failure occurs in a component or circuit. This term encompasses all of the above terms i.e. a defect must be present to cause a fault in a component. This may affect the performance causing a failure to occur.

It is important to note that a defect will not always result in a fault occurring in a circuit and that the consequential effect of that fault may not lead to a failure of the device or circuit.

2.2 Failure Mechanisms

Failures in an integrated circuit may arise during the fabrication of the IC or during its operational lifetime. Failures that occur during fabrication result in a less than perfect yield figure for the circuit and should be detected during the many testing procedures that take place during the fabrication and packaging of the IC. Failures occurring during the operational lifetime of the IC constitute a reliability hazard for the component. They will probably be detected by the failure of the complete system containing the device. Alternatively, failures may be detected during circuit operation by built-in-test circuits if the latter have been included on the circuit. In both cases the result is a system failure unless fault tolerance has been built into the system or device.

The defects which can occur in production are generally different from those that occur during the operation of the device. We will now consider the two cases.

2.2.1 Production Defects

The term "production failures" is used to indicate any failure that may occur in the IC before it leaves the manufacturer. The defects that cause these failures may arise from

incorrect processing, or contamination of the wafers or chemicals that are used. A defect may lead to a fault in a circuit and this will generally have an effect on the electrical performance of the circuit.

The effects of physical defects that can arise in integrated circuits have been investigated by many workers. Fewer papers have been published on the actual defects that can occur. However, one interesting paper in this area by Fantini and Morandi [2.1] includes failure data from a range of sources. Defects are classified into the following categories: i) crystal defects; ii) charge trapping in the gate oxide; iii) gate oxide pinholes; iv) contact surface defects; v) conductor thinning, contamination, and hillock formation; vi) poor wire bonding; vii) poor die attach. These defects all result from imperfections in the fabrication process generally associated with the material layer or surface preparations. The authors do not mention photolithographic errors due to particulate contamination, or uneven etching or deposition over a wafer. The authors identify some fault effects which do not have a corresponding physical defect. These include random ionising radiation events, electrostatic discharge damage and hot carrier problems. These will be discussed in a later section when fault effects are considered.

Another paper which discusses defects, Burgess et al [2.2], is based on analysis of failed NMOS circuits from a British Telecom production line. Many of the defects illustrated arise from photolithographic errors caused by particulate contamination, the source of defect found to be insignificant by Fantini and Morandi. From this we may surmise that the distribution of production defects among the various classifications can be quite specific to a production line. Consequently, a single source of information on IC defects is not a sufficient basis for the formation of general conclusions.

In another paper [2.3], the same authors identify a list of MOS faults which includes device length and width variations. The defects that are most likely to give rise to these faults are uneven etching or poor photolithography. Further support for particulate contamination leading to photolithographic related defects is provided by Kemp, Poole and Frost [2.4]. In this paper the effect of track narrowing due to particulate contamination during photolithography is investigated. This defect is found to result in early lifetime failures of metal tracks resulting in a reliability hazard. This is an example of a production defect leading to a latent fault which subsequently causes failure during circuit operation.

Other papers on the subject of physical failures and defects ([2.5], [2.6], [2.7], [2.8]) do not reveal any further categories of defect. The following list summarises the defects identified in the literature survey described above. All of the defects listed may occur in a standard CMOS process.

i) *Crystallographic defects*: these appear in the substrate regions of transistors. The defects result in incomplete bonding within the crystal structure producing

sites for charge trapping and for the accumulation of dopants. These can lead to increased substrate and leakage currents.

ii) *Particulate contamination:* this causes errors in the photolithographic process used to define features in the layers of the integrated circuit. Particles may result in additional or missing material on a layer causing malformed features or the complete absence of small features, such as interlayer contact cuts.

iii) *Surface contamination:* this can result in poor definition of features and contamination of interfaces such as interlayer contacts. This defect is particularly difficult to identify as the surface contaminant often disappears in subsequent processing.

vi) *Oxide defects:* inhomogeneous growth of thermal oxide can result in "pin holes" in oxide layers, particularly in thin gate oxides. The holes, although very small (typically less than 0.01μm), can have a serious effect on circuit performance. The gate oxide thickness may also deviate from the required value altering transistor characteristics. Charge trapping in oxides can also occur resulting in mirror charges in device structures.

v) *Packaging defects:* these can include a) *poor wire bonding,* resulting in open circuit input / output pins or short circuits between adjacent pins; b) *poor die attach,* leading to broken bonds during lifetime and overheating of the circuit; c) *damage to package (e.g. cracks),* which can result in non-hermeticity giving rise to corrosion of bond wires and other circuit layers.

2.2.2 Lifetime or Field Failure Defects

The term "lifetime failure" or "field failure" is used to indicate any failure that may occur in a device once it has been put into service in a system or simply stored on a shelf after purchase from a supplier. Many of the mechanisms associated with lifetime failures, such as electromigration, occur over an extended time period while others, such as electrostatic discharge damage or ESD, happen in a very short space of time. Some mechanisms are more common at the start of a component lifetime while others occur more frequently after many years of operation.

Information on the causes of field failures of integrated circuits appears only occasionally in the literature. It is often associated with specific components or environments and requires careful scrutiny to extract any general information. Many papers on the subject of reliability failure mechanisms therefore base their analysis on theoretical consideration of materials and device structures, although other papers do discuss field failure data.

One of the earliest papers with emphasis on MOS circuits is by Courtois [2.9]. The author identifies six categories of phenomena that can result in field failures of ICs. The categories are based on a classification issued by the International Council of Scientific Unions [2.10] which has not apparently been used by other workers in this field. The field failure categories identified by Courtois are:

i) *Transport phenomena* or the movement of atoms, impurities or faults within a structure. This phenomena is generally referred to as electromigration and Courtois notes that the typical consequences are hillocks and voids and that step-coverage and misalignment are the primary causes;

ii) *Surface growth* such as dendrites and whiskers. This is initiated by electrochemical reactions and is primarily a problem confined to devices with fusible links (such as programmable read only memory);

iii) *Dielectric breakdown.* Courtois describes this as resulting from the movement of positive ions within the oxide which accumulate and hence increase the electric field locally to the point at which the intrinsic electric field breakdown threshold for the oxide is reached. The process is time and temperature dependent.

iv) *Conductor breakdown* can occur due to electrical overstress and results from excess current flow which causes fusing of metal tracks. ~~Electromigration is not discussed by Courtois.~~

v) *Thermomechanical effects* result from the differing thermal expansion coefficients of layers in the IC structure and cause fractures and stresses in various layers of the circuit.

vi) *External agents.* Courtois groups together all extrinsic failure mechanisms and latent instrinsic failure mechanisms as *external agents.* These include oxide contamination and pinholes (already discussed). Extrinsic factors include radiation (causing *soft errors*) and vibration (causing mechanical failure of bonds and die attach). Chemical corrosion is also given as an external agent which can cause failures. This is caused by high humidity which can occur as a result of non-hermeticity of packages.

A paper by Woods [2.11] identifies four primary life-time failure mechanisms, two of which (radiation induced soft-errors and oxide breakdown) are covered by Courtois. Woods extends the category of *conductor breakdown* to include electromigration. This is an intrinsic form of defect that occurs in polycrystalline aluminium tracks and will be discussed in more detail in a later section. Woods notes that *contact electromigration* is activated by somewhat different mechanisms from track electromigration and treats it as a separate though related failure.

A new failure mechanism described by Woods is *hot carrier degradation.* This effect becomes more significant as device dimensions are reduced and electric fields within devices increase. It is understandable that in 1980 Courtois did not consider this effect to be significant but by 1986 hot carrier injection was generally recognised as a problem, particularly as device dimensions are scaled. Hot carrier (or more specifically hot electron) effects have been noted by many workers in the field of MOS circuit reliability. A recent book on the subject [2.12] emphasises the significance of the effect on current and future sub-micron MOS devices.

Woods also identifies *electrical noise* on ICs as a source of failure. The noise may originate from the circuit itself or from the external environment. Internal sources of noise include capacitive cross-coupling of signals, power supply fluctuations, radiation generated or thermal noise. External sources include electromagnetic interference and power supply noise. These factors are rarely considered in the literature. Electrostatic discharge damage, a specific case of electrical over-stress, is also noted as a source of field failure in MOS ICs.

The previous papers provide useful theoretical analysis to help to predict the probable sources of defects in CMOS IC's. An alternative source of information is field failure data which is discussed by several authors. Boulaire et al [2.13] present field failure data from ICs operating in telecommunications equipment. Three of the four mechanisms described as being of primary concern have been noted in the previous papers (Electrical over-stress and electrostatic discharge damage, electromigration and corrosion).

The new failure found from analysis of failed ICs returned from operations is the loss of program information in EPROM memory circuits. The causes of the EPROM failures are discussed in the literature (for example Crisenza [2.14]). While this thesis is concerned with failures in standard CMOS IC's, the EPROM failures do provide some relevant information. One cause of the failures in dielectric breakdown and this is noted by Hawkins and Soden [2.15] for more standard ICs. Latent defects have been identified as the dominant failure mechanism in radiation hard CMOS ICs fabricated on a particular production line.

The defects identifed by Boulaire (and by other field failure analysis such as Sinnadurai [2.16]) confirm, to a large extent, the conclusions presented by Courtois and Woods. However, the conclusions on the significance of the defects do differ. The field failure data that is presented indicates that corrosion is one of the most significant defects that gives rise to field failures. Electrical overstress is also a major concern. In the more theoretical analyses, the effects that are considered to be of primary importance are oxide breakdown, electromigration and hot electron effects. These two conclusions are not contradictory as the field failure data is, by its nature, historical (dating back to 1980 for Boulaire's earliest samples) and the effects currently being discussed in the literature become more significant with reduced device dimensions [2.11]. A further factor, as we shall see in

the next sections, is that the latter set of defects do not always cause catastrophic failures of circuits and are therefore more difficult to identify in failure analysis.

2.2.3 Faults and the Electrical Effect of Production Defects.

The electrical effects of the defects identified in the previous sections are wide ranging. The defects may lead to a total malfunction of the chip or to a slight shift in a transistor parameter. The electrical effect of the defect depends on various factors: the type of defect, the extent of the defect, its physical location on the circuit and its position within the electrical configuration of the surrounding components. The electrical effect of faults has been extensively investigated and reported in the literature.

We will now consider the faults arising from each of the production defects described and the electrical effect of these faults. The defects to be considered are:

i) Crystal defects
ii) Particulate contamination
iii) Surface contamination
iv) Oxide defects
v) Packaging defects

i) Crystal defects. These give rise to fault effects located in the device substrates. Field data on integrated circuits from a range of technologies (NMOS, CMOS, bipolar) presented in [2.1] indicate that only a small percentage (between 0 and 4%) of failures can be attributed to this kind of fault.

Defect: Crystallographic faults, e.g. stacking faults.

Fault: An accumulation of interstitial and lattice dopants in the region.

Fault effect: High substrate currents which can appear as junction leakage currents if they are located within a *p-n* junction.

Failure: The effect of leakage is not always apparent. The most serious result is that the refresh time of dynamic circuits, and in particular memory circuits, may be significantly shortened. This can lead to a logical error if the refresh time cannot be reduced. Increased substrate current leakage can also result in an increased quiescent supply current requirement which can be a serious problem in IC's required for use in low power or standby systems. Furthermore, it is feasible that the increased substrate currents due to crystal defects may raise the substrate potential sufficiently in critical regions so that latch-up can occur.

Strack et al [2.17] have made measurements on MOS capacitor structures with oxidation induced stacking faults. These indicate an increase for the reverse saturation current for a particular *p-n* junction of two orders of magnitude. While the measurements were for a large area diode with a fault free current of 10^{-10}A they do indicate the relative increase that might be expected in the current. From these results it can be seen that the increase in substrate current due to crystal defects, is insufficient to create a significant substrate potential where the relevant resistances are of the order of kilo-ohms.

ii) ***Particulate contamination.*** The fault effects produced by particulate contamination during photolithography or layer deposition are varied. They depend on the layer or layers affected, and on the position and size of the defect with respect to device features. We will consider some examples of particulate contamination and attempt to draw some general conclusions from these.

Generally, particulate contamination during photolithography or layer deposition will result in additional or missing material. Missing material can result in an open circuit in, for example, a track in active, polysilicon or metal. For an open circuit to occur the defect must completely cross a track. Alternatively, a defect which leaves a thin section of a track remaining might result in a high series resistance in a track which would normally have a negligible resistance. The effect of this fault could be to increase an RC time constant associated with a track which will slow down the circuit response and cause a failure.

Figure 2.1 illustrates the effect of a circular particle at a few example locations on a typical CMOS cell. At this point we only wish to illustrate the range of possible defects that may arise. More detailed analysis of particulate contamination is discussed in Chapter 3. The cell used for the current analysis is a transparent D-latch circuit fabricated in a 1.5μm *n*-well technology. The circuit of the cell is shown in figure 2.2. A D-type has been chosen as it is representative of many cells that occur on a typical IC. Also, it is more complex in its operation than a simple logic gate and illustrates some of the subtle effects that can occur. The choice of circular particles is quite arbitrary but serves to illustrate the effect of particulate contamination on CMOS circuits and devices.

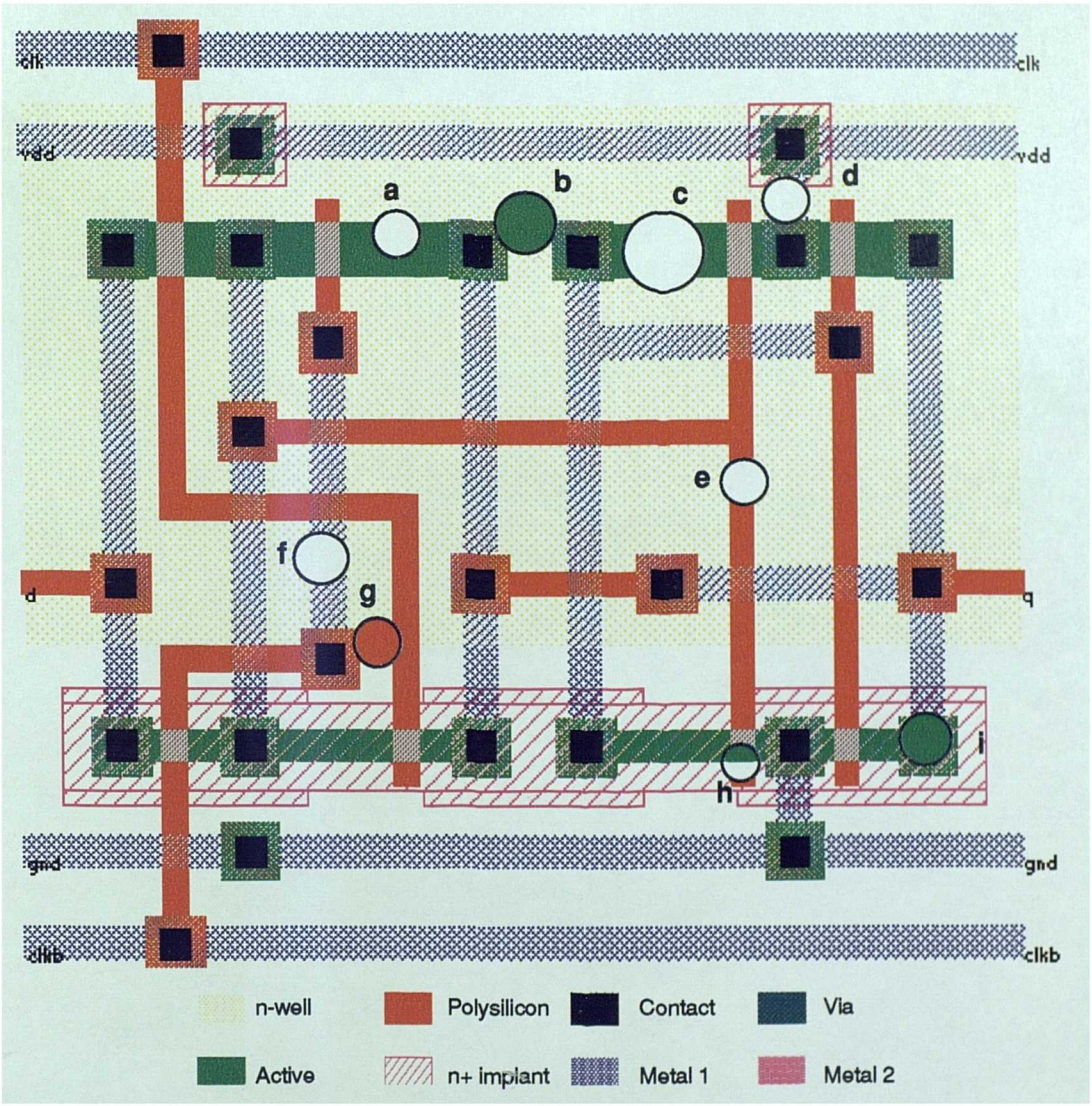


Figure 2.1 Particulate contamination of a D-latch cell

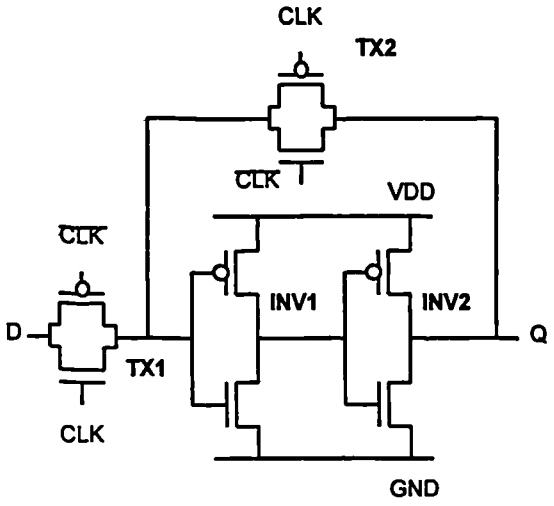


Figure 2.2 The D-latch circuit

The failures caused by the defects labelled (a) to (i) on Figure 2.1 are now described in order to show the wide range of effects that may occur.

(a) **Defect:** Narrowing of active track.

Fault: High resistance in drain/source of *p*-channel transistor in TX2 (transmission gate 2)

Fault effect: There is little effect on the circuit operation even if the fault becomes an open-circuit. In this case the fault makes the input of INV1 (inverter 1) dynamic for the voltage range V_{DD} to V_{DD} - V_{Tn} (where V_{Tn} is the *n*-channel threshold voltage) because the transmission gate acts as a simple *n*-channel pass transistor with its inherent V_{Tn} drop when $V_{DS} = V_{GS}$. Therefore the input of INV1 is not driven in the voltage range V_{DD} to V_{DD} - V_{Tn} and hence the circuit is acting dynamically. This would not be a problem under normal operating conditions, for even if the inverter input voltage decayed to V_{DD} - V_{Tn} the *n*-channel device in TX2 would switch on to prevent further discharging ensuring that a logic 1 appears at the input to INV1 as required.

Failure: No failure occurs.

(b) **Defect:** Additional active region causing bridge between two areas.

Fault: Low resistance path between input and output of INV2.

Fault effect: The output high voltage for the latch will be less than V_{DD} and the output low voltage will be greater than 0V. If the inverters are designed for symmetric operation (same rise and fall propagation delays) then the output voltage would be close to $V_{DD}/2$ in both cases. The actual output voltage would be determined by the relative resistances of the bridging resistance and the device channels.

Failure: A timing failure may occur if the bistable is in a critical path. If the resistance of the bridge is sufficiently low, the output may appear to be stuck at logic 0 or logic 1.

(c) **Defect:** Missing active region causing break in a track.

Fault: Open circuit in the drain of the *p*-channel transistor in INV2.

Fault effect: The Q output of the latch cannot be charged to V_{DD} through the *p*-channel device. However, TX2 provides a possible path for charge sharing from the input of INV1. The output high voltage for Q will be less than V_{DD} but may be sufficient to appear as a logic 1.

Failure: Possible logic fault at the latch output - may appear to be stuck at logic 0.

(d) **Defect:** Missing metal causing a break in a track.

Fault: Open circuit in the V_{DD} supply to the inverters.

Fault effect: The latch output cannot be charged to logic 1. Charge sharing may occur as described for defect (c).

Failure: Possible logic fault at the latch output - may appear to be stuck at logic 0.

(e) **Defect:** Missing polysilicon causing a break in a track.

Fault: Unconnected gate of *n*-channel transistor in INV1

Fault effect: Unknown gate voltage and conductance of *n*-channel transistor.

Failure: The transistor may be in a conducting or non-conducting state. Capacitive coupling from adjacent tracks or circuit nodes may cause gate potential to vary. Failure mode is difficult to predict.

(f) **Defect:** Missing metal causing break in a track.

Fault: Open circuit in the gate terminal of the *p*-channel transistor in TX2.

Fault effect: The gate terminal is unconnected resulting in an indeterminate conductivity for the device. The transistor may remain open circuit resulting in the extreme case described for defect (a).

Failure: No failure occurs.

(g) **Defect:** Extra polysilicon causing bridge between two tracks.

Fault: Low resistance path between the CLK and CLKB (the inverted clock signal).

Fault effect: The CLK and CLKB signals will be at the same potential of approximately $V_{DD}/2$ resulting in both transmission gates being permanently on.

Failure: The latch will be permanently transparent with the CLK signal appearing to be stuck at logic 1.

(h) **Defect:** Missing polysilicon resulting in shortened gate track.

Fault: Drain-source short circuit on the *n*-channel transistor of INV1.

Fault effect: Low resistance of the drain-source region will prevent the output high voltage from reaching V_{DD}. The extent of the degradation will depend on the width of the short circuit region.

Failure: The latch output may be stuck at logic 1.

(i) **Defect:** Missing contact cut.

Fault: Drain connection missing on the *n*-channel device of INV2.

Fault effect: When the INV2 output should discharge to 0V it will only do so through charge sharing with the INV1 input node through TX2. It is likely that the latch output node will have a larger capacitance than the INV1 input and therefore the output will stay close to V_{DD}.

Failure: The output will possibly be stuck at 1.

The faults that have been described are a random selection, but illustrate the wide range of fault effects and failures that can arise from "simple" particulate contamination. The defects all affected a single layer and yet the same defect in two locations can produce very different failures. Even if the faults produced by the defects are the same, the resulting failures can be very different. Consider, for example, defects (a) and (c). An extreme case of (a) causes an open circuit in the drain of the TX2 *p*-channel transistor, the same fault as (c), but in a different device. In the case of defect (a) no failure is detected by electrical testing. For defect (c) the fault may be detected by a functional failure as the output appears to be stuck at logic 0 irrespective of the D input value.

From this example it is clear that an analysis of the effect of particle defects on circuit performance cannot be undertaken without considering the topology of the circuits concerned. This, however, is a time consuming task. To reduce the complexity of the problem, the failures caused by defects can be abstracted to simpler "fault models" which represent the electrical effect of a defect.

From the above brief discussion, it is apparent that the effectiveness of simple models in describing failures must be questioned, as the situation is by no means simple. This is the main subject of this thesis and will be discussed in depth in later sections.

iii) Surface contamination. The defects that arise from surface contamination will tend to be specific to a process line. They will primarily depend on the cleaning and layer removal processes that are used. Some examples are given below.

(a) **Defect:** Regions of chemical residue on the wafer surface.

Fault: Poor adhesion of photoresist or deposited layers.

Fault Effect: Missing or extra material within or between features on a layer. The faults effects are therefore very similar to those arising from particulate contamination.

Failure: As for particulate contamination.

(b) **Defect:** Poor layer removal, for example, metal oxides in contact windows.

Fault: High resistance ~~or rectifying contacts~~.

Fault effects: Increase in contact resistance~~, possibly for only one direction of current flow~~. If this occurs in a signal line, ~~one of the signal transitions (rising or falling) may be more significantly affected than the other, leading to~~ a timing defect may occur. ~~on one signal transition~~.

Failure: Possible timing failure if the contact resistance is high enough.

iv) Oxide defects. Imperfections in oxide layers can take several forms resulting in low resistance paths, ionic contamination and layer thickness variations. The fault effects caused by oxide defects are often parametric rather than logical. The various forms of defect will now be discussed.

(a) **Defect:** "Pin-hole" in gate oxide over channel.

Fault: Low resistance path from gate to channel. The resistance may be terminated by a junction diode in the channel region due to dopant diffusing from the polysilicon gate. The formation of the diode will depend on the doping received by the substrate and the polysilicon which may or may not be of the same type.

Fault effect: The effect of this fault depends on the position of the defect with respect to the drain and source, and on the resistance of the conducting path between channel and gate. The first parameter will determine whether the gate potential is

more strongly influenced by the drain or source potential. The resistance of the short will determine the strength of the interaction of drain/source and gate potential. For example, if the defect creates a very low resistance path close to the drain end of the channel then the gate potential will be close to the drain potential when a channel exists. This assumes that the impedance of the source driving the gate (i.e. another logic gate) is high compared to the resistance of the short. A similar effect will occur if the defect is close to the source. If the resistance of the oxide short is high then the channel potential at the defect location will have less effect on the gate voltage.

In most circuit configurations the fault will introduce negative feedback into a logic circuit. The effect of negative feedback generally in logic circuits is to induce "weak" output voltages for logic levels. The effect depends on the strength of the feedback, but it does not necessarily result in logical errors or even timing errors.

SPICE simulations have been performed using a simple model of a gate oxide short as shown in Figure 2.3. Results show that even for negligible resistance of the gate oxide short (as these defects are often called), the worst case feedback only results in weak logic levels that are restored to full strength after two subsequent gates. From these results it can be seen that pin hole defects in gate oxide do not generally result in significant fault effects. Gate oxide shorts have been modelled in detail by Syrzycki [2.18]. The sophisticated model used in this work confirms the results obtained using the simple model shown in Figure 2.3.

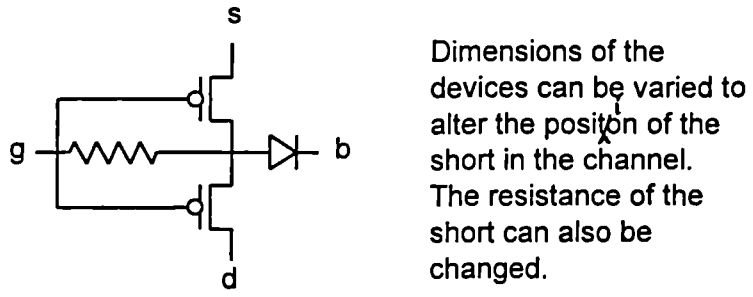


Figure 2.3 A model for a p-channel device with gate oxide short used in SPICE simulations.

A further effect of this form of defect is also noted by Syrzycki. The quiescent power supply current to circuits containing gate oxide shorts is significantly increased above the normal leakage current. Two paths exist for this current:

i) current flow between the supply rails (not noted by Syrzycki) due to negative feedback to the gate that will cause both *n*- and *p*-channel networks of the faulty logic gate to be conducting for one input state (i.e. logic 0 or logic 1);

ii) current flow from the defective device channel, through the gate oxide short to the gate of the driving logic circuit and hence to the appropriate supply rail through the driving gate network.

Raised quiescent supply current is also reported by Hawkins and Soden [2.19] although more in the context of reliability failures. The increased current can be of the order of 100 μA or more and so is significant in the operation of low power CMOS circuits.

Failure: Excess quiescent power supply current for some input conditions. Timing errors may occur if logic gates are in critical delay paths.

(b) **Defect:** Pin hole in gate oxide over source or drain overlap region.

Fault: Low resistance path from gate to drain or source. The resistance may be terminated with a junction diode as for defects over the channel.

Fault effect: The effect is similar to that described for the pin hole defect located over the channel region, the main difference being that the channel does not have to be present for the negative feedback to occur. The strength of the feedback is again determined by the resistance of the gate oxide short.

Failure: Timing failures may occur. The quiescent power supply current will be increased for some input conditions.

(c) **Defect:** Ionic contamination in gate and thick oxides. This form of defect is generally thought to be insignificant but is still reported as the source of failures in recent papers (for example [2.1]).

Fault: Ionic contamination will cause a shift in the threshold voltage of a device from the required value if the contamination occurs in gate oxide. Contamination in field oxide can result in parasitic MOS transistor formation.

Fault effect: Threshold voltage variation will cause devices to switch at different gate voltages resulting in changes in propagation delays for logic gates. Transistors may also be conducting when required to be off causing excess current flow.

Failure: Timing failures may occur if threshold voltage shifts are significant. Also excess quiescent power supply current will flow for some input combinations.

(d) **Defect:** Gate oxide thickness variations.

Fault: MOS transistor threshold variation.

Fault effect: Effects are as for ionic contamination in gate oxide.

Failure: Timing failures possible and increased quiescent power supply current may occur for some input combinations.

v) Packaging defects. Packaging defects arise as a result of production errors and should be detected by the many test procedures that are executed before an IC leaves the production plant. They generally give rise to gross faults which have a catastrophic effect on circuit performance. The defects can be categorised as follows.

(a) **Defect:** Poor wire bonding.

Fault: Signal or power supply bond wire open circuit, high resistance or short circuited to nearby circuits.

Fault effect: Output signals may be missing and input signals ignored.

Failure: Incorrect operation of circuit, possibly stuck-at 0 or 1 on faulty inputs. Power supply faults will cause the complete circuit to fail.

(b) **Defect:** Poor die attach.

Fault: The die may move in the package.

Fault effect: Broken bond wires which can cause open or short circuits as described above.

Failure: As for poor wire bonding.

(c) **Defect:** Package damage, for example cracks.

Fault: Mechanical stress in the die, broken bond wires or pins and also to non-hermeticity.

Fault effect: Mechanical stress in the die can effect electrical performance of the semiconductor and broken bond wires and pins result in the failures previously described. Non-hermeticity is not generally detected until devices have been in operation in the field for some time.

Failure: Normally gross failure of the circuit, possibly after some considerable delay.

The detection of the defects discussed in this section depends to a large extent on the testing procedures and the fault models that are used to generate and assess the tests. Fault models will be discussed in detail in Section 2.3.

2.2.4 The Electrical Effects of Field Failure Defects

The defects to be considered in this section are:

i) Electrical overstress;

ii) Electrostatic discharge damage;
iii) Corrosion
iv) Electromigration
v) Gate oxide shorts
vi) Hot carrier effects

The first two are often characterised as *extrinsic* failures, the latter four as *intrinsic* failures. The defects which give rise to extrinsic failures are external to the IC and so will not be described in the following discussion.

i) Electrical overstress. (EOS) This title encompasses a wide range of conditions, each of which may be transient or continuous. The most common of these will now be described.

(a) **Fault:** Excess supply voltage.

Fault effect: The effects produced depend on the strength and duration of the event. They include latch-up, increased electric fields in devices and increased supply currents.

Failure: Power supply tracks or bond wires may fuse causing total failure of the circuit. Failure due to electromigration is more likely due to increased currents. Threshold voltage shifts, caused by charge injection into the oxide, may cause increases in propagation delays and hence timing failures. However, increased supply voltage will tend to decrease propagation delays, possibly reducing the above effect.

(b) **Fault:** Excess input voltage.

Fault effect: The effect of overstress on signal inputs depends on the *input protection* structures that are present. A simple circuit commonly used for CMOS ICs is shown in figure 2.4.

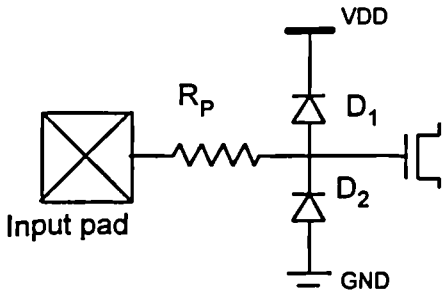


Figure 2.4 Typical input protection circuit for CMOS ICs

Most input protection circuits that are used on CMOS ICs operate by clamping the input voltage (D_1 and D_2) and limiting the input current (R_P). With such circuits present, the effect of excess input voltages either as transient pulses, or static levels

are minimised. The maximum voltage that can appear on the gate of the input transistor is $V_{DD} + V_F$ where V_F is the forward diode drop (typically 0.6V). A 0.6V overstress on the gate of an MOS transistor will not normally cause any undesirable effects. Other protection circuits would have a similar effect on the stress.

If the electrical overstress on a protected input is severe, the protection diodes may suffer damage resulting in a short circuit across the junction.

Failure: Possible stuck-at 0 or 1 faults if the input protection devices are damaged.

(c) **Fault:** Output short to the supply rails.

Fault effect: The output transistors will pass excessive current.

Failure: Fusing of power supply tracks or of the bond wire may occur if the output driver is of sufficiently low output impedance. In such cases, the output will appear to be stuck at the logic value corresponding to the remaining power supply rail.

ii) Electrostatic discharge (ESD) damage. This can be distinguished from electrical overstress by two factors: i) EOS is generally used to describe modest stresses which may be encountered by incorrect connection, power supply noise or other events that generate a few volts of noise. For ESD the over voltage stress is generally over 100V; ii) EOS may be a transient or static phenomenon, ESD is a purely transient effect. ESD occurs when a static charge is suddenly deposited onto the gate of an MOS transistor. The accumulated charge may be generated by the human body or by equipment.

Fault: A very high voltage pulse, in the range of hundreds to tens of thousands of volts, is applied to a signal input. Extremely high electric fields (of up to 10^{12} Vm^{-1}) would be ~~are~~ produced across the gate oxide.

Fault effect: The effect of ESD on a CMOS circuit input depends on the strength of the discharge event. Events which develop a gate voltage of more than 20V are potentially hazardous for a gate oxide thickness of 25nm, as the intrinsic breakdown field strength for SiO_2 ($8x10^8$ Vm^{-1})is exceeded at this voltage. Oxide breakdown by this means results in a conductive path from the gate electrode to the source, drain or substrate of the device.

The effect of a gate oxide short is to increase the input current significantly and also possibly to induce timing or logical errors depending on the severity of the short as described in section 2.2.3.

Failure: Input may be stuck at logic 0 or 1. Increased propagation delays may occur on input stages.

iii) ***Corrosion.*** This may affect individual tracks or large regions of a circuit. The defect may be precipitated by external agents, such as moisture, or may be a purely intrinsic phenomenon. Some examples are given.

Defect: Electrochemical reaction catalysed by impurities such as chlorine in passivation layers resulting in corrosion, or the interaction of gold and aluminium at wire bonds [2.20]. Phosphorus may also lead to corrosion of aluminium, which is unfortunate, as it is present in phosphosilicate glass which is used as a passivation layer.

Fault: Open circuits and high resistance in aluminium tracks and at bond pads. The fault is equally likely to occur in signal and power supply tracks.

Fault effect: Increased signal propagation and open circuit faults.

Failure: Timing faults, signals stuck at logic 0 or 1.

iv) ***Atomic transport phenomena.*** There are a number of mechanisms which result in atomic transport. The most important occur in metal tracks.

Defect: The most commonly discussed transport phenomena is *electromigration.* This is the movement of atoms in conductors resulting from momentum transfer from electron flow [2.21]. A similar defect is stress migration which is the movement of atoms in conductors due to a stress gradient in the track. Contact migration is caused by concentration gradients of different materials at interface regions which produce a diffusion of atoms to reduce such gradients. This effect occurs in aluminium to silicon contacts where aluminium can migrate into silicon and vice versa.

Fault: Voids or hillocks in tracks producing open circuits and possibly short circuits. Contact migration may result in the formation of rectifying contacts [2.22].

Fault effect: Voids in tracks generally act as an open circuit. It has been demonstrated [2.23] that tunnelling conduction can occur across such voids. The current flow in such cases is insufficient to maintain correct operation of the circuit at normal operating frequencies. However, this can affect the operation of "floating" nodes as will be shown later. The effect of the open circuit will depend on its location and the function of the track that is affected.

The effect of an open circuit in a signal track is generally assumed to cause the circuit node being driven to, remain at a logic level (either 0 or 1). This is an unrealistic representation of the effect of the failure and one of the main aims of this thesis is to establish theoretically and experimentally the effect of open circuit defects in signal lines.

Failure: Timing failures, floating nodes and signals stuck at logic 0 or 1.

v) ***Gate oxide shorts.*** A gate oxide short consists of a conductive path from the gate to the drain, source or substrate (channel) of an MOS transistor. The production sources of this defect and its electrical effect were discussed in detail in earlier sections. The defect was also considered in the current section, where it arises as a result of electrical overstress or ESD. A further cause of gate oxide shorts is a gradual wearout mechanism generally referred to as time dependent dielectric breakdown (TDDB). This is the mechanism described by Courtois [2.9]. Anolick and Nelson [2.24] have demonstrated both electric field and temperature dependence of this phenomenon. Further studies [2.25, 2.26] have shown it to be an intrinsic breakdown mechanism resulting from Fowler-Nordheim tunnelling of carriers into the gate oxide. The effect of the fault has been described in section 2.2.3.

vi) ***Hot carrier effects.*** Degradation of MOS transistors due to the generation of hot carriers has only recently become a significant phenomenon. Despite the considerable literature on the subject at the time (for example Hess and Sah, 1974 [2.27], Ning et al., 1977 [2.28] and Hu, 1979 [2.29]) Courtois did not recognise the effect and it has not been identified explicitly in other field failure data studies. There are two probable reasons for this: i) the devices being analysed date from before 1980 and up to 1985. Most CMOS devices at this time were not susceptible to hot carrier effects due to the relatively low electric fields present in the device channel; ii) the effects of hot carriers on device performance may be attributable to other causes making identification difficult. However, as device dimensions have been reduced, hot carrier effects have been recognised as a potentially dominant failure mechanism and have received considerable attention in the literature (for example Eitan and Frohman-Bentchkowsky [2.30], Tam et al [2.31], Hsu and Tam, [2.32]). Despite the lack of field failure data for this mechanism, the effect has been studied both experimentally and theoretically from around 1982 (for example Takeda et al [2.33], Katto et al, [2.34]).

Hot carriers are generated in the inversion and pinch off regions of MOS transistors under the influence of high lateral electric fields. Under such conditions, the energy of electrons and holes is greater than would generally be expected - such carriers are referred to as *hot* carriers. These can degrade devices in several ways.

(a) **Fault:** Impact ionisation in the inversion layer or pinch off region of an MOST can create electron-hole pairs.

Fault effect: Holes are swept into the substrate under the influence of the device electric fields and constitute a substrate current.

Failure: Increased power consumption.

(b) **Fault:** Photons may be generated by electron-hole recombination, or by Brehmstralung from high energy electrons.

Fault effect: Re-absorption of photons in the charge storage region of dynamic RAM circuits can cause loss of charge.

Failure: Transient or intermittent logical failure of such devices.

(c) **Fault:** Hot carriers may enter the gate oxide by tunnelling or surmounting the Si-SiO_2 potential barrier.

Fault effect: Gate oxide breakdown or threshold voltage shifts.

Failure: Failures associated with gate oxide breakdown as previously described. Also parametric failures such as degradation of transistor transconductance resulting in timing failures. The consequences of parametric shifts and other effects resulting from hot electron injection are discussed in the literature, for example Aur et al [2.35], Pimbley [2.36], but precise details will depend on circuit configurations.

2.2.5 Summary of Failure Mechanisms

We have seen that a wide range of defects can arise in CMOS integrated circuits and that the electrical effect of such faults can range from insignificant to total chip failure. Some production defects are latent, resulting in lifetime failures. Others, while being a significant departure from the designed structure, have little, if any, effect on circuit performance.

Field failure defects may cause parametric shifts to occur, for example, threshold voltage shifts due to hot electron injection, resulting in intermittent faults. Alternatively, complete failure of the circuit may occur in a very short time as in the case of electrical overstress or electrostatic discharge damage.

The complex physical nature of IC defects and the often subtle electrical effects that are produced require abstraction to a high degree to enable the efficient design of tests for their detection. This is the purpose of *fault models* which will now be discussed in relation to the defects identified in this section.

2.3 Fault Models

The aim of fault models is to simplify the wide range of effects that arise from the large number of possible defects in integrated circuits. This allows test generation to become a practical task instead of the time consuming procedure that would result from detailed physical analysis of all possible defects in a circuit, and their consequences. As shown in the previous section, the physical mechanisms that produce defects, and their subsequent electrical effect are very wide ranging. To derive a test for a defect based on the physical processes of its evolution would be immensely time consuming because of the large amount of detail involved. It is therefore necessary to abstract the fault to a simple electrical effect that can be detected by simple observation of, for example, the logic levels at the output pins of the circuit under test.

Ideally, a fault model should encompass many physical defects. This allows a test generation methodology based on the model to be applied to many faults. In fact, the dominant techniques for test generation do not consider defects at all. They simply aim to establish that fault effects, represented by models of defects, can be identified if present. This has the advantage that the physical analysis of defects can be ignored thus greatly simplifying the test generation procedure.

The disadvantage of this methodology is that information is lost by the high level of abstraction. Consequently, a test failure does not identify or locate the defect that is present, but simply notes that one (or more) of a wide range of possible defects is present. Generally this information is sufficient however and the detailed failure analysis is left to specialists if required. In production testing it is sufficient to know that a chip has a fault and must be discarded.

Throughout this discussion of faults and fault models, we will only consider the case of a single fault in the circuit. Multiple fault detection and test generation is not considered, as it is generally thought to be an intractable problem. The number of possible multiple *stuck-at* faults (this term is explained in the next section) in an n node circuit is $3^n - 1$. Even for modest circuits this number of faults cannot be considered as the test generation process would simply take too long and be too complex. For example, a circuit containing 40 nodes has more than 10^{19} possible multiple fault conditions. The assumption that only single fault occurs will therefore be used in this discussion and throughout the thesis.

2.3.1 Classical Fault Models

The most commonly used fault models for digital systems are the *stuck-at 0* and *stuck-at 1* models. In these, defects are abstracted to the logic level. All detectable faults are

assumed to cause an output or input to be stuck at the appropriate logic level for all input combinations. The model is used to generate tests in the following way.

> *To test for a circuit node being stuck-at 0, for example, the inputs are set to establish a logic 1 on the node. If, when the test is applied, the node does not adopt its fault free value (logic 0) then a defect is present which is causing the node to be stuck at logic 0.*

Each node in a circuit has two possible fault conditions: stuck-at 0 or stuck-at 1. Both inputs and outputs of logic gates are tested for the fault and so for an m input logic gate there are $2(m + 1)$ faults to test for; for example, a three-input NAND gate has eight possible fault conditions. This is considerably smaller than the total number of possible defects that could arise in the circuit. The reduction is achieved because of the high level of abstraction of the failure modes.

In fact it is unnecessary to generate individual tests for each of the fault conditions possible. The complete set of faults can be *collapsed* to a minimal set which still covers the complete set while requiring fewer tests. Consider a three input NAND gate with inputs *a, b, c* and output *f*. The faults *aS0* (*a* stuck-at 0), *bS0*, *cS0* and *fS1* (*f* stuck-at 1) are said to be *equivalent* because, if any of these faults are present the output will be at logic 0, the fault conditions cannot be distinguished. This will occur because if any input of a NAND gate is at logic 0 the output will be logic 1. A single test can therefore be used to detect all of these faults simultaneously. For the NAND gate the only possible test vector for these faults is (*abcf*) = (1110) (i.e. all inputs set to logic 1 and the fault free output is logic 0).

Of the remaining faults only *aS1, bS1* and *cS1* need to be tested because each of these *dominates* the fault *fS0*. For example, the only test for *aS1* is (*abcf*) = (0111). This test will cause the output to be at logic 1 if the fault *aS1* is not present and so the test will also cover *fS0*. The same argument applies to the *b* and *c* inputs. The faults, however, are not *equivalent* because there are three possible tests for *fS0* (i.e. (*abcf*) = (0111) or (1011) or (1101)) and selecting any one of these to cover *fS0* will only cover one other fault, not all three. A sufficient set of tests for the three-input NAND gate are shown in Table 2.3

Test: *a b c f*	Faults covered
1 1 1 0	*aS0, bS0, cS0, fS1*
0 1 1 1	*aS1, fS0*
1 0 1 1	*bS1, fS0*
1 1 0 1	*cS1, fS0*

Table 2.3 Fault coverage for a two input NAND gate

Based on the stuck-at model, the tests for a NAND gate are simple to generate and apply. The compression of testing in this way can be extrapolated to complete chips and this to a large extent explains the popularity of the stuck-at fault model. We should however investigate how thorough this reduced testing is on a real circuit.

The adequacy of the stuck-at fault model has been investigated by many workers in this field, for example Wadsack [2.37], Al-Arian [2.5], Abraham [2.38], Banerjee and Abraham [2.7], Galiay et al [2.6], Hayes [2.39]. The analysis is generally performed by introducing faults at the electrical level i.e. into a transistor level circuit diagram. Simulation or analysis is then used to show that certain defects do not produce a stuck-at node on input or output. The papers listed above do not take the analysis to the point at which the percentage of faults that are not covered by the tests generated using the stuck-at model, can be determined. They simply identify some defects that are not detected and then concentrate on establishing a fault model for those cases.

To perform the quantitative analysis suggested, two further stages of refinement are required. A distribution of defects must be determined and these must be introduced randomly into the *physical layout* of the circuit. A study of this kind has been reported by Shen et al [2.40] in which a technique called Inductive Fault Analysis is described. The technique is used to investigate the effect of circular defects of varying diameter on the lithography of integrated circuits. These effects are then translated into circuit behaviour. The technique has only been applied to photolithography defects and so the results cannot necessarily be extrapolated to the host of defects which were identified in the previous section. However, the paper presents some interesting results which will now be summarised.

The defect size density distribution (often misleadingly called the *defect density distribution*) that is used for the analysis is described in a paper by Maly, [2.41], and is based on production line data reported by Stapper [2.42]. Clearly some caution is needed here as the precise form of the defect size probability density function is dependent on local conditions in a cleanroom. However, other publications on the subject (for example [2.43]) show an approximate $1 / d^3$ (where d is the defect diameter) form for the distribution function which can be modified to fit most production line data that has been published. (The probability density function actually consists of two regimes: a linearly increasing region from $d = 0$ to d_0 and the $1 / d^3$ region for $d > d_0$. The parameter d_0 varies with production lines but is always smaller that the smallest feature size.)

Shen considered defects to be generated according to the distribution function and scattered randomly over the metal, contact, polysilicon and active layers of a simple nMOS circuit containing 29 transistors. It was assumed that a defect may add material to a layer or remove it. In the simulation reported, only 10% of the 4800 physical defects that were

generated have a significant electrical effect on the circuit performance. Of the 476 electrical faults, only 132 produce a clearly defined stuck-at failure. This is 28% of the total electrical fault effects that occur. The other electrical effects that are produced are characterised as transistor stuck on or off, floating lines, bridging faults and miscellaneous effects.

It is not correct to assume from these results that a test set based on the stuck-at model will only detect 28% of the defects as some of the other defects will also be detected by these tests. This may occur because, in some circumstances, a defect will act as a stuck-at fault. Alternatively, the test pattern generated may, by chance, have the property of detecting timing delays or sequential faults (these are described in the next section). In conclusion Shen et al predict that 64% of the photolithography defects that cause electrical malfunction are likely to be detected by stuck-at tests. This result is closely in agreement with another experiment reported by Ferguson et al. [2.44]. Burgess et al [2.45] report a higher percentage coverage by tests generated using stuck-at fault models, but they show that this cannot be guaranteed and depends on the chance sequencing of test patterns to detect certain faults. Some new work in this area is presented in Chapter 3.

2.3.2 CMOS Stuck-Open Fault Model

The most common failure of the stuck-at model that is reported in the literature is its inability to model the effect of *stuck-open* faults. This effect arises in CMOS circuits because of the complementary nature of the networks in static logic gates. The term "stuck-open" normally refers to an open circuit between the output of a CMOS logic gate and one of the supply rails. Effectively one of the transistors in that path cannot be closed to create the path i.e. it is stuck open. The fault, however, may arise from a wide range of defects including missing metal, contact or active regions or a transistor gate stuck-at 0 (for *n*-channel, stuck-at 1 for *p*-channel). The effect is most easily explained using an example.

Consider the two input NOR gate in figure 2.5 which has a defect at **d** that causes an open circuit in one path between the output and 0V line. For input combinations (*ab*) = (11) and (01) the gate will operate correctly and the output will be at logic 0. The input combination (*ab*) = (00) will also produce the correct output of logic 1. For the input combination (*ab*) = (10) however the output is isolated from both power supplies due to the complementary operation of CMOS circuits and the presence of the fault. The dynamic charge storage capability of CMOS will cause the output to retain the logic level corresponding to the previous input combination. Therefore, the output value in the fault condition will be determined by the previous test vector. Consequently the fault may or may not be detected.

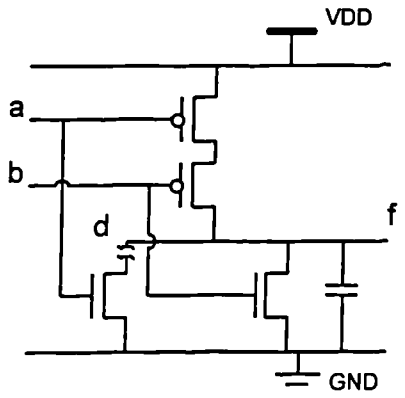


Figure 2.5 Two i/p NOR gate with stuck-open defect

The following example illustrates the problem. To test for the gate output being stuck-at 1 it is necessary to establish a logic 0 at the output which may be achieved using the test (*abf*) = (100) - the output should be discharged through the faulty path. If the previous test in the sequence had coincidentally set the output to logic 0 by applying (*ab*) = (11) or (01) then when the test vector (100) is applied the output will remain at logic 0 (the fault free value) and the fault will not be detected. If, however, the previous test caused the output to be a logic 1 then this could not be discharged to logic 0 through the faulty path and the fault would be detected.

From the above example we can see that to test for stuck-open faults in CMOS circuits it is necessary to establish an initial value at the gate output prior to the test. This will reveal the defect if it is present when the test is applied. For example, to test for a stuck-open fault in the pull-down path of a logic gate the output must be preset to logic 1 in the test cycle preceding that in which the stuck-open test is applied. Therefore a two vector sequence of tests is required to test for stuck-open faults. Furthermore, it is necessary to test each pull-up and pull-down path independently for complete coverage of all stuck-open faults.

This last conclusion illustrates the inadequacy of not only stuck-at test generation but also schemes such as "toggle testing" in which all gate outputs are toggled. For complete coverage, all gate outputs must be toggled by activating all possible pull-up and pull-down paths, a condition not normally imposed on such testing schemes.

Stuck-open faults were first noted by Wadsack [2.37]. In this paper the author notes the inadequacy of the stuck-at fault model and proposes logic gate models of combinational circuits which allow stuck-open faults to be modelled as stuck-at faults. This is achieved by including a bistable element in the circuit to model the charge storage feature of a stuck-open fault. This is a conceptual modification to the circuit which allows fault simulation to be carried out based on stuck-at models. A similar approach is suggested by Al-Arian and Agarwal [2.5] in which some very high level functional models of stuck-open faults in simple gates are reported. The test sequences generated by these techniques are more complex than simple stuck-at tests and result in the two vector sequential tests as previously described.

An ordering scheme for the exhaustive test set of combinational circuits is proposed by Bates and Miller [2.46]. It is shown that Eulerian cycles of test vectors can cover all stuck-open faults in combinational circuits. An improved scheme is presented by Bensouiah, Johnson and Morant [2.47] in which a reduced test sequence is used to test for all stuck-open faults in combinational circuits. The sequence has the advantage of producing a trivial (toggling) output response allowing simple analysis for built-in-self-test circuitry.

Several authors (for example Reddy and Reddy [2.48] and Liu and McCluskey [2.49]) suggest modifications to circuits to improve stuck-open testability. Generally the modifications provide a simple means of pre-setting the output prior to application of the test vector. Layout and circuit modifications are suggested by Soden et al [2.50] and Moritz and Thoren [2.51] to reduce the likelihood of a stuck-open fault occurring. Other authors have addressed the problem of test invalidation by timing skews in circuits [2.48]. This can occur if timing delays in a circuit cause short pulses ("glitches") to appear at a gate input after its output has been preset for a stuck-open test. The pulse may cause the output to be charged or discharged through a fault-free path before or whilst the test vector is being applied. Consequently, the gate may appear to be fault free when a fault is present.

Thorough stuck-open testing not only covers stuck-open faults but also stuck-at faults. This may account for the large amount of literature on the subject which has been generated mostly since 1986. Shen et al [2.46] claim that only 3% of faults generated by inductive fault analysis are stuck-open faults and suggest that alternative defects should receive more attention. However, as a step towards achieving complete coverage of all defects, the published work on stuck-open faults is a significant contribution.

2.3.3 Parametric Fault Models

The stuck-at and stuck-open fault models that have been described are *logical models* representing the effects of defects as incorrect logic values. Many of the defects and subsequent electrical effects identified in this chapter do not result in a logical error. Their effect is more subtle, causing a shift in a circuit parameter such as propagation delay, voltage level or current consumption. Detecting analogue shifts in such parameters is potentially a more complex and time consuming task than logical testing (which is itself a difficult procedure) and the subject has only recently been addressed significantly in the literature.

Parametric faults have a wide range of causes. Two of the most common defects that are discussed with regard to their parametric effects are *stuck-closed* faults and *bridging* faults.

i) **Fault:** Transistor stuck-closed. A transistor is stuck in a permanently conducting state and therefore the device is termed *stuck-closed* or *stuck-on.* The defects leading

to this fault include transistor gates stuck-at 1 or 0 (for *n*-channel and *p*-channel respectively), missing polysilicon gate or drain-source short circuits. Threshold voltage shifts may also lead to this form of fault.

Fault effect: Stuck-closed faults tend to result in intermediate voltage levels on gate outputs due to both networks in the CMOS gate being active simultaneously. The precise value of the output voltage depends on the relative impedances of the pull-up and pull-down networks and may vary with different input combinations. Consider the two-input NOR gate shown in figure 2.6 which has a stuck-closed fault on transistor M2.

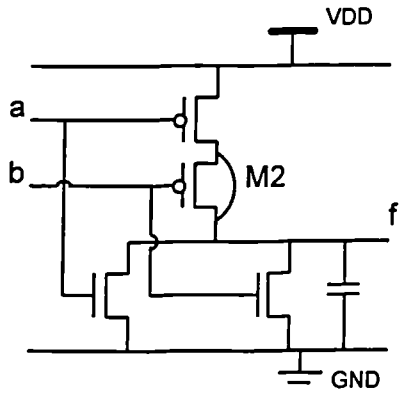


Figure 2.6 Two input NOR gate with *stuck-closed* fault on M2

Assume that the gate is designed for symmetric rise and fall delays and that the *on resistance* of the transistors is R for *n*-channel devices and $R/2$ for *p*-channel. We will also assume that the fault causes transistor M2 to have an resistance of $R/2$ - i.e. it is stuck "fully on". Given these conditions, the output voltage for the four input combinations is as shown in Table 2.4.

a	b	f (Volts)
0	0	V_{DD}
0	1	$^1/_2\ V_{DD}$
1	0	0V
1	1	0V

Table 2.4 Output voltage for all input combinations for a faulty NOR gate

Clearly the output voltage for the input combination (ab) = (01) is not a definite logic 1 or 0. The interpretation of this signal will depend on the transition voltage of the succeeding gates. If we remove the constraint that the fault produces a transistor with a nominal resistance of $R/2$ and allow the resistance to range from zero to open circuit (strictly a stuck-open fault), then the output voltage ranges from $^2/_3 V_{DD}$ to 0V. It is

clear, therefore, that this fault effect cannot be considered to be a logical effect as the output logic value under fault conditions depends on circuit parameters.

The stuck-closed fault has been addressed by several authors. A valuable study has been reported by Banerjee and Abraham [2.7, 2.52] in which they introduce a multi-valued algebra for the modelling of intermediate voltage levels arising from parametric faults. The results are interesting, although the discussion of floating gate faults is inaccurate in spite of the fact that this is one of the most thorough papers on non-logical fault effects that has yet been written.

A qualitative description of the effect of *stuck-on* (stuck-closed) faults is given by Ferguson et al [2.44]. The specific case of transistor stuck-on faults caused by floating gate defects is discussed. However, the conclusions of this discussion are unjustified and incorrect because of the simplistic model used for the floating gate transistor. The operation of the faulty floating gate transistor is one of the main topics of this thesis and will be described in detail in subsequent chapters.

Wadsack [2.37] models the stuck-closed fault as a *don't know* logic value in his logic gate models of CMOS faults. This is an adequate, although possibly pessimistic, approach to stuck-closed modelling. Liu and McCluskey [2.53] describe a circuit modification that can be used to ensure that stuck-closed faults are detected. The technique requires the addition of two transistors to each logic gate and two additional control lines. It is assumed that the stuck-closed transistor has a sufficiently low resistance to allow the output node to charge to a faulty value in a normal test period. This is not necessarily the case and the problem is not addressed in the paper.

One of the most interesting techniques for detecting stuck-closed faults is quiescent supply current monitoring or I_{DDQ} testing. Power supply monitoring as a means of fault detection was first suggested by Nelson and Boggs [2.54]. The technique is particularly suitable for CMOS circuits because of the low quiescent power dissipation of fault-free circuits. I_{DDQ} monitoring has been suggested as a testing technique for many CMOS defects: Levi [2.55], short circuits, transistor leakage (drain-source), open circuits; Hawkins and Soden [2.19], gate oxide shorts; Hawkins and Soden [2.64], open circuits; Rodriques-Montanes et al [2.57], bridging faults; Richardson and Dorey [2.58], reliability indicator. It is apparent from this list that the technique is particularly suitable for detecting defects that have a parametric rather than logical effect.

ii) **Fault:** Increased switching delay. Several papers refer to the additional propagation delay in logic gates that can be attributed to stuck-closed faults. (For example [2.7]). This occurs if the resistance of the faulty transistor is greater than the on resistance of the fault-free device. The effect will not be symmetric with only one

transition edge affected depending on the location of the fault. Timing faults can also be caused by other defects such as high resistance contacts or active regions, or threshold voltage shifts.

Fault effect: The effect of increased propagation delays in logic circuits is variable. Some logic gates can tolerate a large increase in switching delay without affecting overall circuit performance, while a small increase in others can cause a total failure of the system. It is also possible that increased delays could lead to intermittent circuit behaviour as circuit delays are sensitive to input patterns and a delay fault may be masked by the use of an incorrect test vector.

Delay fault simulation has been proposed [2.59] and could provide a useful but time consuming solution to the problem described above. Essentially the technique allows the observability of delay fault effects at primary outputs to be assessed. A distinction is made by Smith [2.60] between global or path delay faults and local or single gate faults. In this paper a procedure is described which can identify the paths that are tested for delay faults by a set of test patterns. Both of these techniques provide some assistance in assessing the coverage of delay faults for a given test set. Test generation for delay faults is not described.

iii) **Fault:** Bridging fault. This term is normally used to describe a short circuit between two or more inputs of a logic gate. This could arise, for example, due to a photolithography error producing a metal or polysilicon short between adjacent input lines. Shen et al [2.40] suggest that this is a most significant fault accounting for 30% of the faults generated in their inductive fault analysis experiments.

Fault effect: The effect of defects modelled by this fault is to cause the affected nodes to adopt intermediate voltages. The voltage on a node due to the presence of a bridging fault is determined by the output impedance (often called the *drive strength*) of the logic gates driving the shorted nodes. The effect is similar to that caused by a stuck-closed fault with the competing pull-up and pull-down networks distributed over more than one gate.

2.4 The Correlation Between Failure Mechanisms and Fault Models

In the preceding sections of this chapter we have investigated the possible defects that can occur in CMOS circuits and analysed their effect on circuit performance. The wide range of effects that are possible make testing for individual defects a very difficult task. Consequently, fault models have been developed to simplify test generation and assessment. The fault models currently used have been described in section 2.3. It is now necessary to

investigate the relationship between failures and fault models to establish which models can be used to generate tests for each fault.

Table 2.5 summarises the defects that have been identified, and suggests appropriate fault models or testing techniques for their detection. The defects that are described do not constitute the complete set of possible defects, particularly in the list of photolithography errors, but have been selected as a representative subset from the analysis of the literature described above.

One of the most significant features highlighted by the analysis is that a large number of defects cannot be detected by tests based on the stuck-at model. This is confirmed by physical defect analysis reported by Shen [2.40] and more recent work reported in the next chapter. A further significant fact is that many faults have a parametric rather than logical effect. Consequently, timing fault models and I_{DDQ} testing should receive more attention.

Several of the defects discussed have an unpredictable effect because of their parametric nature. One of the most unpredictable faults is the unconnected transistor gate terminal or *floating gate*. This fault has been investigated by Renovell et al. [2.61], Maly [2.62] and Henderson et al [2.23] who have presented experimental results on some aspects of floating gate faults. However, most authors make incorrect assumptions concerning the effect of the fault, assuming a fixed potential for the floating gate of a faulty transistor and suggest that the fault can be treated as a stuck-open or stuck-closed fault. There is no justification for this assumption. Capacitive coupling from adjacent tracks and devices will affect the potential of the floating gate. The coupling is to a large extent layout dependent although (as will be shown) the most significant coupling is contributed by capacitances within the faulty device. This fact allows the fault to be characterised more accurately than is suggested in the literature.

Other significant defects, such as bridging faults, which are not detected by stuck-at testing are strongly influenced by circuit topology and yet their effect is more predictable. The majority of bridging defects will produce intermediate output voltages resulting in stuck-closed defects which can be detected by I_{DDQ} and timing tests.

There is little distinction to be made between defects that arise during the lifetime of a component and those generated during production. Whilst the types of defect in each case are somewhat different, their effects are as varied in each case, showing no trends to particular types of electrical effect in either case. Consequently, the same testing techniques should be applied to test circuits at production and during lifetime. Built-in-self-test circuitry therefore should implement all of the production testing techniques, i.e. stuck-at, stuck-open, timing faults and I_{DDQ}.

Defect	Fault	Fault effect	Test or fault model
Crystallographic	Charge traps and dopant inhomogeneities	Excess substrate and junction leakage currents	I_{DDQ}
Photolithography			
a) Narrow active track	High resistance in drain or source of p-channel in TX gate	Weak voltage at TX gate output for short period	Transient I_{DDQ}
b) Active bridge	Low resistance from input to output of inverter	Intermediate voltage at inverter output	I_{DDQ}, timing fault
c) Missing active region	Open circuit in V_{DD} supply.	Q output cannot be charged to V_{DD} through pull-up	Output st-0. May not be detected due to charge sharing
d) Missing metal	Open circuit in V_{DD} supply	As above	As above
e) Missing polysilicon	Break in gate track of n-channel device in inverter	Unknown gate voltage	Possibly stuck-closed, stuck-open, I_{DDQ}
f) Missing metal	Break in gate track of p-channel in TX gate	Unknown gate voltage	Not detectable
g) Polysilicon bridge	CLK and CLKB shorted	Clock potential stuck at $V_{DD}/2$	St-1 on CLK and CLKB

Table 2.5a Relationship between faults and fault models

Defect	Fault	Fault Effect	Test or fault model
h) Missing polysilicon	Shortened gate track	Source to drain short on *n*-channel of inverter	Stuck-closed
i) Missing contact	No connection to drain of *n*-channel in inverter	Cannot discharge Q output to 0V	Q output st-1. may not be detected due to charge sharing
Surface contamination	Missing features and breaks in tracks	Similar to particulate contamination	Similar to particulate contamination
Oxide defects			
a) Pin hole in gate oxide	Short between gate and channel / drain / substrate	Feedback to gate from channel, drain or source	I_{DDQ}, timing fault, stuck-closed
b) Ionic contamination and gate oxide thickness variation	V_T shift	Change in gate propagation delays	Timing fault, I_{DDQ}
Packaging defects			
a) Poor wire bonding	Signal or supply pins open circuit	Missing signal or no power supply	Simple functional test
b) Poor die attach	Broken bond wires	As above	As above
c) Package damage	Broken bond wires, non-hermecticity	As above, corrosion	As above, Stuck faults, I_{DDQ}

Table 2.5b Relationship between faults and fault models

Defect	Fault	Fault effect	Test or fault model
Field failures			
Electrical overstress	Over-voltage on supply or inputs, outputs shorted	Possible latch-up, fusing of bond wires or tracks	I_{DD} (for latch-up), stuck-at 0 / 1, stuck-open
Electrostatic discharge damage	Gate oxide breakdown, thermal breakdown of silicon / metal	Feedback to gate from channel, drain or source, open circuits in tracks	I_{DDQ}, timing faults, stuck-closed
Electromigration, stress migration, contact migration	Open circuits in tracks and contacts	Depends on location	Stuck-at 0 /1, stuck-open, stuck-closed
Gate oxide breakdown	Short between gate and channel / drain / substrate	Feedback to gate from channel, drain or source	I_{DDQ}, timing fault, stuck-closed
Hot carriers	Increased substrate current, charge accumulation in gate oxide, gate current	Latch-up V_T shift - change of propagation delays Charge storage leakage	I_{DD} Timing faults, I_{DDQ} I_{DDQ}, stuck tests

Table 2.5c Relationship between faults and fault models

2.5 Chapter Summary

A thorough survey of the literature on physical defects, faults and fault models has been described. A subset of the dominant failure mechanisms has been identified and the adequacy of current testing techniques in detecting these defects has been analysed. The main deficiencies in testing remain in detecting parametric faults for which the most useful tests are I_{DDQ} monitoring and delay fault testing.

A particular fault that deserves further investigation is the floating gate fault. This defect is poorly analysed by most workers who make grossly simplified assumptions of its effect. The significance of floating gate faults, in terms of its possible rate of occurrence both in photolithography errors and wearout mechanisms, and the considerable lack of previous investigation suggests that a thorough investigation of the operation of the faulty floating gate MOS transistor would be a worthwhile task.

In the following chapters, the floating gate fault will be analysed experimentally and a model for the operation of an MOS transistor with floating gate will be developed. The aim is to produce a justified and definitive description and model of the fault to facilitate reliable test generation for this significant and often ignored fault.

Chapter 3
The Floating Gate Fault

The review of failure mechanisms and fault models presented in chapter 2 identified a range of defects that are not adequately modelled by the classical stuck-at fault model or the more refined CMOS stuck-open model. One of the most significant defects in this category is the floating gate fault in which the MOS transistor gate is left with no connection to the rest of the circuit. The effects of this fault are difficult to predict, depending on a range of circuit and device parameters. It is however clear that this defect can rarely be modelled as a stuck-at or stuck-on fault. This is mainly because the floating gate voltage is not stuck at a particular value but can change under the influence of electric fields in the MOS transistor and the surrounding circuitry.

In this chapter we will investigate the floating gate defect in detail. The defects which give rise to the fault will be identifed and the significance of the fault, in comparison with other failures, will be established. New analysis of the effects of particulate contamination is described and the relative numbers of the main categories of fault are determined. From this work the floating gate fault is shown to be a significant form of fault. The effect of the fault will then be investigated and some test circuits, which have been designed to reveal the effect of this fault, will be described.

3.1 Causes of the Floating Gate Fault

A floating gate will occur when the connection between the gate terminal of an MOS transistor and the driving source becomes open circuit. This can arise in several ways.

i) *Layout design error.* These should be detected at design time by design rule checking (DRC), electrical rule checking or netlist extraction and comparison. In the context of open circuit faults, design rule checks will only detect small gaps in tracks which violate a minimum separation rule. A missing section of track or a missing contact will not be detected by DRC. Electrical rule checks should detect a floating

gate as a transistor gate with no path to VDD or GND (the supply rails). There are no exceptions to this. If the rule is checked then the error will be detected. Network consistency checks should detect an unconnected gate. Again, there are no exceptions to this and all errors should be identified. However, incomplete checking may result in floating gate faults being missed.

ii) *Photolithography defects.* These will occur when particulate contamination affects any of the layers of a circuit. Defects in any of the interconnect or contact layers, which result in missing material, can result in floating gate faults. The layer must be generated using negative photoresist (i.e. exposed regions are left after development) for particulate contamination to result in missing material. Layers generated using positive resist will be susceptible to bridging defects caused by particulate contamination.

Particles may contaminate the layer surface prior to photoresist deposition causing subsequent lift off of the resist and hence missing material. Alternatively particles may contaminate the surface of the photoresist after spinning, as shown in figure 3.1, or the surface of the reticle for the layer. Contamination in all of these locations has a similar effect to that illustrated and could cause missing material in any of the layers listed above, resulting in an open circuit or floating gate defect.

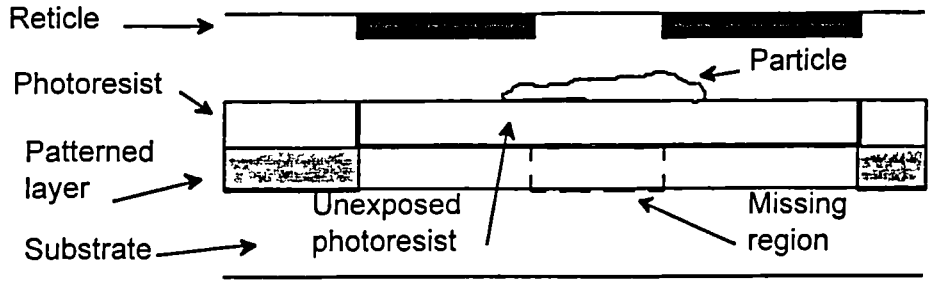


Figure 3.1 Generation of open circuit defects in the photolithographic process due to particulate contamination on the surface of the photoresist. (Note that the vertical scale is compressed for convenience)

If the defect is too small to cause a complete open circuit then a narrow track which may have high resistance will be created. This is likely to present a reliability hazard to the circuit causing an open circuit and hence a floating gate to occur during the lifetime of the component.

The location of a defect in the circuit layout will have a significant effect on the fault produced. Figure 3.2 shows a set of possible fault locations giving rise to a range of different floating gate faults which are listed below.

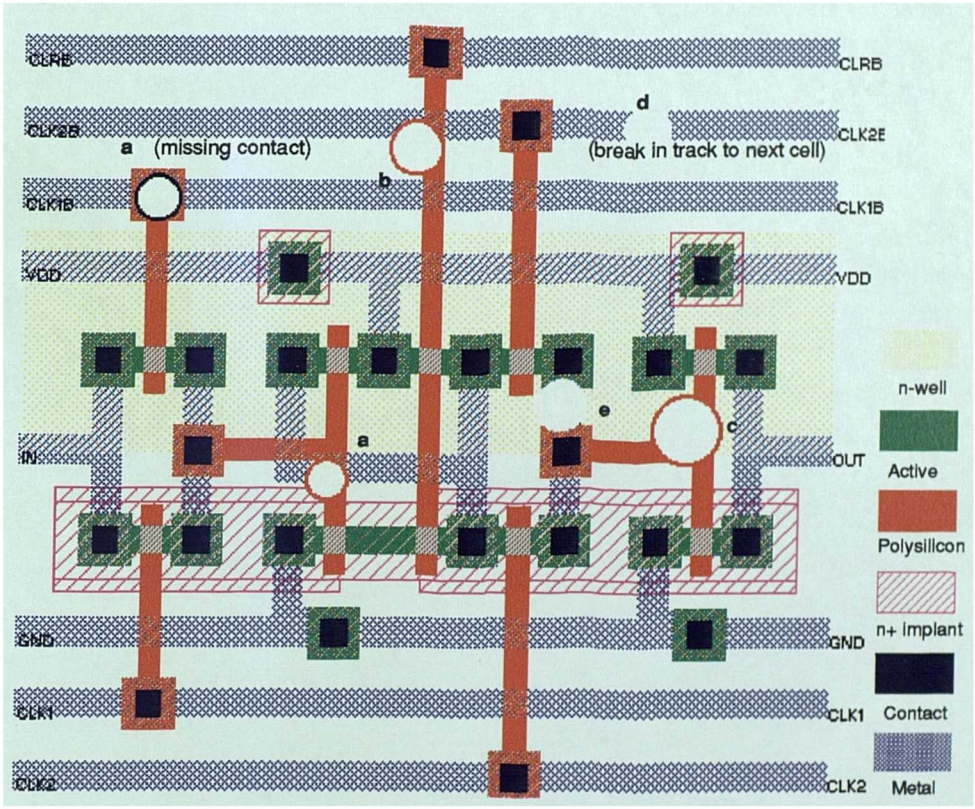


Figure 3.2 Possible defect locations giving rise to floating gate faults

a) Single floating gate (*n*- or *p*-channel)
b) Double connected floating gates (*n*- and *p*-channel)
c) Double independent floating gates
d) Set of double connected floating gates
e) Half floating gate - equivalent to the *stuck-open* fault

The effect of the faults produced are discussed in section 3.3.

iii) *Electromigration.* This effect is normally associated with power supply tracks that have a relatively high unidirectional current flow. It was suggested in chapter 2 that electromigration may occur in signal lines under certain circumstances. In particular, asymmetric rise and fall delays in a logic gate will result in unequal current densities for charging and discharging the load capacitance of succeeding gates. We will now investigate this proposal.

It is generally accepted (for example Bauer [3.1]) that ionic transport due to the electron wind in a conductor, J_{e^-}, is proportional to the current density to the power n as shown in equation 3.1.

$$J_{ion^+} \propto J_{e^-}^n \tag{3.1}$$

where the power n is generally accepted to be approximately 2 (Black [3.2]). The total ionic flux conveyed in a time t is therefore given by:

$$F_{ion^+} \propto J_{e^-}^2 t \tag{3.2}$$

It is important to note that it is not simply the net momentum transfer from electrons that determines the ionic flux. If this were the case the flux would be directly proportional to the current density.

The non-linear nature of equation 3.2 will result in a difference in the ion flow in the two directions in a conductor in which positive and negative currents are not equal. Therefore for the case of asymmetric charging and discharging of a load capacitance there will be a preferential ionic transport in one direction along the signal track resulting in electromigration. Consider, for example, a logic gate in which the output rise time is t_r and the output fall time is t_f. If the peak current flowing to the load capacitance is I_c then the net ion flux towards the driving gate will be as given in equation 3.3.

$$F_{ion_c^+} \propto I_c^2 t_c \tag{3.3}$$

where t_c is the period for which the peak current can be considered to be flowing. This is given by $I_c.t_c = I_d.t_d$, such that the total charge for charging and discharging the load capacitance is the same. The equation for the ionic flux for the discharging cycle is the same with subscript d instead of c. By substituting the charge equality equation (above) into equation 3.3 we obtain:

$$F_{ion_c^+} \propto \frac{I_d^2 t_d^2}{t_c} \tag{3.4}$$

which is not equal to the discharging flux $F_{ion_d^+}$. We therefore see that there is a net flow of ions towards the load capacitance as if a steady current were flowing.

While electromigration would appear to be feasible in CMOS signal lines, it is possible that *electromigration in supply lines will* occur *before signal lines fail* due to the higher unidirectional current flow in such tracks. Large capacitive loads which require large charging currents and signal asymmetry are necessary for electromigration in signal lines, but the failure is a significant possibility.

Most signal lines in CMOS circuits have two sections of track in which unidirectional current flow occurs. These sections lead from the transistor drains to the vertex with the main signal track. If either of these sections fails then a *half floating gate* will occur. Under this fault condition, the two transistor gates driven by the affected gate output are floating for one logic condition (logic 0 if the failure is on the *n*-channel branch, logic 1 for the *p*-channel branch). This is precisely the situation which arises when a stuck-open fault occurs. It is sensible to characterise this fault as a half floating gate as the gates are indeed floating for typically 50% of the time. Consequently, when in this state, the fault will act as a fully charged or discharged floating gate. The effect of this on circuit operation will be described in section 3.4.

iv) *Contact migration.* This may take several forms, three of which can result in the creation of voids and hence open circuits at a contact. The step coverage regions (i.e. the edges of a contact cut) are common sites for simple aluminium electromigration. The step coverage often produces a thinning of the metal layer making it more susceptible to voiding. Aluminium may also migrate into the silicon under the influence of the electron wind, again generating voids. Furthermore, silicon may migrate into the metal leaving voids beneath the contact. Each of these mechanisms may lead to a floating or half floating gate. The latter two mechanisms may lead to the formation of rectifying contacts, for example aluminium migrating into *n*-type silicon will create a *p-n* junction diode, or even short circuits through doped regions to the substrate. These mechanisms do not lead to floating gate faults.

The arguments presented concerning pulsed bidirectional current flow apply to contact migration.

v) *Stress migration.* This defect can arise in any metal track independently of current flow. It arises through the release of tensile stress induced by differential thermal expansion in metals and other layers in the IC. Signal lines are as susceptible to this failure mechanism as any other metal tracks and open circuits, and hence floating gates, may result.

vi) *Corrosion.* Electrochemical reactions may cause corrosion of metal tracks resulting in open circuit defects. This form of defect is often quite global in nature and would generally result in multiple faults in a circuit including floating gates.

From the range of defects presented that can cause floating gates to occur, it is apparent that this type of fault is potentially important. We will now investigate the significance of the fault further and attempt to identify its probability of occurrence relative to other failure modes.

3.2 The Significance of the Floating Gate Fault

Floating gate faults can arise from a wide range of defects as discussed in the previous section. However, it is necessary to establish that they also occur in significant numbers when compared to other forms of fault such as nodes stuck-at 0 or 1 or transistors stuck open. In this section we will attempt to quantify the significance of this fault.

3.2.1 Layout Design Errors

It is unlikely that layout design errors will cause any form of fault on an integrated circuit since very thorough checking is performed before mask generation occurs, using the tools outlined in the previous section. In addition, the layout for most IC's is produced automatically providing less opportunity for the introduction of errors than with full-custom layout. For these reasons, while is it possible that layout errors can result in faults in a circuit, it is unlikely that this is a major cause of failures.

3.2.2 Photolithography Defects

Photolithography defects are generally considered to be the most significant cause of IC faults. Some work has been published on the relative significance of different faults that can occur in nMOS circuit as discussed in chapter 2 [2.46]. A procedure called Inductive Fault Analysis is used to investigate the susceptibility of the layout of a logic cell to photolithography defects. In this process, random defects are generated on the cell layout and their effect is automatically assessed. In the experiment, 4800 defects are generated on an

nMOS logic cell. Of these, only 476 produce a fault effect that would be detectable by normal test techniques. A summary of the distribution of faults generated by the procedure is reproduced in table 3.1.

Fault type	Number of occurrences
Single stuck-at 0 or 1	108
Multiple stuck-at 0 or 1	24
Transistor stuck closed	49
Transistor stuck-open	11
Multiple stuck-open or closed	10
Floating gates	101
Bridging faults	144
Supply tracks	16
Other faults	13

Table 3.1 Fault distributions for photolithography defects (from Shen et al [2.46] Note: Fault nomenclature has been changed for consistency with this thesis)

The results show that only 28% of faults directly produce stuck-at 0 or 1 behaviour. Stuck-open defects represent a very small percentage while bridging faults and stuck-closed constitute around 40% of the faults. Floating gates represent 21% of the faults generated.

These results show the inadequacy of using the stuck-at fault model for test generation in MOS circuits. While many of the faults which are not classified as stuck-at faults will be detected by stuck-at tests, the high percentage of these faults indicates that tests generated using the stuck-at fault model will provide inadequate fault coverage for this circuit. Over 60% of the faults (bridge, stuck-closed and floating gate) may not produce a logical fault effect and so are unlikely to be detected by stuck-at tests.

The results presented by Shen are most useful but relate to a single nMOS logic cell. As is suggested in the paper, the relative frequencies of faults produced are likely to depend, to some extent, on the circuit topology. It is not clear therefore, as is suggested by Shen, that the results can be applied to CMOS technology in which the topology is inherently different and additional layers are used. In a later paper [3.3], Ferguson and Shen describe a more extensive experiment based on three CMOS circuits. The categories of fault that are used in this paper are reduced and so the detailed information that can be extracted is limited.

However, it is apparent that bridging and stuck-closed faults and breaks in tracks are the dominant forms of failure for these circuits.

3.2.3 CMOS photolithography fault distribution experiments

It was decided that fault distribution experiments should be performed for some typical CMOS cells to compare with the results available in the literature. This would enable the details omitted in the literature to be determined. The aim of the experiments was to establish the relative significance of various fault types in typical CMOS cells.

To achieve this, a computer program was written to generate defects randomly over cell layouts displayed on the computer. The cell layout with superimposed defects was then inspected visually to determine the results of the faults. While automation of the inspection process is possible, it was decided that visual inspection would be adequate for the experiments envisaged and would in fact be superior, in some cases, as it would allow unforeseen fault effects to be detected and interpreted.

3.2.3.1 Generation of Defects

Photolithography defects may occur in any layer of an integrated circuit. For a typical CMOS process, extra or missing material may occur in *n*- (or *p*-) well, active, polysilicon, p^+ implant, n^+ implant, contact, metal 1, via or metal 2 layers. The effect of the defects in each of these layers needs to be assessed using defects of two types, extra and missing material, generated for all layers. It was assumed that extra or missing material was equally likely to occur for each layer. This assumption is not strictly correct as the distribution between the two types of defect will depend on the type of photoresist (positive or negative) used in creating a layer. However, as this factor varies between processes, it was decided that an even distribution between the two types of defect would produce the most general results.

Defects are assumed to be circular. This is an approximation, as many defects have quite random shapes. However, there is no information available on shape distributions and the wide range of sources of particles probably makes it rather meaningless to try to characterise the shapes. (The author recalls a lengthy discussion between C.H. Stapper and W. Maly on this subject in which it was concluded that circular defects were the only sensible choice [3.4]).

Defects are assumed to be scattered uniformly over the cell area. Clustering is not taken into account, as the cells being considered are smaller than the area occupied by a typical defect cluster on a wafer.

The radii of the defects generated are allowed to vary in the range $r_0 < r < r_{max}$. The maximum diameter is essentially infinite. The distribution (or probability density function)

over this range was chosen to vary as $1/r^3$. Defects, therefore, have a maximum probability of having a radius $r = r_0$ and the probability of the defect radii being greater than r_0 decreases as $1/r^3$ to 0 at $r = \infty$. Defects of less than r_0 were not generated. The distribution is illustrated in figure 3.3. This defect radii distribution is suggested by Stapper [3.5] (and others, for example Gandemer et al [3.6]) and is supported by production line data.

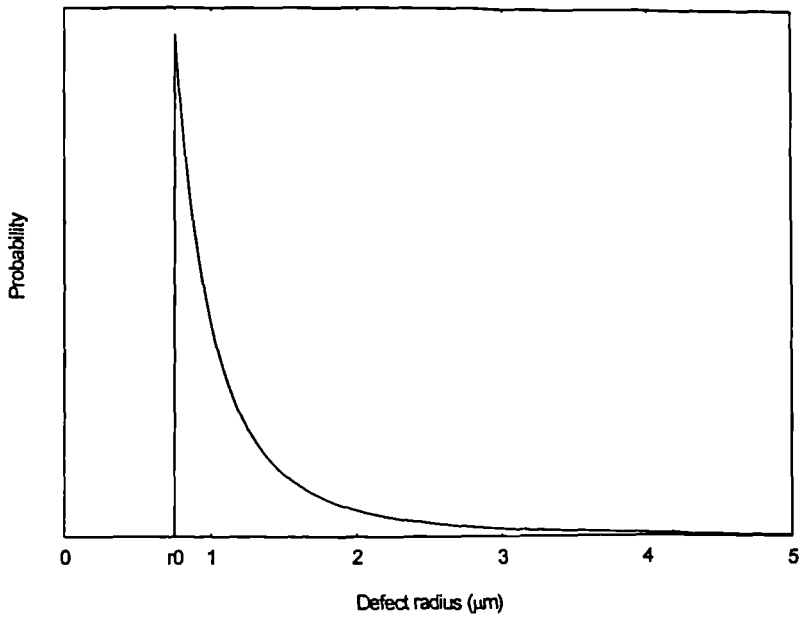


Figure 3.3 Defect radii distribution function used for the defect simulations

It was decided that r_0 should be set to half the minimum feature size for the technology. Defects generated with a diameter significantly less than the minimum feature size could not produce logical fault effects, as an open or short circuit is required for this to occur. Timing errors may occur if a significant fraction of the track is missing and these faults are taken into account in the analysis described. By choosing r_0 to be less than half of the minimum feature size, a large number of defects would be generated that had no effect on circuit performance. This would reduce the confidence level for the results generated (or increase the length of the experiments) justifying the choice of r_0 as half the minimum feature size for the technology being considered.

There is, however, a possibility that by changing r_0 the fault type distribution may be altered. This may occur because of the different minimum dimensions of features on different layers. For example, by reducing r_0 a larger number of small defects will be generated. Therefore, defects in layers which have larger minimum feature sizes such as metal or implant layers, will have relatively fewer defects. This shift in the defect distribution between the IC layers may change the fault distribution which is the subject of this investigation.

An experiment was therefore performed to investigate this potential effect. Three sets of defects were scattered for a cell using values for r_0 of 0.5, 1.0 and 1.5 μm. The defects were generated for metal and polysilicon layers in which the minimum dimensions are 2.4 and 1.5 μm respectively. A total of 4200 defects were generated for the experiment. Analysis of the data shows, as expected, that the relative number of faults generated for the polysilicon layer with respect to the metal layer increases as the minimum defect radius is reduced. However, as can be seen from the fault type distributions presented in figure 3.4, there is no significant change in the shape of the fault distribution of the different values of r_0. We therefore conclude that, over a limited range, the parameter r_0 does not have a significant effect on the fault type distributions for the cells. As we shall see later, far more significant changes are caused by other factors.

To summarise, the defects are uniformly scattered over the layout with a radius variation as described above. They may represent extra or missing material in any layer of the integrated circuit.

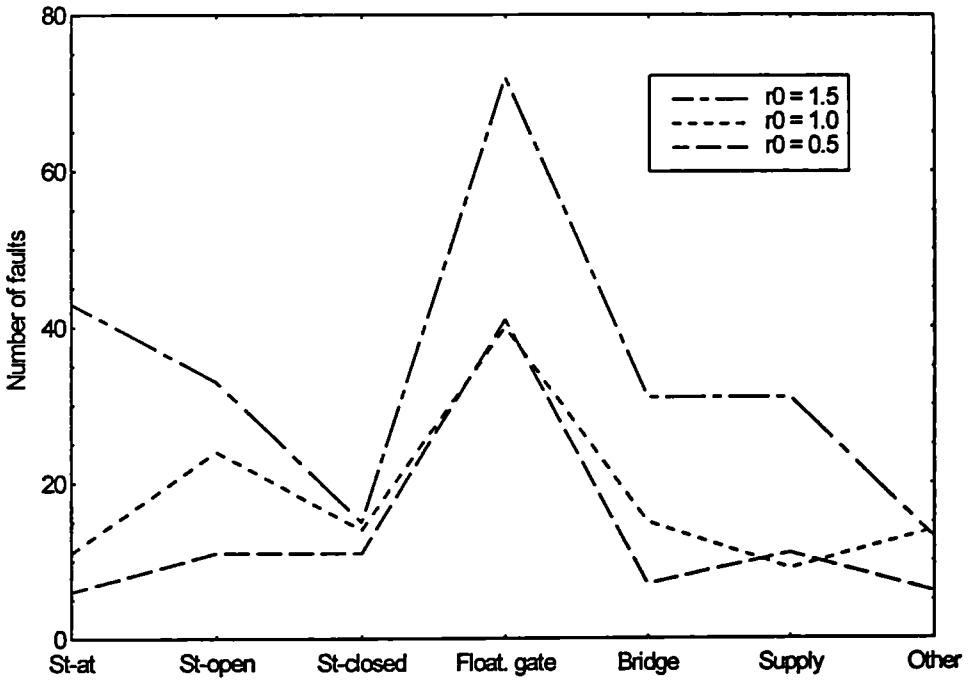


Figure 3.4 Variation of fault type distribution for a range of values of r_0

3.2.3.2 Assessment of Defects

It is not sufficient to only consider the interaction of a defect in a layer with other features in that layer. For example, if extra polysilicon is deposited over an active region a new transistor with floating gate may be formed. Defects on each layer must therefore be

considered in the context of other layers present as well as their location in the circuit. The interactions between defects and circuit layers are summarised in Table 3.3.

Defect layer	Interaction layer	Fault
n-well	p^+ active	Missing well intersecting p^+ active track or *p*-channel region will cause stuck-at 0 on track or channel
	n^+ active	Extra well across *n*-channel will cause a transistor stuck-closed fault. Extra well can also cause a bridge between adjacent n^+ active tracks.
	well tap	Missing well over entire n^+ active in well tap will isolate the tap (*p*-n junction created with substrate)
	Substrate tap	Extra well over entire p^+ active in substrate will isolate the tap (*p-n* junction created with substrate)
p-well	As above	Faults similar to above except stuck-at 1 faults and *p*-channel stuck-closed
Active	Active	Open circuit, short circuit (bridge faults) and high resistance tracks can be generated
	Polysilicon / channel region	Missing active across entire width of device will cause stuck-open - partially missing channel can cause narrow device
	Contact and metal	Missing active covering overlap with metal - contact intersection causes open circuit - normally stuck-open transistor - also missing well / sub. tap
Polysilicon	Active	Extra polysilicon crossing active track creates a floating gate transistor - narrow track created if polysilicon does not completely bisect the track. Missing polysilicon in the channel region can cause stuck-closed transistor

Table 3.3 (a) Layer interaction in the generation of faults by defects on a CMOS IC

Defect Layer	Interaction layer	Fault
~~n-well~~ Polysilicon	Active - metal contact	Extra polysilicon covering contact region causes open circuit between active and metal - normally stuck-open. Also bridge faults if contact to metal not completely isolated (gate - drain / source short possible)
	Polysilicon	Open circuit, short circuit (bridge faults) and high resistance tracks can be created
p^+ implant	p^+ active in n-well (includes contact region)	Missing across track causes open circuit in track - stuck-open in p-channel device
	n^+ active in substrate	Extra across track causes open circuit - stuck-open in n-channel device
	n^+ active - metal contact in substrate	Extra covering contact region causes metal track stuck-at 0
	n^+ active - metal contact in n-well	Extra covering contact region causes open circuit in well tap.
n^+ implant	n^+ active in substrate	Missing across track causes open circuit - n-channel stuck-open
	n^+ active - metal contact in substrate	Missing over contact region causes metal track stuck-at 0
	n^+ active - metal contact in n-well	Missing over contact region causes open circuit - missing well tap
	p^+ active in n-well	Extra across track causes open circuit - stuck-open in p-channel device
	p^+ active - metal contact in n-well	Extra over contact region causes stuck-at 1 on metal track.

Table 3.3 (b) Layer interaction in the generation of faults by defects on a CMOS IC

Defect	Interaction layer	Fault
Contact	Active - metal contact	Missing over contact causes open circuit - normally stuck-open *n*- or *p*-channel or supply open circuit
	Polysilicon - metal contact	Missing over contact causes open circuit - normally floating gate
	Metal over active or polysilicon	Extra contact over intersection will cause short circuit - bridge fault or stuck-at 0 or 1
Metal 1	Metal 1, contacts to active, polysilicon or metal 2	Missing across track or over contact / via causes open circuit - stuck-open, floating gate, supply faults
	Metal 1	Extra between tracks causes short circuit - bridge faults
Via	Metal 1 - metal 2 via	Missing over via region causes open circuit - supply open, floating gate faults (stuck-open unlikely)
	Metal 1 and metal 2	Extra over intersection causes bridge fault, stuck-at 0 or 1
Metal 2	Metal 1 and via	Missing causes open circuit - supply open or floating gate faults
	Metal 2	Extra between tracks causes short circuit - bridge faults or stuck-at 0 or 1

Table 3.3 (c) Layer interaction in the generation of faults by defects on a CMOS IC

In general if a track width or separation is reduced to less than 0.5 μm the defect is considered to cause a fault. For contacts, if the remaining area of the contact is less than 50% of the minimum contact area, or a new contact is at least this size then the defect is considered to cause a fault.

From Table 3.3 the expected set of faults can be established. The set of faults identified as the most significant are:

Node stuck-at 0	*p-channel stuck-closed*	*Independent floating gates*
Node stuck-at 1	*n-channel floating gate*	*Bridge fault*
n-channel stuck-open	*p-channel floating gate*	*VDD open circuit*
p-channel stuck-open	*Connected floating gates*	*GND open circuit*
n-channel stuck-closed	*Multiple connected floating gates*	*Others*

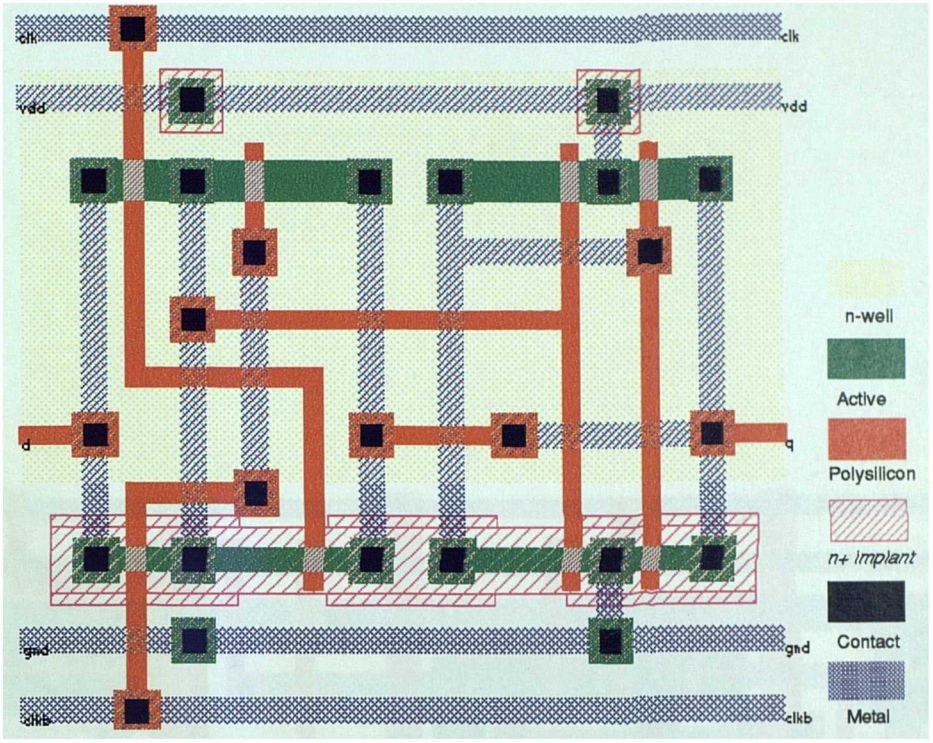
The final category of faults (Others) is used to sum all other fault conditions and serious potential or non-logical faults. These include high resistance tracks, narrow / short channels, very wide / long channels, missing transistor, missing substrate or well taps, extra devices and VDD / GND short circuit. Some of these defects would cause degradation of the timing performance of the circuit (e.g. narrow / long channels) or increased susceptibility to latch-up. Others may lead to a latent defect which would result in reduced reliability of circuits. Faults in this category include short channels, leading to hot electron degradation or punch-through and narrow tracks resulting in electromigration failures.

3.2.3.3 Defect Simulations

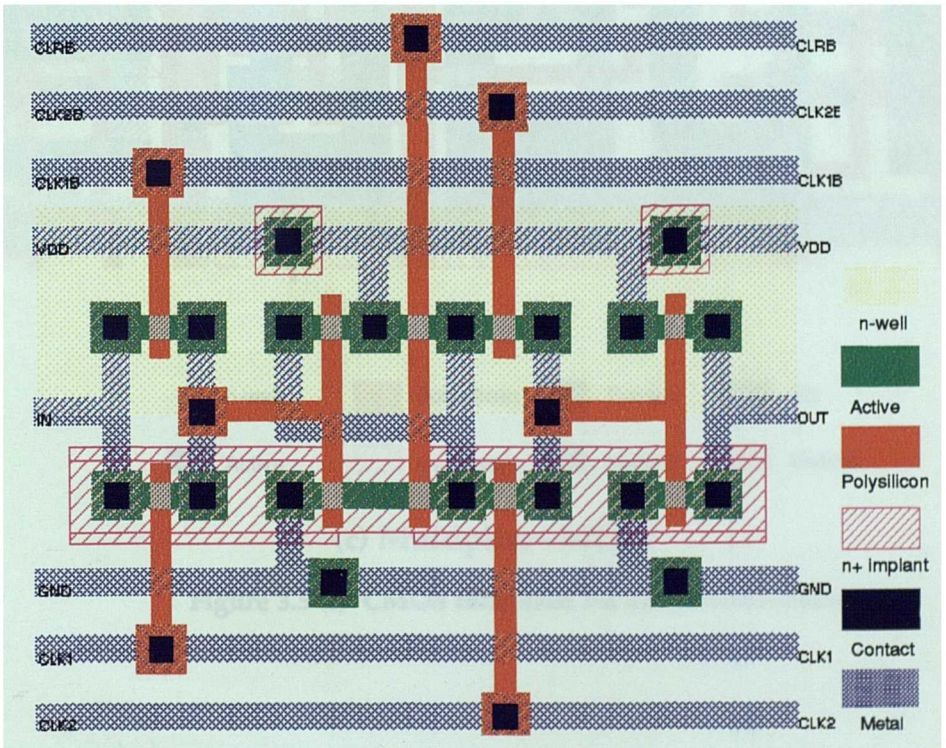
Defect scattering experiments were performed on the three CMOS cells shown in figure 3.5.

Two of the cells were from a 1.5 μm *n*-well CMOS process (DLATCH and SREG). These cells were generated automatically from sticks diagrams. DLATCH is a fairly sparse cell containing four *n*- and four *p*-channel transistors in a total cell area of 4510 μm^2 (69.2 x 65.2 μm). The cell is a static transmission gate transparent D-type circuit. It is representative of a class of cells that do not require extreme compaction. The cell includes metal links between transistor gates. The SREG cell is more compact and contains five metal tracks, which span the entire width of the cell, used for intercell connections. The circuit contains five *n*- and five *p*-channel devices in a total cell area of 3865 μm^2 (64 x 60.4 μm) and is a dynamic shift register element.

The third circuit simulated was a full-custom designed parallel array multiplier cell. The cell is implemented in domino logic and the technology is 3 μm *p*-well double metal CMOS. The cell is very compact and includes wide transistors and a significant use of active tracks for interconnect. There are 25 *n*- and 6 *p*-channel transistors in a cell area of 26320 μm^2 (235 x 112 μm). The cell includes metal and polysilicon intercell signal tracks and the supply tracks are 10 μm wide. This cell also includes some metal 2 tracks used for intercell signal connections.

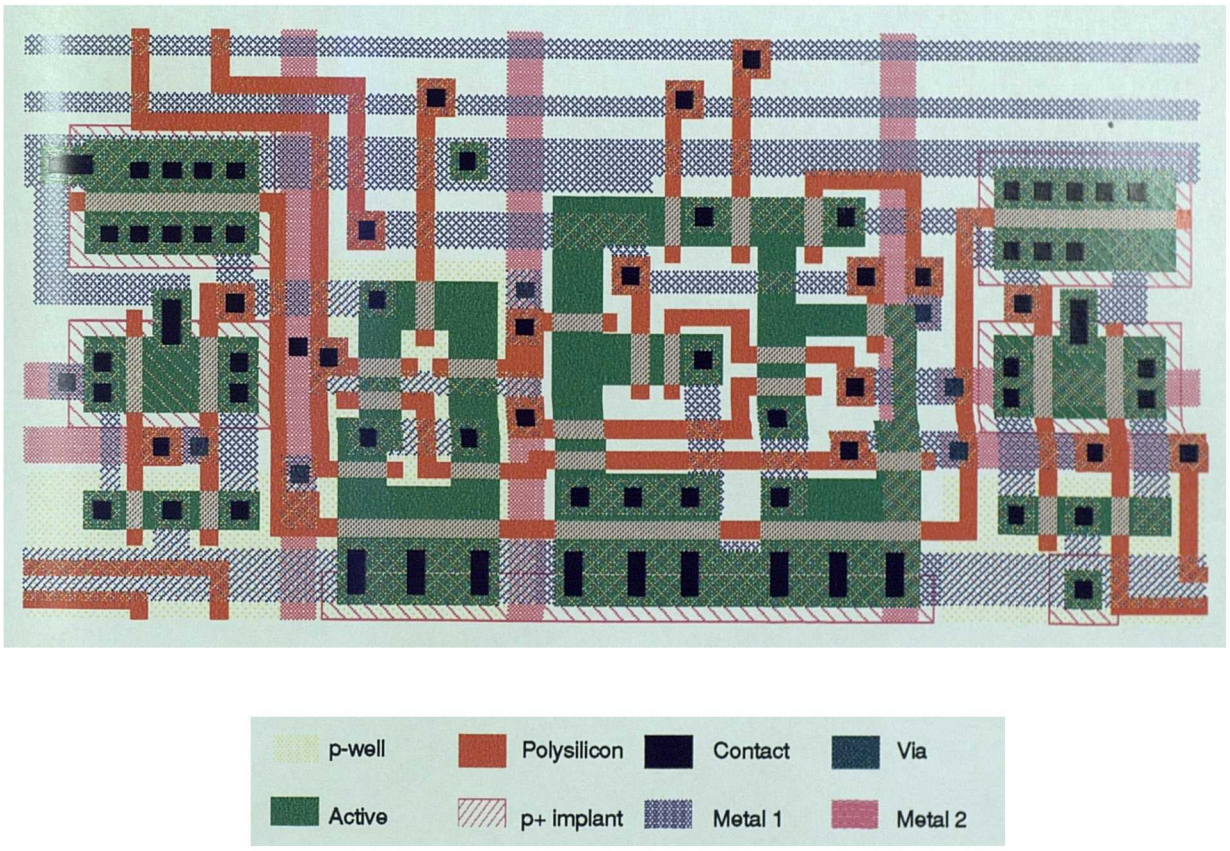


(a) D - latch - DLATCH



(b) Shift register - SREG

Figure 3.5 (a-b) CMOS cells used for defect simulations



(c) Multiplier - MULT

Figure 3.5 (c) CMOS cells used for defect simulations

The cells were selected to represent typical integrated circuit cells that have somewhat different characteristics. One of the limitations of Shen's results is that they apply only to a single cell example and, as will be shown, the fault distribution produced depends to some extent on the topology of the cell used for the simulations. For the 1.5 μm technology a value of 1.0 μm was used for r_0. This is larger than half the minimum dimension, but for most layers the minimum dimension was greater than 2 μm and so this seemed reasonable. For the 3 μm cell r_0 was set to 1.5 μm. It has been demonstrated that the value of r_0 does not significantly alter the distribution of the faults produced by the defects.

Extra and missing material defects were generated for each layer in the cell. A total of 4900 defects were generated for cells DLATCH and SREG and 6300 for cell MULT (700 defects for each layer). This large number of defects (16100 in total) ensures a high level of confidence in the distributions calculated from the results. The results of the simulations are presented in Tables 3.4 and 3.5 and figures 3.6 and 3.7

	S@0	S@1	n st-op	p st-op	n st-cl	p st-cl	n fl. gate	p fl. gate	Con. fl.g	* Con fl.g	Ind. fl.g	Bridge	Vdd open	Gnd open	Others	Total
cdlatch																
n-well	13				10							1			4	28
Active		1	11	9	1	1						4		1	9	37
Poly	2		2		6	8	15	6	7			9		0	9	64
p+	10	6	7	7											2	32
n+	4	2	13	2											9	30
Contact	1		1	2								12	1	7		24
Metal 1	6	3	9	13				3	1	8		6	6	3	5	63
sreg																
n-well	20				15										6	41
Active	1	1	12	15		2						1	1		5	38
Poly			5	2	1	7	9	10	6		1	14			2	57
p+	14	8	5	4												31
n+	11	10	2	6												29
Contact	1			7				1	1			4			5	19
Metal 1	8	8	3	6				2	1	43		15	10	13	3	112
mult																
p-well		63				13									7	83
Active			7									2			19	28
Poly	1	2	3		19	7	14	8	2	12		18			9	95
p+	10		2													12
n+	5	9	1													15
Contact		1	2				1			2		10				16
Metal 1	3	17	2				4		1	9		13			2	51
Via	5	2					1					4				12
Metal 2							5			8						13

Table 3.4 Number of defects in each category for each layer in all cells

Cell	S@	St-open	St-closed	Floating gate	Bridge	Supply	Other	Total
cdlatch	48	94	26	40	32	18	38	296
sreg	82	91	25	74	34	24	21	351
mult	118	17	39	67	47	0	37	325
All cells	248	160	90	181	113	42	96	930

Table 3.5 Summary of number of defects in all layers for each cell

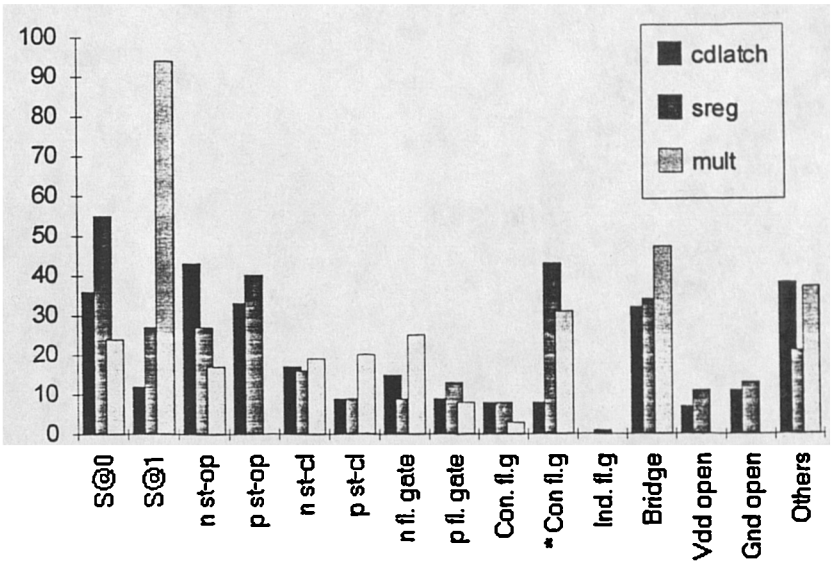


Figure 3.6 Fault type distribution for each cell

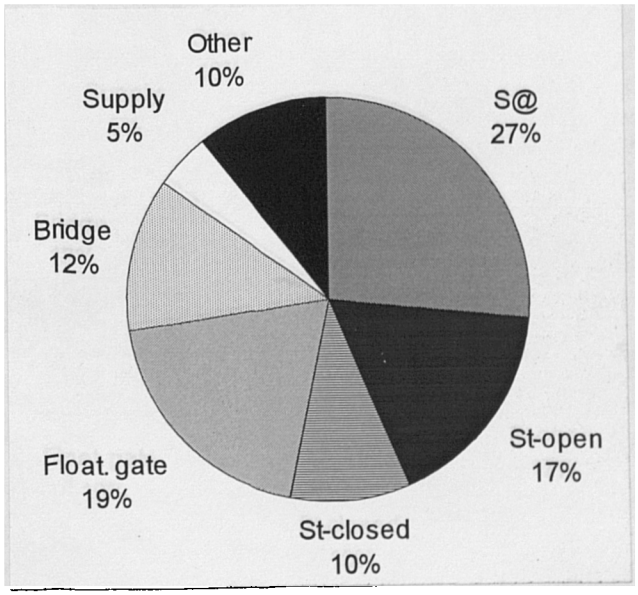


Figure 3.7 Summary of fault distributions for all cells

3.2.3.4 Results of the Simulations

The 16100 defects simulated produced 930 faults in the categories tabulated. The histogram in figure 3.6 highlights two significant anomalies in the distributions. There is a large number of stuck-at 1 faults in the MULT cell. As can be seen from Table 3.4, the majority of these are generated by missing layer defects in the *p*-well. The large number occurs because of the large amount of active area in the well, compared to that in cells DLATCH and SREG and is to be expected. However, some caution should be expressed for this figure for the following reason.

The minimum defect diameter for the MULT cell was set to 3.0 μm and so most defects would be this diameter or slightly larger (4.0 - 5.0 μm). Owing to the early deposition of the well implant in a CMOS process there is significant diffusion of the dopant in the well in subsequent stages of the processing. This would be typically 3 - 4 μm isotropically. Therefore holes with a diameter of less than 6.0 μm in the well will "self heal" and may disappear. The well would, however, be lightly doped in this region and the spot is likely to be a point of soft breakdown, such as excess junction leakage current which may lead to latch-up. The precise effect of small holes in the well is therefore difficult to predict without accurate data on the amount of well diffusion. The worst case of the hole remaining as a fault was therefore taken as the effect of such defects. From this argument any reduction in the number of faults arising from this defect should apply to other cells also so that the high proportion of this fault with respect to the other cells is correct.

The second anomaly is the relatively high number of multiple connected floating gate faults in cell SREG, particularly when compared to the DLATCH cell. These arise because of the large amount of intercell signal connection within the cell and this large number is a correct representation of the significance of the fault.

These anomalies are apparent when the fault distributions illustrated in figure 3.6 are examined. The MULT cell has a high level of stuck-at faults when compared to the other cells. The percentage of floating gate faults in the cell SREG is also relatively high. There are other significant variations between the distributions. The MULT cell has few stuck-open faults. This is probably because the cell contains wide active tracks and devices with parallel contacts for drain and source interconnects which will be less susceptible to open-circuit than the minimum dimension tracks with single contacts used in other cells. There are no supply faults in the MULT cell because the supply tracks are 10 μm wide and multiple contacts are used. Conversely, cell DLATCH has a large number of stuck open faults due to the large area of the drain-drain links in the cell and the use of single contacts. The use of minimum width transistors also contributes to this bias in the distribution.

The percentage of faults in the floating gate category is very significant in all cases. The average of the three distributions is given in figure 3.7. If we rank the average number of

faults in each category in order of significance we see that floating gate faults are the third most significant fault effect produced by photolithographic defects in CMOS circuits. Bridging, stuck-closed, supply and other defects are, individually, considerably lower in significance. Stuck-at, stuck-open and supply faults generally account for around 49% of all faults. Floating gates account for 19% and bridging, stuck-closed, supply and other faults constitute around 32% in total. We should note that these distributions are for the faults that are generated by photolithographic defects. They do not necessarily represent the effect of the faults which may be different. This will be investigated in section 3.3.

3.2.3.5 Summary of Fault Distribution Experiments

It is clear from the results presented that the stuck-at fault is not representative of the majority of faults produced by photolithographic defects in CMOS circuits. Stuck-open faults (including supply faults) account for around 22% of CMOS faults. This certainly justifies the significant effort devoted to model improvements in this area in recent years and contradicts Shen's conclusions in this matter. The nMOS cell used by Shen is similar to the MULT cell in layout characteristics with many *n*-channel devices interconnected largely through wide active tracks. This inevitably leads to a small number of stuck-open faults as shown by the results presented above.

The floating gate fault is clearly a very significant fault effect and it requires further investigation. The range of possible manifestations of this fault, the effect of which will depend on circuit interconnection and layout topology, makes it difficult to predict the wide range of effects that may result.

Bridging and stuck-closed faults, which have similar characteristics in some circumstances, are also significant and must be accounted for in a comprehensive IC test strategy.

3.2.4 Electromigration

It is difficult to be precise in quantifying the significance of any ionic transport phenomenon. For electromigration, the likelihood of occurrence in a signal line compared to that in a power track needs to be considered. This can be achieved by considering two separable parameters: i) the relative probability of electromigration in a signal line compared with a power supply track; ii) the relative proportion of signal tracks to power tracks on a typical IC. Although electromigration in polysilicon is possible [3.7] we will discount this failure as it is unlikely compared to electromigration in metal.

To consider the first parameter we need to resume the argument presented in section 3.1 (iii). In this section it is demonstrated that electromigration may occur in signal tracks in

which unequal current densities flow in opposite directions due to asymmetry in propagation delays. From equation 3.3 and the similar equation for the discharging flux, we can see that the net ionic flux in a signal track is:

$$F_{net} = c.\left(I_c^2 t_c - I_d^2 t_d\right) \tag{3.5}$$

where c is a proportionality constant which includes a factor of $1 / A^2$ where A is the cross-sectional area of the track. By substituting the charge equality equation we obtain:

$$F_{net} = c.I_d^2 t_d \left(\frac{t_d}{t_c} - 1\right) \tag{3.6}$$

From the equation we can see that when $t_d = t_c$ (i.e. symmetric operation) the net ionic flux is zero. However, for $t_d < t_c$ or $t_d > t_c$ the net ionic flux is non-zero and electromigration can occur. For example, consider the inverter shown in figure 3.8 in which the delay when the output is falling is one third of the rising delay (this is approximately the situation for an inverter in which both n- and p-channel devices are minimum size).

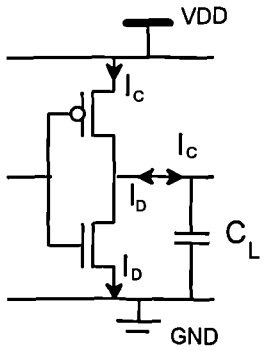


Figure 3.8 Inverter showing charging and discharging current flows

(Note. The direct current between VDD and GND which flows while both n- and p-channel devices are on has been ignored for clarity in this argument and is not shown on the diagram).

For this circuit the net ionic fluxes for the GND, SIGNAL and VDD tracks are:

$$F_{GND} = c.I_d^2 t_d \qquad F_{SIGNAL} = \frac{2}{3} c.I_d^2 t_d \qquad F_{VDD} = c.I_c^2 t_c = \frac{1}{3} c.I_d^2 t_d \tag{3.7}$$

If we assume that the signal track width is minimum width and the supply tracks are twice this, then the mean-time-to-fail (MTTF) for the signal track, which is proportional to the ionic flux desnsity, is 0.125 times that of the V_{DD} track and 0.5 that of the GND track. This means that the signal track will fail before either of the supply tracks if we consider this circuit in isolation. In most circumstances, the GND and V_{DD} tracks will supply current to more that one logic gate which will decrease their respective MTTF's. However, it is the author's

opinion that these simple calculations show that electromigration in signal tracks is a potential source of failure that should not be ignored.

Other factors also affect the likelihood of failure. The current density in power tracks is generally higher than that in signal lines, as power tracks normally service more than one logic gate. On the other hand, signal tracks are generally minimum width, whereas power tracks are often three times this or significantly more. The morphology of the surface that a track traverses also affects its susceptibility to void formation and will obviously vary between tracks.

As can be seen, there are several unknown and highly variable parameters in electromigration. The number of logic gates that are asymmetric and the extent of the asymmetry are variable design parameters. The power distribution techniques used in the IC design are wide ranging. The ratio of power tracks to signal tracks in also difficult to estimate, although as each power track is generally used to service more than one logic gate, and each logic gate produces at least one signal track, it is reasonable to assume that there will be more signal tracks than power tracks on an IC. The length of signal tracks will, however, generally be shorter than power tracks.

Clearly, without detailed analysis of IC layouts, it is difficult to reach a quantifiable conclusion on the relative importance of electromigration in signal and power tracks. It is the author's opinion that the probability of failure of a signal track due to electromigration is significantly greater than is generally accepted owing to the proposed theory. Consequently, while the relatively low significance of this effect should be noted, the fault effect should not be discounted.

3.2.5 Contact Migration

The arguments presented in the previous section can generally be applied to contact migration. That is, it can be assumed that contact migration on signal lines is a significant possibility. In favour of this proposition is the fact that most signal tracks rely on a single contact, whereas multiple (parallel) contacts are often used for power supplies. This is not a requirement for the previous arguments to apply. It is however, equally difficult to quantify the probability of occurrence of the fault. Therefore, it can only be concluded that the failure mechanism is sufficiently significant to be considered as a source of floating gate faults.

3.2.6 Stress Migration

Stress migration is not dependent upon current density. Therefore, all tracks with the same width have equal probabilities of stress migration, which can lead to voids and open circuit faults. Most signal tracks are narrower than power tracks and this would tend to

increase the probability of failure of signal tracks as there is less material to be removed before voids can be created.

The above argument may not hold as track widths are reduced to submicron dimensions because of the "bamboo effect". This phenomenon is based on the generally accepted theory that migration occurs along grain boundaries in the metal. In these regions, atoms are less tightly bound and are therefore more ~~easily ionised~~ mobile. They are therefore able to move more freely at boundaries than within the crystalline region. As track widths approach the grain size of the track material, grains start to occupy the full width of the track and so the paths for ionic migration along the track are removed. Therefore the likelihood of migration is reduced in such tracks. This theory has not been applied specifically to the description of stress migration, but it is widely used in the description of electromigration. However, the author can see no reason for the arguments not to apply to stress migration failures.

For current technologies it is reasonable to assume that stress migration is more likely to occur in narrow signal tracks than in wide power tracks. The other parameter to be considered is the relative numbers of tracks, or more specifically the total lengths of the signal and power tracks that are present in a circuit. As indicated previously, it is difficult to reach a general conclusion for this parameter. The generation of floating gate defects due to the presence of stress migration must be considered to be a significant occurrence.

3.2.7 Summary of the Significance of Floating Gate Faults

From the discussion presented in this section, and in particular the results of the defect simulations, it is apparent that a significant number of defects which occur during production, or which may rise during the operation of an integrated circuit will result in floating gate faults. Furthermore, the floating gate fault effect is third in ranked order of photolithographic faults for the cells examined. From these conclusions it is apparent that the fault should be investigated in detail and its effects understood.

3.3 The Electrical Effect of the Floating Gate Fault

A *floating gate* is defined as an MOS transistor gate which has no d.c. path to the VDD or GND supplies. This definition includes: individual *n*- or *p*-channel transistor gates that are completely isolated; pairs of *n*- and *p*-channel gates that are connected together but otherwise isolated and multiple sets of floating gates that are connected together but isolated from all other nodes. Floating gates may also occur for which the above definition does not strictly apply, but in which the path to the supplies is a very high impedance. An example would be a gate which is connected to an active layer track which is itself isolated from other nodes. A reverse biased diode actually connects the gate to one of the supplies but this represents a very high impedance.

A half floating gate may also arise in which the path to only one of the supplies is an open circuit (or very high impedance). Such faults would normally be called stuck-open but, as the fault causes a floating gate when in one logic state, the title is a good description of the fault.

The effect of a floating gate fault is often assumed to cause the gate to be stuck-at 0 or 1. Consequently, the affected transistor is said to be either stuck-open or stuck-closed. This assumption is made by many authors (for example [2.5, 2.7]). However, it is incorrect for two reasons:

i) The potential adopted by the floating gate will depend on the charge on the gate and the capacitance to the other nodes in the circuit. As both of these are unknown quantities no assumption can be made about the conducting state of the channel.

ii) Capacitive coupling of the floating gate to other nodes in the circuit will cause the gate potential to vary, resulting in a variable conductance for the transistor rather than a single stuck state.

The following example illustrates the inadequacy of the stuck fault model for a floating gate transistor. Consider two inverters in series in which a photolithography defect has resulted in a missing contact cut between the metal output of the first inverter and the and polysilicon input of the second. The two tracks are coupled via a capacitance due to the overlap of the metal and polysilicon layers in the contact region. The fault and resulting circuit are illustrated in figure 3.9. The coupling capacitance for a typical 3 μm CMOS process would be approximately 2.5 fF.

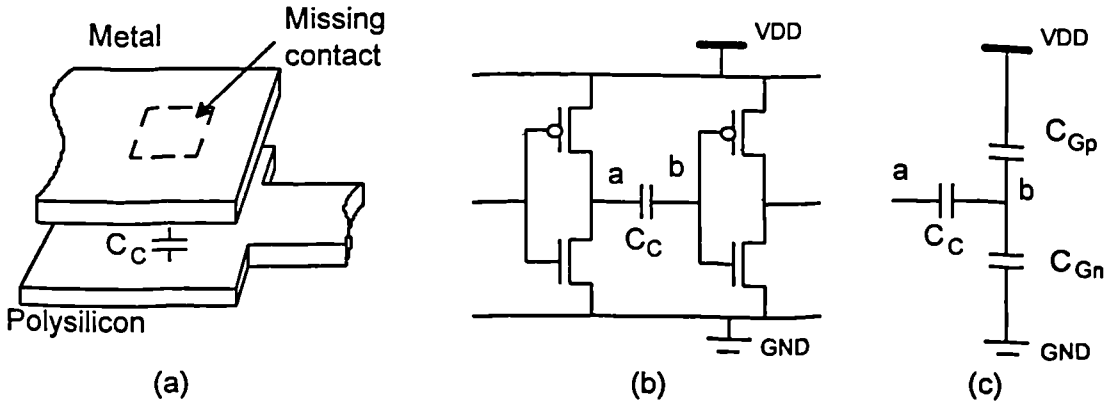


Figure 3.9 (a) coupling capacitance created by missing contact cut; (b) capacitively coupled floating gate inverter; (c) capacitive potential divider formed at the floating gate node

The series capacitor, C_C, forms a capacitive potential divider with the gate capacitances, C_{Gp} and C_{Gn}, as shown in figure 3.9(c). Changes in the voltage at *a* are coupled through to node *b* and are attenuated according to the ratio given in equation 3.8:

$$\Delta V_b = \Delta V_a \cdot \frac{C_c}{C_c + C_{gn} + C_{gp}} \tag{3.8}$$

Typical gate capacitance values for a 3 μm process are C_{Gn} = 8fF (W = 4 μm, L = 3 μm, T_{ox} = 50 nm) and C_{Gp} = 20 fF (W = 10 μm). Consequently, the attenuation ratio for the circuit of figure 3.8 is 0.076. This is quite a large attenuation factor but it produces a significant result. If the output of the first inverter changes by 5 V, the floating gate potential will change by 385 mV. If the floating gate potential is initially near to either of the supply rails this will not have a significant effect on the circuit operation and the gate will appear to be stuck-at 0 or 1. However, if the floating gate potential is initially near the transition voltage for the inverter, a change of 385 mV is very significant and when amplified by the gain of the logic gate can cause the inverter to produce the fault free logic outputs. This result has been verified by simulation using the PSPICE circuit simulator.

This simple experiment illustrates that when the floating gate potential is allowed to vary, the effect of the fault will change significantly compared to the "stuck conductance" model. A more sophisticated model of the floating gate behaviour is developed in Chapter 6 of this thesis.

The absolute value of the potential adopted by the floating gate depends on the amount of net charge on the gate electrode. If we consider all nodes on an IC to be connected to 0 V then the floating gate potential is simply given by $V_{FG} = Q_{FG} / C_{TOTAL}$, where V_{FG} and Q_{FG} are the floating gate potential and charge respectively, and C_{TOTAL} is the capacitance of the floating gate to all other nodes in the circuit. We must now answer the question: "What value of charge or potential will the floating gate adopt?".

For the case of an isolated gate resulting from a processing defect, the resultant charge is difficult to predict because of the variations in processing steps used in producing an IC. The use of the polysilicon gate as a mask in a self-aligned process however, is likely to result in net charge on the isolated gate after ion implantation of the active regions. The charge due to ion implantation will generally be positive as the wafer is normally held near to the ground potential of the system during implantation.

The residual charge deposited in the drain and source active regions can leak away after implantation through the substrate. However, the residual charge in the polysilicon gate, does not have a leakage path and consequently it remains trapped within the gate material. Subsequent wet processing steps such as photoresist removal may discharge the gate. However, plasma etching is often used to remove photoresist thus exposing the gate region to further sources of charge which may increase or reduce the existing charge on the gate. Furthermore, subsequent implant processes may affect the stored charge.

From the range of processes that can occur after creation of the gate region it is difficult to predict how much positive or negative charge will reside on a fully processed floating gate. However, the large number of sources of charge available during the normal processing stages of an integrated circuit provide many opportunities for the deposition of residual charge. It is therefore most likely that the floating gate will have some residual charge present at the end of processing and that this charge may give rise to a wide range of initial floating gate potentials.

The floating gate potential is likely to be different for isolated *n*- and *p*-channel gates as these are exposed to different implant processes. Also, connected floating gates will experience a different charging sequence as will multiply connected floating gates. It cannot therefore be assumed that the floating gate potential is the same for all floating gate devices on the same circuit.

A form of floating gate which occurred several times during defect simulations is illustrated in figure 3.10. As shown, the gates of M3 and M4 are not completely isolated but have a path to the supplies through the reverse biased drain and source *p-n* junction diodes. The potential adopted by the gate will depend on the relative values of the ~~reserve~~ reverse saturation currents for the two diodes. If the diode that is connected to the V_{DD} supply has a larger reverse current than that of the diode connected to the GND supply, the gate will charge towards V_{DD}. The converse will also be true. The gate may therefore be at V_{DD} or GND potential, but in both cases is biased by a very high impendance source.

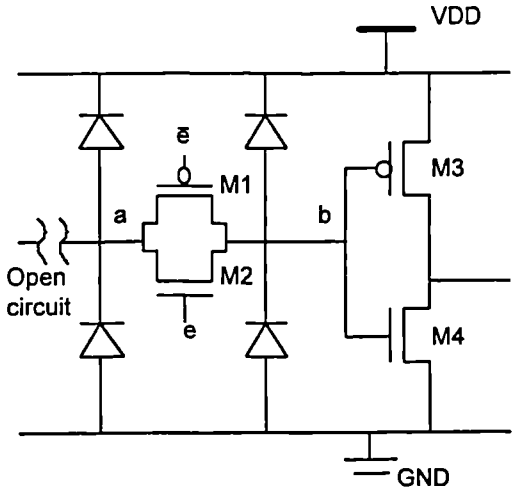


Figure 3.10 Transmission gate stuck-open resulting in a connected floating gate

The potential of a floating gate which arises as a result of a wearout failure such as electromigration, may also have a range of values. Henderson et al [2.23] have demonstrated that tunnelling conduction can occur across a void in a metal track caused by electromigration. Such currents are small but they will allow the gate potential to follow the track potential over a relatively large time scale: of the order of 10^{-4} s. The gate potential will

be a time averaged value of the track signal integrated over this period. The gate potential could therefore adopt any potential between 0 V and V_{DD} for this form of failure.

For all of the floating gate faults envisaged, mechanisms therefore exist which allow the gate to adopt a wide range of potentials and there is no reason to suggest that any particular potential should be favoured. It was decided that experimental verification of the operation of floating gate transistors may lead to a greater understanding of the failure mechanism and its effect. A set of test circuits were therefore designed for this investigation and they will now be described.

3.4 Test Circuits for the Analysis of Floating Gate Faults

The aim of the test circuits was to provide data on the operation of (i) floating gate transistors and (ii) circuits containing floating gate transistors. It was intended that this data would provide supporting evidence for the models and simulations described in this thesis. In particular, the circuits were designed to provide information on the following factors:

a) the value of the floating gate charge;
b) the variation of the floating gate potential with the drain or source potential;
c) the effect of the overlap capacitance on the drain/source to gate coupling;
d) the effect of capacitive coupling to the floating gate from overlapping or adjacent tracks.
e) the effect of single *n*- and *p*-channel and connected floating gates on logic gate operation;
f) the effect of stuck-open or half-floating gates on circuit operation.

A test structure was designed to provide an environment for the circuit under test which closely emulated the core of an integrated circuit whilst allowing the observation of inputs and outputs and the monitoring of supply current. Additionally, the structure was designed to allow voltage contrast measurements in a scanning electron microscope (SEM). It was intended that this technique would be investigated as a means of analysing the floating gate operation as it potentially allows the measurement of voltages on buried layers in the IC in a non-intrusive way. Little modification of the circuit was required to enable voltage contrast measurements to be made other than to ensure that voltage reference nodes would be visible within an SEM image for all regions to be analysed. Consequently, extra GND stubs in the second metal layer were included on the layout at strategic locations.

To emulate the core of an IC the test circuit was buffered with input and output inverters. In addition, the output of the circuit under test could be directly probed. Separate

VDD supply pads were provided for each of the test circuits designed, but common GND and input pads were used for all circuits. The basic structure of the test circuit is shown in figure 3.11.

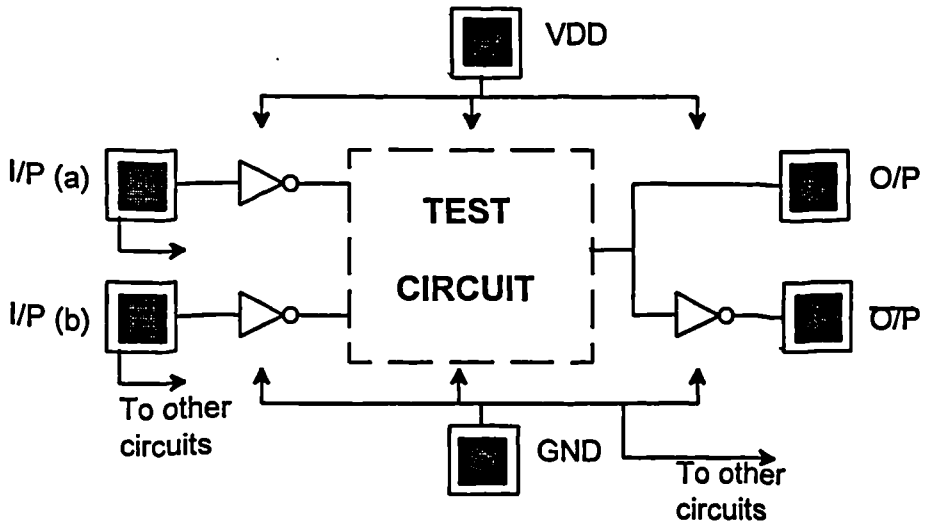


Figure 3.11 The basic test structure providing a core environment for the test circuit.

The choice of the circuit to be used for the introduction of faults required careful consideration. The circuit should be sufficiently complex to provide for the range of fault effects that can occur in any commonly used circuit whilst being simple enough to allow clear observation of the fault effects produced. An inverter would be of little use as it provides very limited control of the gate output under fault conditions. Properties such as charge sharing between the drains and sources of series connected devices cannot be investigated, and the influence of a fault free transistor in parallel with a faulty device cannot be assessed.

A simple two-input logic gate (i.e. NAND or NOR) overcomes these problems. Control of the output state is provided via the fault free input and the CMOS gate structure provides both series and parallel connected devices. All of the fault conditions and parameters listed can be analysed using such a gate. Increasing the number of inputs to three or more does not increase the number of fault conditions or parameters that can be considered.

Increasing the complexity of the gate to, for example, a four input AND-NOR complex gate does not increase the utility of the circuit for floating gate analysis. However, four such gates were included in the test circuits to allow analysis of other fault conditions such as bridging faults which change the function of the affected logic gate.

Having selected the two-input NOR gate as the most useful test circuit, the faults to be introduced were considered. After some consideration, it was decided that the faults would be primarily confined to those that emulate photolithographic errors in a predictable way. Faults that might produce marginal errors such as narrow gaps in tracks (i.e. $< 0.5\mu m$) were

not included as their outcome could not be predicted (and space for the test circuits was limited). The faults were chosen to provide results that would assist in the verification of a simulation model. This would allow the prediction of more esoteric fault effects in a controllable way: i.e. with circuit simulation.

Over the course of the work reported in this thesis, four sets of test circuits have been fabricated. Three of these are very similar but they contained modifications intended to overcome a latch-up problem encountered in testing the first circuits and other minor variations. The final set of test circuits was more varied, providing alternative circuits for more detailed analysis of individual devices.

The first set of test circuits designed were a set of twelve two-input NOR gates contained within the test structure described. The NOR gates had the following intentional faults, as shown in figure 3.12.

a) Single floating gate *p*-channel transistor at the top of the series combination created by a large (3.0 μm) gap in the polysilicon track to the gate.

b) As (a) but for *n*-channel transistor.

c) Connected *n*- and *p*-channel floating gates created by a large (3.0 μm) gap in the polysilicon track to the gate.

d) Connected *n*- and *p*-channel floating gates created by a missing metal to polysilicon contact cut.

e) Stuck-open *p*-channel transistor created by missing source active to metal contact cut.

f) As (e) but for *n*-channel device.

g) VDD supply open circuit by omission of an active to metal contact cut.

h) As (g) but fault in GND supply.

i) Short circuit between polysilicon inputs of the NOR gate. The input inverters were modified to provide asymmetric driving conditions.

j) Gate to drain short circuit in *n*-channel transistor created by misplaced contact cut.

k) NOR gate with all transistor gates individually controllable.

l) Fault free gate.

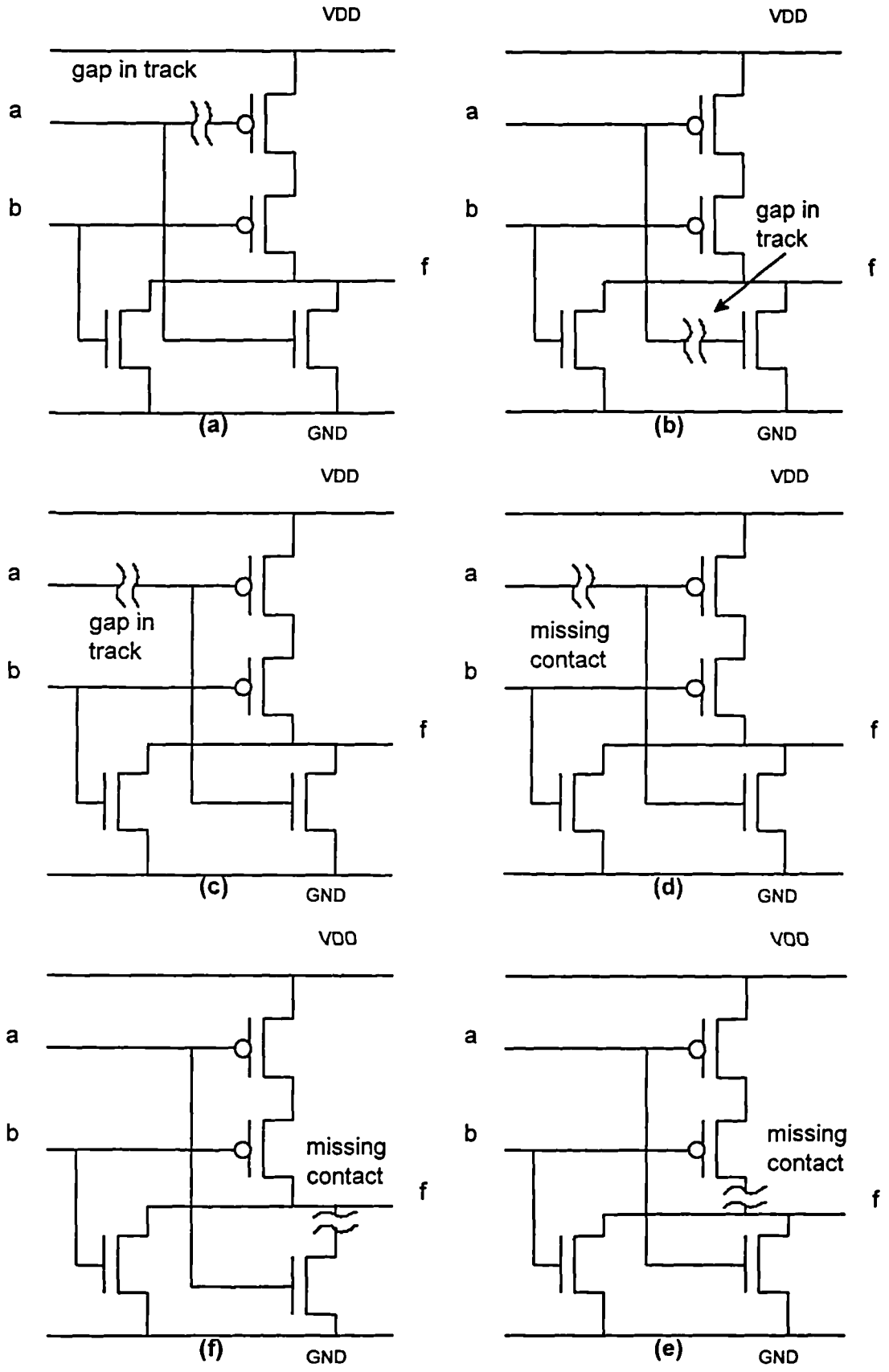


Figure 3.12(a) Faults introduced into NOR gate test circuits

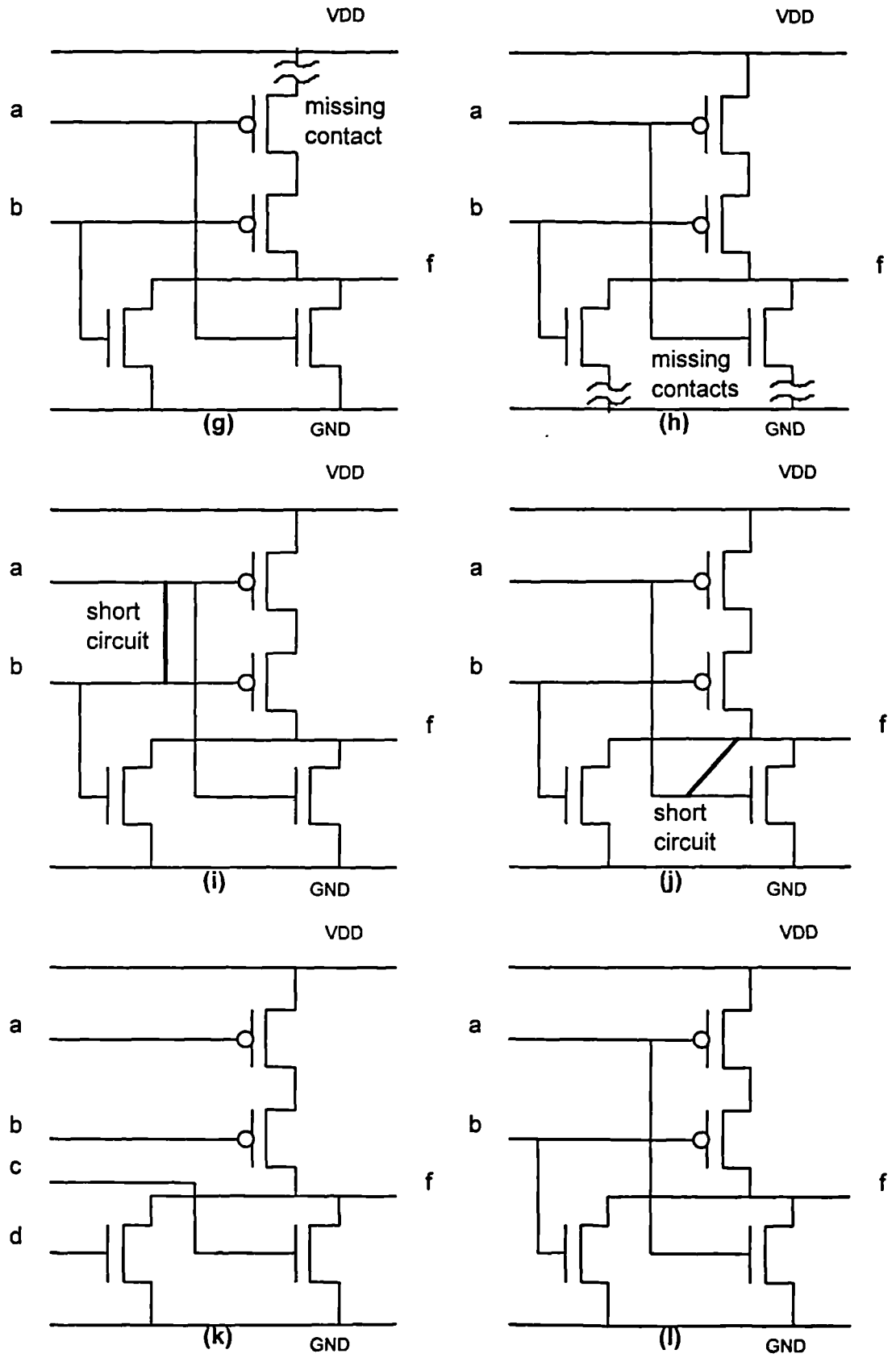


Figure 3.12(b) Faults introduced into NOR gate test circuits

Circuits (a) to (d) allow for the investigation of circuit operation under various floating gate conditions. Also, by selective probing, the operation of individual floating gate transistor can be ascertained. Gates (e) to (h) allow for the analysis of stuck-open or half-floating gate faults. In these circuits, it is the output inverter that has the half floating gate fault condition. Circuits (i) and (j) provide for analysis of typical bridging faults, while circuits (k) and (l) provide complete control of the logic gate inputs and a reference gate.

The second set of circuits included faults (a), (b), (c) and the fault free gate (l) from the previous list. The same faults were also introduced into a two-input NAND gate and an inverter. In addition, a set of *n*- and *p*-channel transistors were included. These allowed for the investigation of the effects of gate to drain/source overlap capacitance variation on the floating gate potential and the observation of signal coupling to floating gates from overlapping tracks. These circuits were included after development of a theoretical model for floating gate MOS transistor operation which is described in Chapter 6. It became apparent during the work on this model that overlap and coupling capacitances are significant factors in the operation of these devices. An example layout diagram for the test circuits designed is given in figure 3.13. Results of measurements on these test structures are given in Chapter 5.

3.5 Chapter Summary

In this chapter, the floating gate fault has been considered in detail. It has been shown that there are a wide range of causes for the fault both from production defects and wearout failures. It has been demonstrated, by detailed simulation and analysis, that the fault is significant in its rate of occurrence in relation to other fault types. It has been found to be ranked third in order of significance with approximately 20% of photolithographic defects resulting directly in floating gate faults. A new model for evaluating electromigration in signal lines has been proposed and described. As electromigration is accepted as one of the dominant wearout failure mechanisms for current and future IC's, it can be concluded that floating gate faults resulting from field failures are a significant possibility.

The forms of floating gate fault that can occur have been shown to be quite varied. Consequently, the effect of the fault is difficult to predict. Additionally, the effects are dependent on variable, and hard to predict, parameters such as residual gate charge and stray capacitive coupling in circuits. These unknown quantities make prediction of the effect of the faults even less reliable. A set of test circuits has therefore been designed to provide experimental data on the unknown parameters of the floating gate transistor. The results of measurements on these circuits will be described in the next two chapters. The application of the voltage contrast technique, using a scanning electron microscope, to the measurement of isolated nodes will be considered Chapter 4. In Chapter 5, electrical analysis of the test structures will be described, and the results of the measurements discussed.

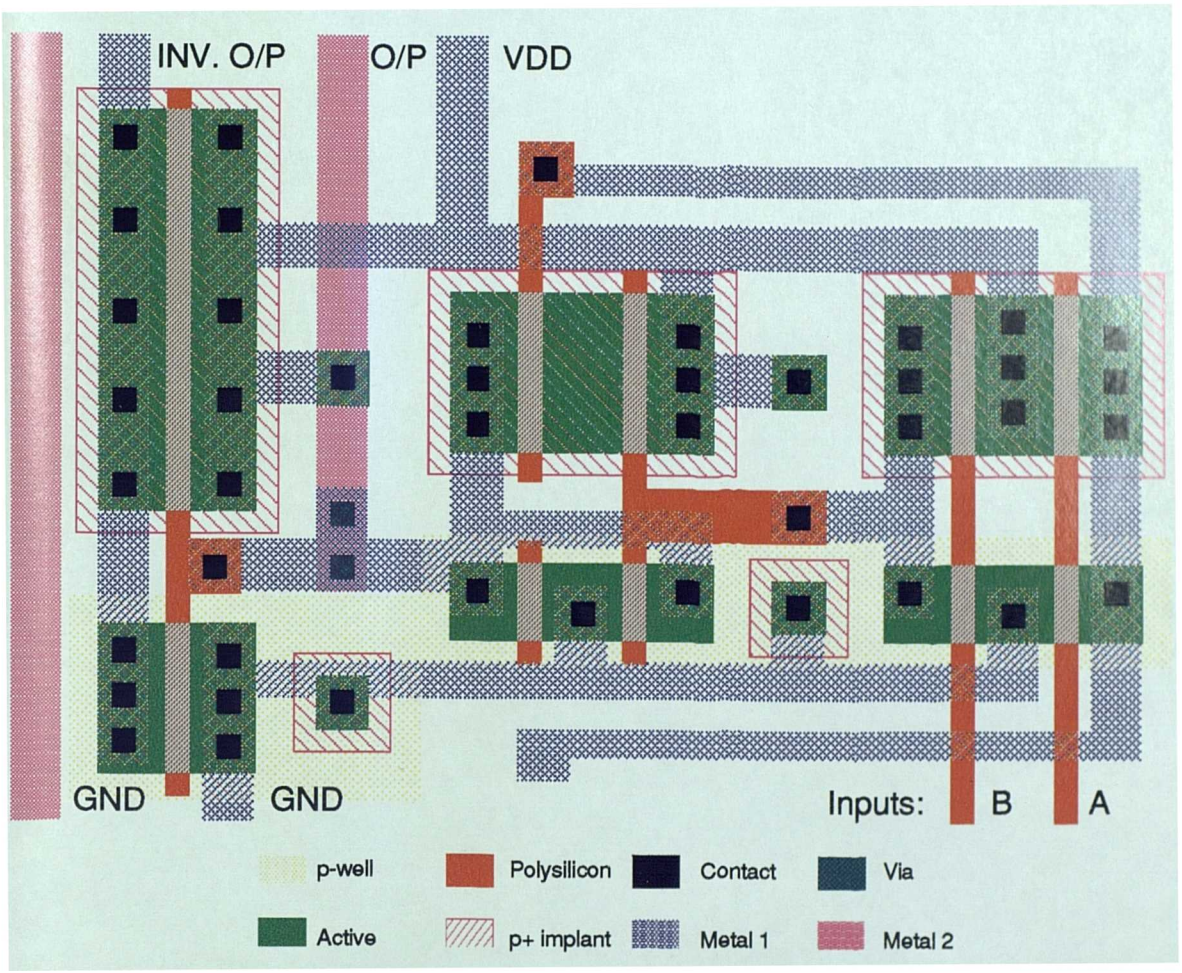


Figure 3.13 Standard test structure with single *n*-channel floating gate introduced into the NOR gate

Chapter 4

Voltage Contrast Microscopy of Integrated Circuits

As described in the previous chapter, a range of test circuits have been designed primarily to provide data on the operation of floating gate transistors. The circuits were fabricated over a period of four years through a silicon brokerage company (Micro Circuit Engineering) and the University of Edinburgh Microfabrication Facility. Minor modifications were made to correct what was initially thought to be a design error and various changes made to the set of devices on each process run. The final set of circuits are somewhat different from the original set and will be described in Chapter 5.

In this chapter we discuss the application of voltage contrast microscopy to the analysis of these circuits and devices. At an early stage in the investigation of floating gate faults it was felt that a "direct" measurement of the floating gate voltage would be useful. This is clearly very difficult for two reasons: i) the floating gate is completely isolated and surrounded by insulating material making direct contact with a probe impossible without damage to the device; ii) employing techniques to penetrate or remove the insulating layers would affect irreversibly the small amount of charge stored on the gate - the quantity which we are attempting to measure. It was realised that a technique known as voltage contrast microscopy had the potential to overcome both of the problems and so it was decided that the technique would be investigated to see if it could provide the required measurements.

4.1 Scanning Electron Microscopy Analysis

The image viewed in a scanning electron microscope (SEM) is normally generated by collecting *secondary electrons* from the sample. These are low energy electrons (a few eV) that are generated by the interaction of high energy *primary electrons* (typically > 1 kV) from the SEM anode. The secondary electrons are gathered by a collector and accelerated in a photomultiplier which is used to generate an output signal. Other imaging techniques are also employed utilising backscattered electrons, auger electrons, x-rays and sample currents, all of

which are induced by the primary electron beam. However, for simple microscopy, secondary electron imaging is normally used.

The low energy of the secondary electrons allows them to be significantly affected by electric fields at the surface of a sample. These fields are normally eliminated by coating the surface with a thin conducting layer (for example, gold) but the effect can also be exploited for the analysis of voltages on integrated circuits (and other devices) in the SEM. This application of the voltage contrast phenomenon was first reported by Everhart et al. [4.1]. The principle of the technique will now be described.

4.1.1 The Principle of Voltage Contrast in the Scanning Electron Microscope

Secondary electrons emitted from a metal track at 0V will not experience any local field. They will be emitted from the surface with the normal cosine variation in trajectory [4.2] and a proportion will reach the collector. Electrons emitted from a track at a positive voltage (say 5V) will experience a local retarding field and those with energy less than 5 eV may not escape the field. These electrons will fall back to the track or nearby surface. The number of electrons reaching the collector from this track will therefore be lower than that from the 0V track. The track with positive bias will therefore appear darker on the SEM image and a contrast will appear due to the potential variation over the surface of the integrated circuit. This process is illustrated in figure 4.1.

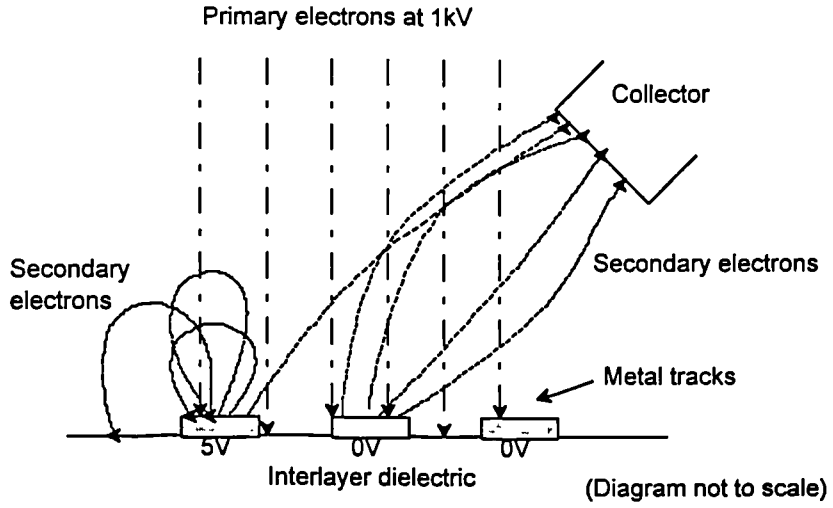


Figure 4.1 Secondary electron paths in the SEM with local electric fields at the sample surface - the voltage contrast principle

The effect of voltage contrast is seen in the microscope image given in figure 4.2 in which the central bond pad is at 5V while many of the surrounding tracks are at 0V. Contrast is clearly visible. The image was generated and captured on the author's SEM image capture system which is described in section 4.2.

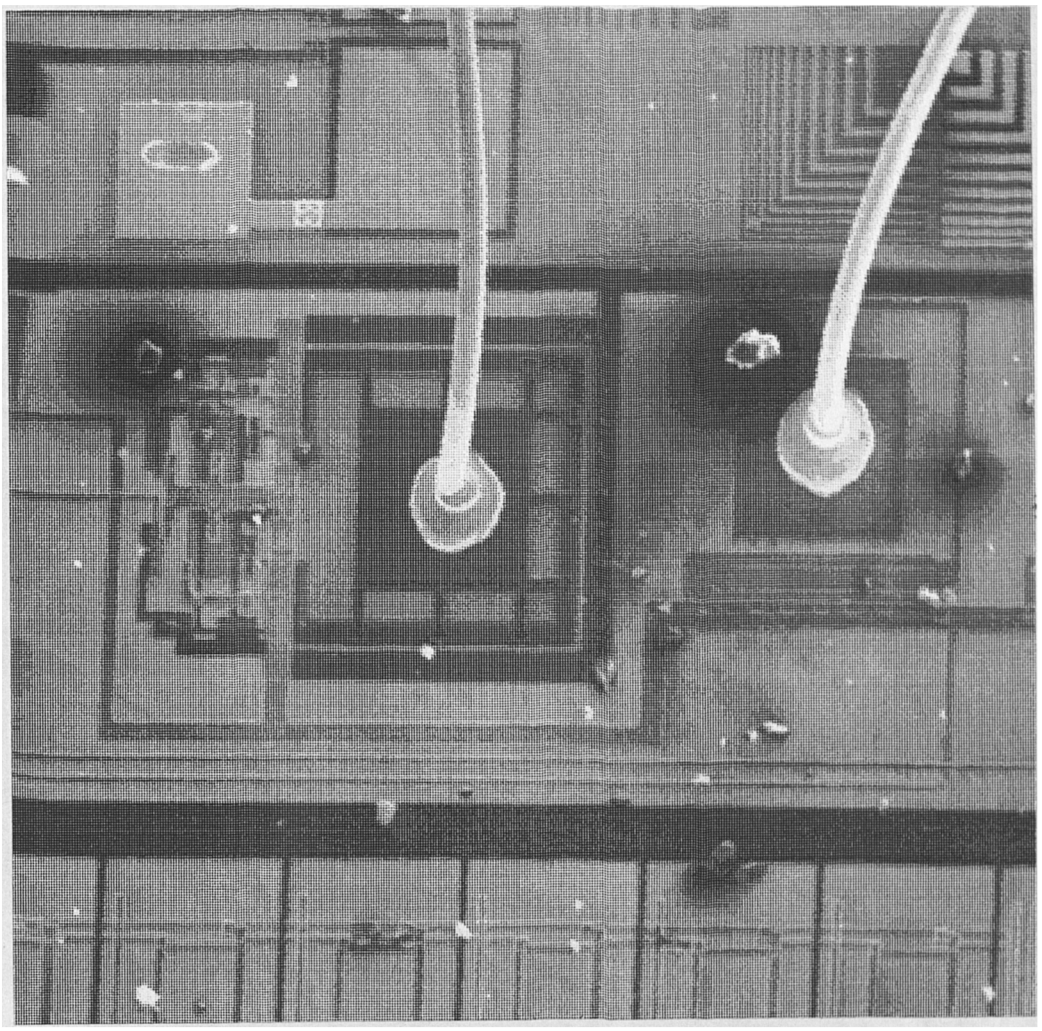


Figure 4.2 Voltage contrast in the scanning electron microscope. The central bondpad is at 5V with many of the surrounding tracks at 0V.

The contrast produced by this phenomenon can be used to determine the voltages on the sample under inspection, and hence the logic state of nodes on an IC can be determined. The technique is extensively employed for failure analysis of integrated circuits. In its simplest form, voltage contrast can be used to identify faults such as breaks in tracks or incorrect logic values due to poor logical or timing design of circuits. The die of the circuit to be inspected must be visible and preferably unpassivated, although the latter condition is not essential. The circuit is simply powered up in the SEM and the appropriate logic values required to reveal the suspected fault are applied.

It is important to note that voltage contrast images are generally composites containing both topographic, material and voltage contrast. Topographic contrast occurs because of the cosine variation of the angular trajectory of the secondary electrons. Elements of the chip surface that are at different angles produce different numbers of secondary electrons in the direction of the collector and therefore appear with different intensities. A good example of this is the edge of a metal track which appears significantly brighter than the centre region. Material contrast occurs because of the different work functions of materials in the sample. This leads to a unique secondary electron spectrum for each material causing differences in the numbers of electrons collected for each type of material. Despite this combination of contrasts, the voltage contrast can be clearly identified and used for circuit analysis.

The technique described is suitable for static or slowly varying signals but cannot be used to analyse high speed signals. This can be achieved by the stroboscopic technique in which the electron beam is scanned in synchronism with the dynamic signals applied to the IC. Provided that the circuit operation is periodic, a "snap shot" of a particular instance in the cycle of the circuit can be captured.

This technique can be extended to capture the node voltage over a period of time by varying the phase of the SEM scanning system and the applied signals. The waveforms of signals on, for example, data lines can be captured and compared with logic simulations for the same period. Such *logic state mapping* [4.3] can identify marginal timing errors in designs as well as logic design errors for faults due to fabrication problems that cannot otherwise be identified.

The techniques described so far can be used to determine the logic value of nodes within an IC. Voltage contrast can also be used to provide quantitative measures of voltages on samples. There are two main techniques that can be used for this: i) simple exploitation of the intensity / voltage relationship or ii) analysis of the energy spectrum of the secondary electrons. The two techniques will now be described.

4.1.2 Quantitative Voltage Measurements Using Image Contrast.

The intensity of a track on an IC in an SEM image is determined by the number of secondary electrons which reach the detector from that track. The number of electrons detected is reduced by a local positive bias on the node. This is illustrated in figure 4.3 which gives the secondary electron spectra for tracks at 0V and 5V.

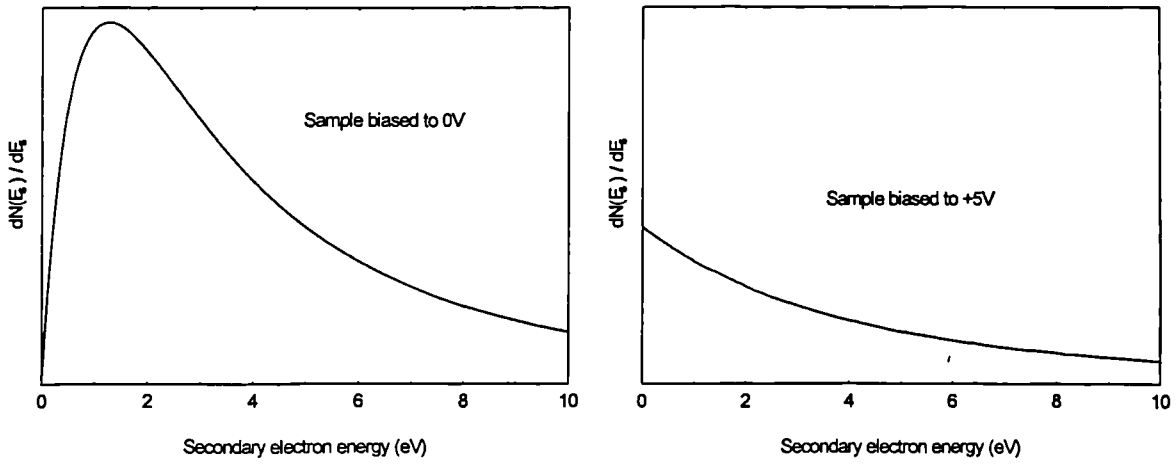


Figure 4.3 Secondary electron energy spectra for a metal track at 0V and 5V

The energy spectrum is described by Chung and Everhart [4.4] and is given by:

$$\frac{dN(E_s)}{dE_s} \propto \frac{E_s}{(E_s + \Phi)^4} \tag{4.1}$$

where E_S is the secondary electron energy, $dN(E_S) / dE_S$ is the number of electrons at energy E_S (this is expressed as a differential because the electron energy is continuous) and Φ is the material work function. For all of the calculations and results reported, the work function for aluminium is taken as 3.8 eV [4.4]. Electrons emitted from the track at 5 V that have an energy of less than 5 eV have a lower probability of reaching the collector than equivalent electrons from the 0 V track. The intensity of a point on the image is determined by the total number of secondary electrons reaching the collector from that point given by the integral of equation 4.1. For a track at potential V, the intensity is given by the integral from qV to infinity:

$$I_V = c\int_{qV}^{\infty} \frac{dN(E_s)}{dE_s} dE_s \tag{4.2}$$

Figure 4.4 shows a plot of intensity against track voltage according to equation 4.2. Voltage measurements can therefore be made by measuring the intensity of a node on the IC and then obtaining the voltage from this characterisation curve. The curve does not however

give an absolute measure of voltage, but normalisation by reference to known voltages (such as the supply rails) eliminates this problem.

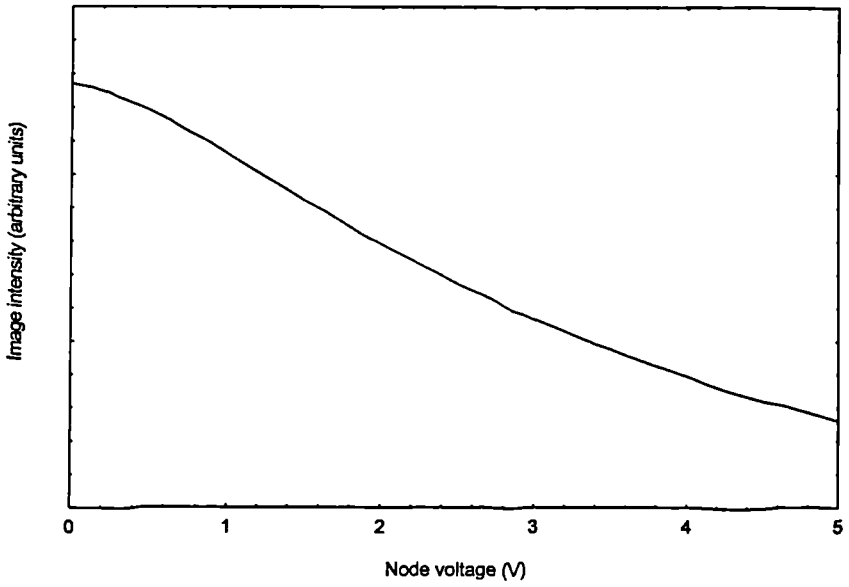


Figure 4.4 Variation of image intensity with node voltage in an SEM image

There are several problems associated with this measurement technique as described by Dinnis [4.5]. The relative nature of the measurement is one problem but this can be eliminated by reference to a track at a known voltage, such as a supply track. The effect of topographic and material contrast can also cause problems and ideally these should be removed from the image. This is not a simple problem to solve. A further significant problem arises from the interaction of secondary electrons with electric fields from adjacent tracks. Such *local field effects* can cause complex interactions between the intensity of neighbouring tracks [4.6] and must be taken into account for accurate measurements.

4.1.3 Voltage Measurements Using Secondary Electron Energy Analysis

The problems described have resulted in most workers in this field utilising secondary electron energy analysis to obtain quantitative voltage contrast measurements. These are based on the fact that the secondary electron spectrum is shifted along the energy axis by the potential of the node being analysed. By measuring this shift, the absolute change in the node voltage can be determined. Two basic types of analyser are used: i) the restricted aperture or band pass analyser such as that described by Wells and Bremer [4.7], and ii) the unrestricted aperture or high pass analyser, an example of which is the hemispherical analyser described by Feuerbaum [4.8].

In both systems a grid is placed between the sample and the SEM collector which is biased with a retarding potential. This potential is varied to maintain peak collector current for the restricted aperture analyser and constant collector current for the unrestricted aperture

system. By means of electronic feedback, changes in the specimen potential are directly followed by changes in the retarding potential and therefore by monitoring this, changes in the specimen potential can be measured.

Theoretically, both systems can provide an accurate measure of specimen voltage changes. However, sources of error exist for both systems. Local field effects can introduce changes in the secondary electron energy spectrum and hence cause errors in measurements. The effective energy distribution of the secondary electrons for the restricted aperture analyser, depends on the energy resolution obtained which introduces further errors and the requirement for characterisation. Furthermore, the limited number of electrons that are available for the restricted aperture case increases the noise in the received signal and hence the measured error.

The theoretical limit of resolution of the two systems has been analysed by Gopinath [4.9]. Ignoring the practical limitations described above, the minimum voltage resolution for the restricted aperture analyser is given as 80 mV. That for a hemispherical (unrestricted aperture) analyser is given as 0.9 mV. Most existing and reported analysers are of the latter type.

Despite the dominance of energy analysis as the technique employed for quantitative voltage contrast in both experimental and commercial systems, the preceding discussion illustrates that the primary difference between the two techniques is the provision of relative or absolute voltage measurements (intensity and energy analysis respectively). Energy analysis also has the advantage of removing the problems of topographic and material contrast.

After investigation of the technique, it was decided that the design and construction of an adequate energy analysis system would consume considerable time with little benefit. (A commercial system was not available for the microscope being used for this work.) Therefore, for the purposes of the work described in this thesis, it was decided that direct intensity and contrast measurements could be used to provide a measure of the floating gate potential. It was intended that the problems of topographic and material contrast and the relative nature of the measurements taken would be overcome by post-processing of the SEM images which would be digitised and captured onto a computer system.

4.1.4 Voltage Measurements on Buried Nodes

Before describing the measurement system used one further principle must be described. The floating gate is completely isolated by silicon dioxide and silicon nitride layers and, in particular, it is covered by an interlayer dielectric. It would therefore seem that the voltage contrast technique cannot be used to measure the floating gate potential as the

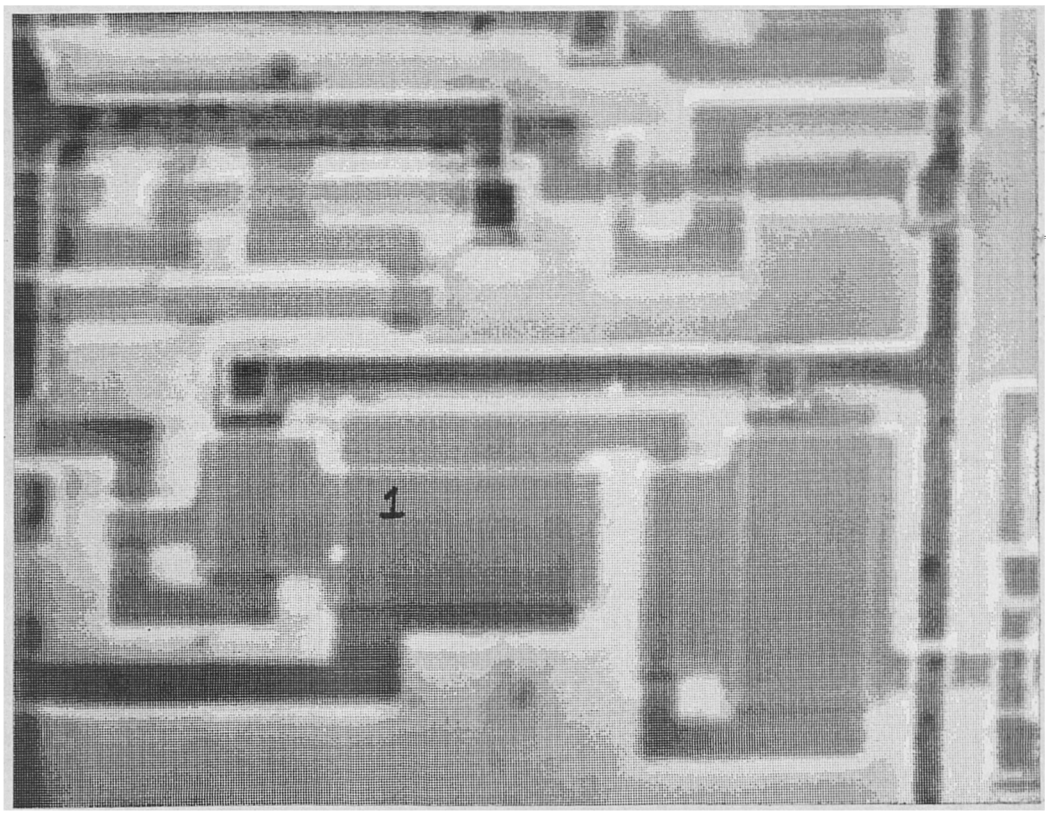
secondary electrons are not generated on the gate electrode but on the oxide surface above the gate. However, there are two solutions to this apparent problem.

The first uses a principle described by Pensak [4.10] in which an insulating film can be made to be conducting by electron bombardment. By setting the primary electron energy to ensure penetration through the oxide to the buried layer, the silicon dioxide is made conducting so that the oxide surface is at the same potential as the buried layer. The resulting voltage contrast can then be used to give a measure of the potential of the buried layer. This phenomenon is illustrated by the SEM images in figure 4.5. The images were generated using an SEM gun voltage (which corresponds to the primary electron energy) of 10 kV.

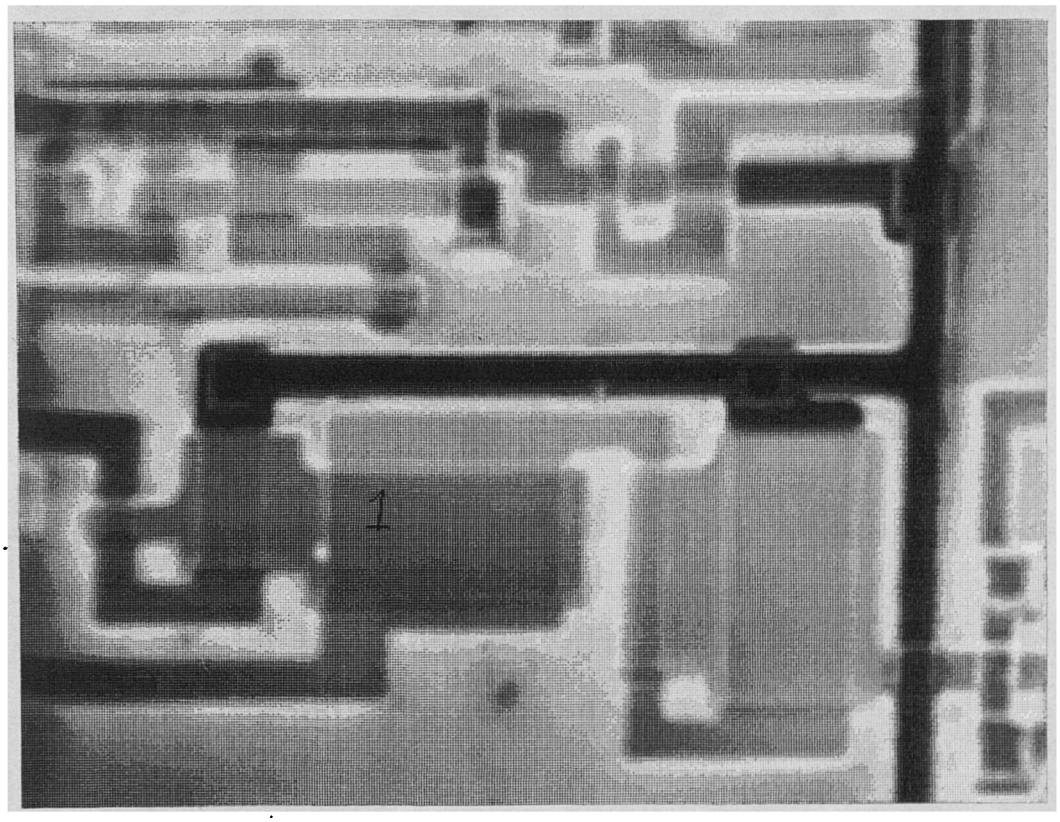
The region indicated by the number **1** on the images is a polysilicon gate which is buried beneath a layer of silicon dioxide. Contrast due to the applied gate voltage (approximately 3.5 V) is clearly visible, demonstrating that the technique can be used to measure voltages of buried nodes. The technique has also been demonstrated by Wolfgang [4.3].

This technique has two disadvantages for the study of floating gates. The first is a general problem which is applicable to any MOS device. The use of high energy primary electrons can cause irradiation damage to MOS structures. The primary result of this is charge trapping in the gate oxide which results in threshold voltage shifts. The second problem is specific to the measurement of floating nodes. The penetration of the primary electron beam to the buried gate will disrupt the amount of charge stored on the gate. As this is the quantity that we are attempting to measure, the effect is highly undesirable and would invalidate any of the results obtained. The use of high energy primary electrons is therefore seen to be unsuitable for the measurement of floating gate voltages.

The second solution to the problem of buried layer measurements relies on the fact that the electric field produced by the gate electrode potential is not confined to the surface of the gate, but extends into and above the oxide. Secondary electrons generated on the top surface of the oxide will therefore be affected by the field due to the gate electrode and hence voltage contrast will be visible. This technique is commonly termed *capacitively coupled voltage contrast*. It is generally explained in terms of a capacitive potential divider (see for example, Richardson [4.11]) and changing voltage levels. It is not however necessary to have a changing electric field for voltage contrast to occur. A static field, such as that produced by the potential on a floating gate will also cause voltage contrast. This technique does not require the primary electrons to penetrate the insulating layer and so the charge on the floating gate should remain unaffected by the measurement. It was therefore decided that this technique should be used to attempt to measure the floating gate potential.



(a)



(b)

Figure 4.5 Voltage contrast on buried nodes: a) no voltage applied; b) 5 V applied to circuit

4.2 The Measurement System

The microscope used in this work was a Cambridge Instruments Stereoscan S180. This microscope can operate at gun voltages down to around 900 V with relatively low beam currents. Both of these conditions are desirable for voltage contrast measurements on IC's which contain large areas of insulating materials. The low gun voltage limits the interaction of the electron beam with the sample. The low beam current reduces the problem of charging of the insulating layers on the IC. At low gun voltages (where the insulating layers of the IC are not conducting) the primary electron current is generally greater than the secondary electron current. Therefore, a net build up of charge occurs which is trapped in the top surface of the dielectric layer. After some time, the build up of charge causes a degradation of contrast in the image and eventually no features can be seen. By using a low beam current, the time taken for the charging to occur is increased and imaging can proceed for tens of minutes before significant degradation of the image occurs. Another solution to the problem of charging is to attempt to make the primary and secondary electron currents equal. This can be achieved by varying the gun voltage and an operating point can be found at which charging effects are minimised.

The S180 microscope has other useful features including a sample stage incorporating a fifty-way electrical connector that allows a forty-pin chip to be mounted and controlled inside the SEM chamber. Slow scan rates could also be selected which is useful for image capture. The images generated by the SEM were captured on an Olivetti Personal Computer (M24) using hardware developed by the author and software initially written by D.J. Brettell [4.12] and later enhanced by the author. Examples of an images captured on the system have already appeared in figures 4.2 and 4.5 The software includes routines for subtraction and addition of images, thresholding and normalisation, filtering and convolution and intensity measurements of points or averaging over rectangular areas.

4.3 Intensity Analysis for Quantitative Voltage Contrast

Before attempts could be made to measure the potentials of floating gates, it was necessary to characterise voltage measurements made using intensity analysis. A contact chain structure was used for this which consisted of 41 metal tracks connecting 40 active regions on an nMOS IC. A section of the structure with 5V applied is shown in figure 4.6. It is clear from this image that the chain gives a set of intensities for known voltages and this can be used for characterisation. To ensure that the voltage variation along the chain was uniform, the metal segments (each of which is 12 x 24 μm) were mechanically probed. The variation was found to be uniform with a voltage step of 125 mV ± 4 mV. Voltage contrast images of the chain were then recorded and the intensity measurements converted to voltage values by means of the characterisation process now described.

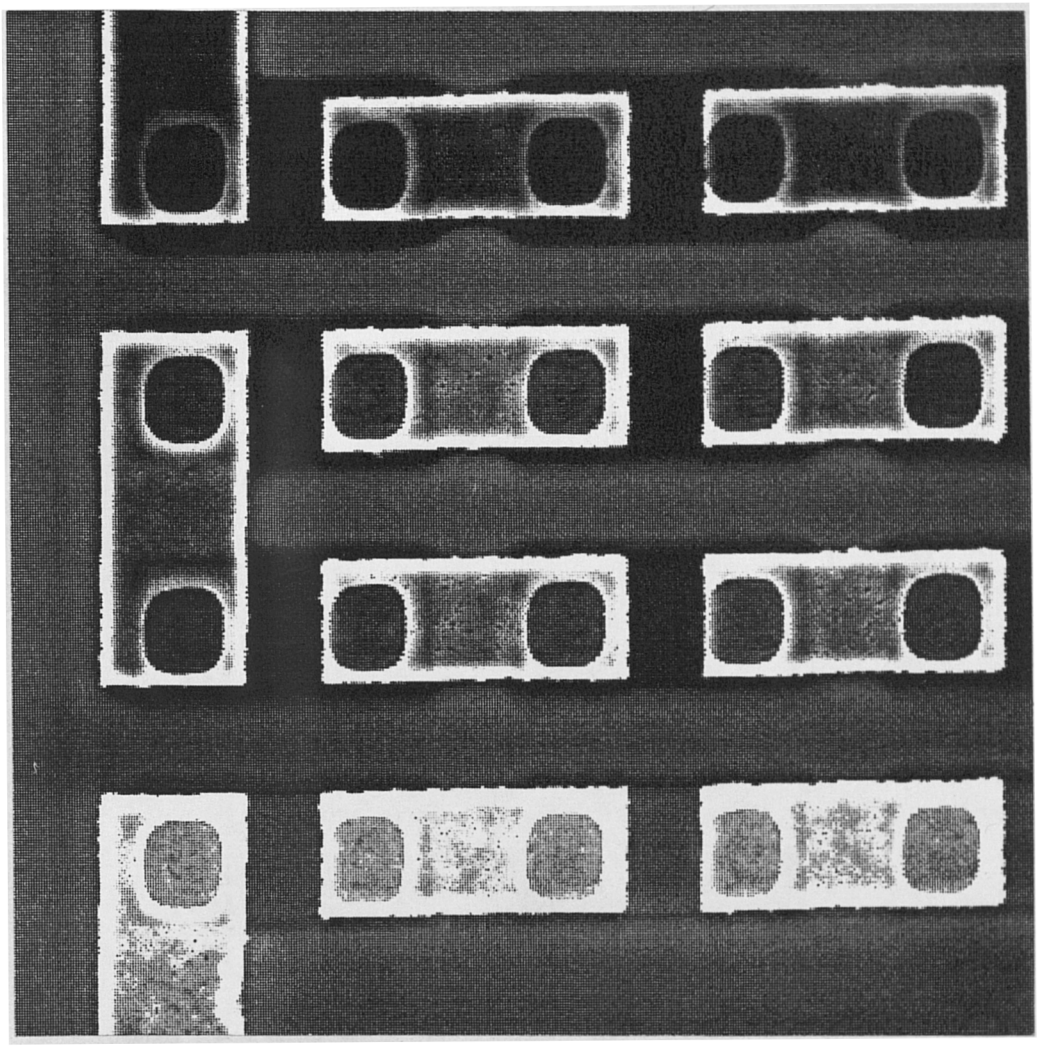


Figure 4.6 Voltage contrast on the contact chain used for characterisation

4.3.1 Voltage Contrast Intensity Measurements for the Contact Chain

To perform intensity analysis, an SEM image is digitised and captured by the computer system. The 400 x 400 pixel image is digitised to 8 bits of resolution and is captured in 4 seconds. The software that has been developed allows the intensity of individual pixels to be measured, although electrical noise in the signal from the SEM and noise from the capture system results in a distribution of intensity values for a region which is nominally at the same intensity.

There are two sources of noise in the recorded signal. The signal produced by the SEM collector and photomultiplier is inherently noisy due to the low number of secondary electrons being generated. This situation is made worse by the requirements of low gun voltage and low beam current. In the image capture system, a total of 160000 pixels are digitised in four seconds giving a sample period per pixel of approximately 25 μS. This is rather a short period for the integration of the SEM signal which is performed by the capture system and consequently, the signal contains a significant amount of noise. The computer used for the image capture also acts as a source of noise primarily from the power supplies and reference voltage used for the analogue to digital converter. While attempts were made to reduce the amount of noise in the captured image, a significant noise signal remained. The error on the measured intensity was typically ± 10 units. The maximum intensity value was 256 units.

The computer system can be used to help to overcome the noise problem by allowing averaging of intensities over a large number of pixels. Therefore, in addition to the point measurement facility, the software allows the average intensity over an area defined by the user to be measured. By averaging over typically 500 pixels, the error can be reduced to less than ± 1 unit. The standard deviation over the averaged region is typically between 10 and 20 units. Further reduction in the noise can be achieved by the application of a median filter which has been included in the software package. This is particularly good at removing spot noise and when applied it reduces the observed error further and the standard deviation drops to between 5 and 10 units typically.

One of the consequences of the noise in the system and the need to perform averaging is that it is necessary to ensure that the features to be measured occupy a significant number of pixels. Consequently, relatively high magnification is used for the measurements described.

Images of the contact chain were captured at gun voltages of 1.2 kV and 15 kV. Between these two voltages significant and rapid charging of the sample occurred. At around 1 kV the system would seem to be near to the balance point at which the numbers of primary and secondary electrons are equal resulting in very little charging of the interlayer dielectric. Unfortunately, at this low gun voltage, the resolution of the microscope was rather poor.

Resolution in an SEM can be improved by operating at higher gun voltages but this moves the system away from the charge balance point. This was observed for gun voltages in the range 1.5 kV to around 8 kV. Above 8 kV the rate of charging was again reduced. It is assumed that this was due to penetration of the insulating layers by the primary electron beam and the resulting conduction as described earlier. At these high energies, significant irradiation damage would be done to MOS devices. The contact chain consisted simply of active regions and metal tracks and consequently would not suffer from irradiation problems.

Figure 4.7 shows a graph of image intensity along the contact chain for an image captured at 1.2 kV. Also shown on the graph is the ideal intensity characteristic which is derived by calculating the integral of the electron energy spectrum equation as given in equation 4.2. For each segment in the chain, the integration is performed from the track segment voltage to a maximum energy above which the number of electron is negligible. The track segment voltages were obtained by mechanical probing.

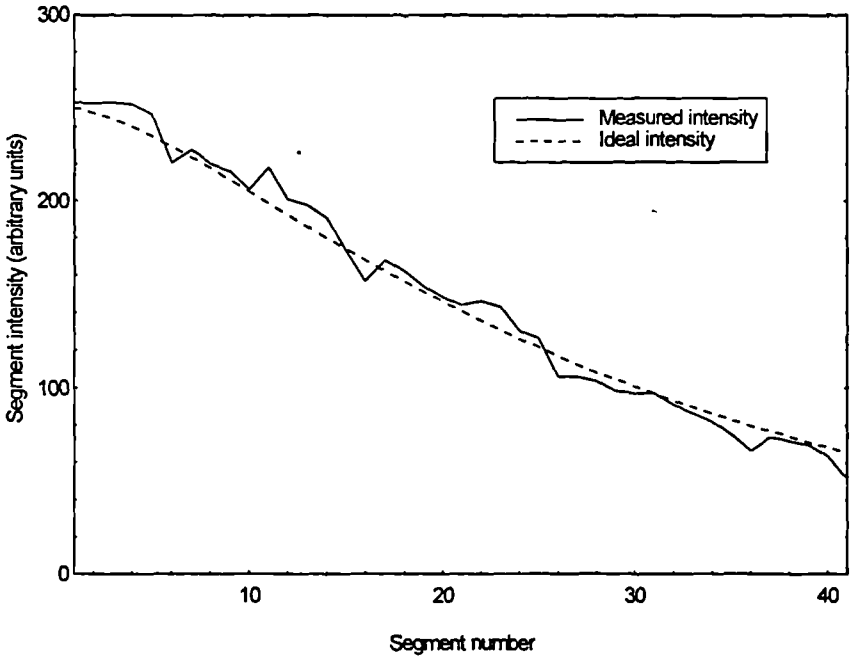


Figure 4.7 Graphs of measured and expected image intensity along the contact chain

The ideal characterisation curve is scaled so that the maximum and minimum intensities correspond to those measured for the 0V and 5V tracks on the SEM image. This is necessary (and reasonable) as, for this measurement technique, we are concerned with the change in the track intensity from a measured reference e.g. 0V or 5V. However, choosing the 0V and 5V tracks for the scaling does not produce the best fit of the measured data to the ideal curve because any error in these two values will shift the entire curve away from the correct value. The best fit was obtained by varying the maximum and minimum intensities over a small range and calculating the standard error for the data sets. By doing this, the error was reduced from 9.05 to 7.77. This highlights a potential problem with the intensity

measurement technique in that it relies on references which may contain an error and this will limit the overall resolution of the measurement technique.

The measured intensities can be seen to follow the ideal characterisation curve quite well although there are some significant anomalies. Several of the points have a large deviation from the expected values particularly at points corresponding to the vertical end segments and to segments in column five of the structure which has a topographic anomaly. The error in the measured values resulting from electrical noise has been reduced to ± 1 unit by averaging over around 200 pixels for each measurement. A further source of error must therefore exist to account for the irregular variation of intensity along the chain.

A possible source of this error is the material and topographic contrast which is present in the image. It should be possible to remove this by subtraction of an image captured with zero voltage applied from the composite image. Figure 4.8 shows the variation of the image intensity along the chain with no voltage applied. Some of the anomalies identified in the voltage contrast image are also apparent in the no voltage image.

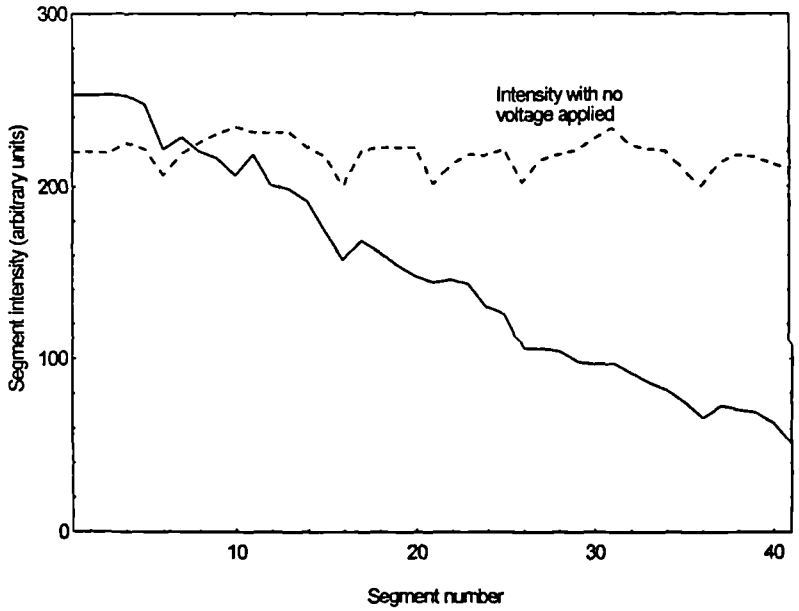


Figure 4.8 Image intensity along the contact chain with and without voltage applied.

It is clear from the graphs that simple subtraction of the zero voltage image from the voltage contrast image would not produce useful results, as it would lead to negative intensity values. Furthermore, scaling of the zero voltage image down to the voltage contrast minimum intensity before subtraction does not remove the topographic (and material) contrast correctly. This is because the image contrast is proportional to the image brightness. Consequently, the topographic contrast in the composite image in bright regions is higher than the contrast in dark regions. Therefore, to correctly subtract the topographic contrast, the intensity of the composite image must be taken into account and the zero voltage image

of the composite image in the region being considered before subtraction. This process must be performed independently for each track segment being considered. The subtraction process is summarised as follows:

i) Find the intensity of the track to be measured from the composite image.

ii) Find the deviation from the mean intensity for this track in the zero voltage image. This gives a measure of the topographic contrast for the track segment.

iii) Scale the deviation by the ratio of the composite and mean zero voltage intensities and subtract from the composite image.

This procedure can be expressed in the following formula:

$$I_{vo} = I_c - \left(\bar{I}_{nv} - I_{nv}\right)\frac{I_c}{\bar{I}_{nv}} \tag{4.3}$$

Where I_c, I_{nv} and I_{vo} are the track segment intensities for the composite, zero voltage and resulting voltage only images respectively and $\bar{I}_{nv}$ is the mean track segment intensity for the zero voltage image. The above process has been applied to the composite image data and the resulting intensity curve is shown in figure 4.9. The process has reduced the standard error for the data from 7.77 to 6.57, a rather small but not insignificant improvement.

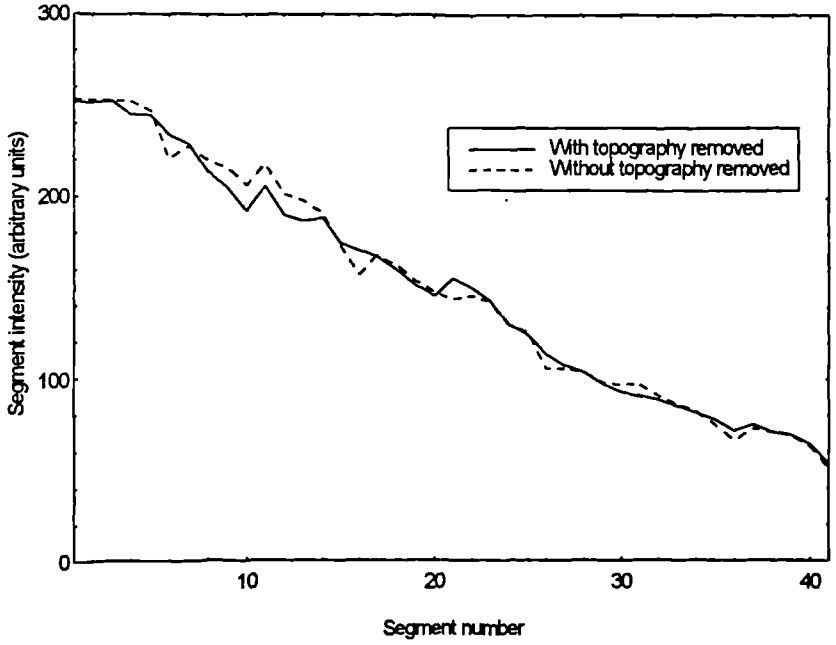


Figure 4.9 Intensity curve with and without topography removed

The intensity curve still contains significant deviations from the ideal characteristic that cannot be attributed to electrical noise, measurement error or topographic contrast. It is felt that the most likely source of these errors is non-uniform charging of the interlayer dielectric

dielectric and a resulting *local field effect* problem. The local field effect has been described by many workers and is recognised as a serious problem in all forms of voltage contrast measurement systems. The problem occurs because the secondary electrons emitted from the track being examined are influenced by the potentials of local tracks and accumulated charge on the dielectric surface in the same manner (although not to the same extent) as the potential of the measured track. The effect is illustrated in figure 4.10 which shows the calculated secondary electron paths at the surface of an IC in the SEM with neighbouring tracks at potentials of 0V and 5V.

The traces shown in figure 4.10 were produced by a simulation program written by the author to investigate local field effects in voltage contrast imaging. The program allows the calculation of secondary electron paths in an electric field which is read in from a file. A program written by R.W. Kelsall was used to calculate the electric field at the IC surface for a range of electrode configurations and potentials. Electron parameters (energy, initial trajectory angle and position on track) are varied over a predetermined range and the resulting electron paths were calculated. This allows the number of electrons reaching the SEM collector to be estimated and thus the change of intensity for various track voltage configurations can be found.

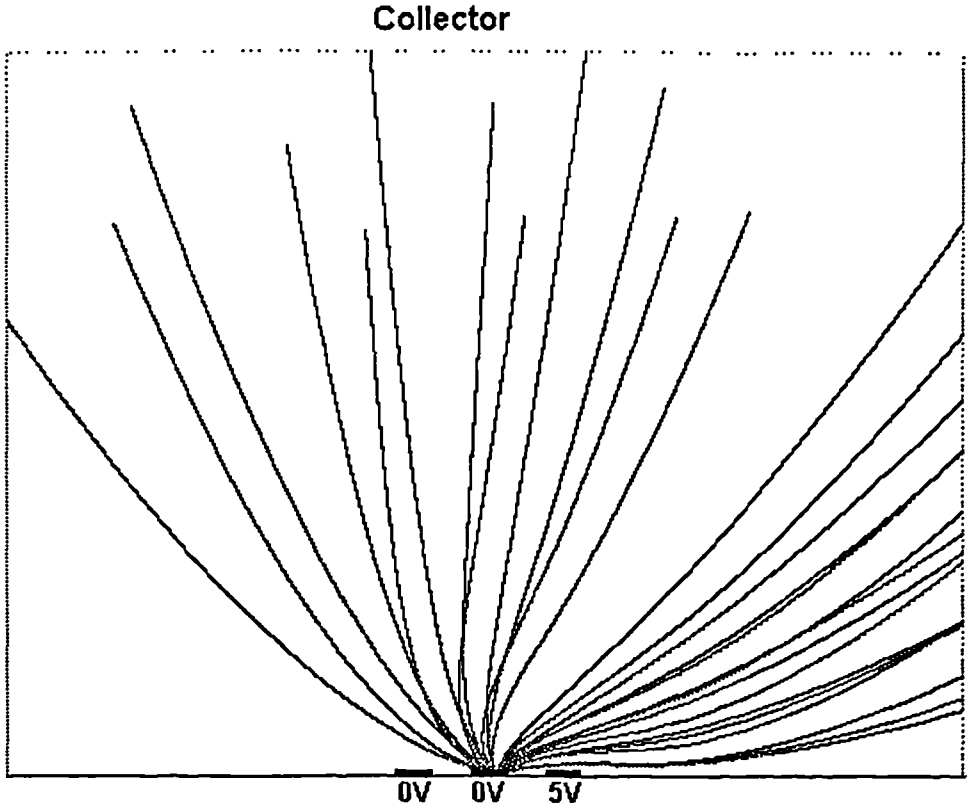


Figure 4.10 Secondary electron paths at the surface of a IC

The example given shows cross-sections of three metal tracks on the surface of an IC. The tracks are 2 μm wide with a separation of 2 μm. It is clear from this example that setting the right hand track to 5V has a significant effect on the secondary electrons emitted from the centre track (which is at 0V). Simulations indicated that the intensity of this track would be reduced by approximately 50% when 5 V is applied to the neighbouring track and thus an error in the measured voltage would result. This effect is not confined to intensity measurements as the secondary electron energy spectrum for the centre track is also affected, resulting in errors in measurements made by energy analysis.

The problem with the intensity measurements seemed to result from non-uniform charging of the surrounding oxide layer. One possible solution to this was to use a higher gun voltage so that charging of the oxide layer did not occur. The measurements were therefore repeated using a gun voltage of 15 kV. This also had the benefit of improving the image resolution significantly which should help to reduce the errors in the image. The data from the 15 kV images was obtained from four pairs of images (with and without voltage applied) at high magnification, each pair covering a section of the chain. The intensity measurements are shown in figure 4.11 with a best fit ideal characteristic superimposed.

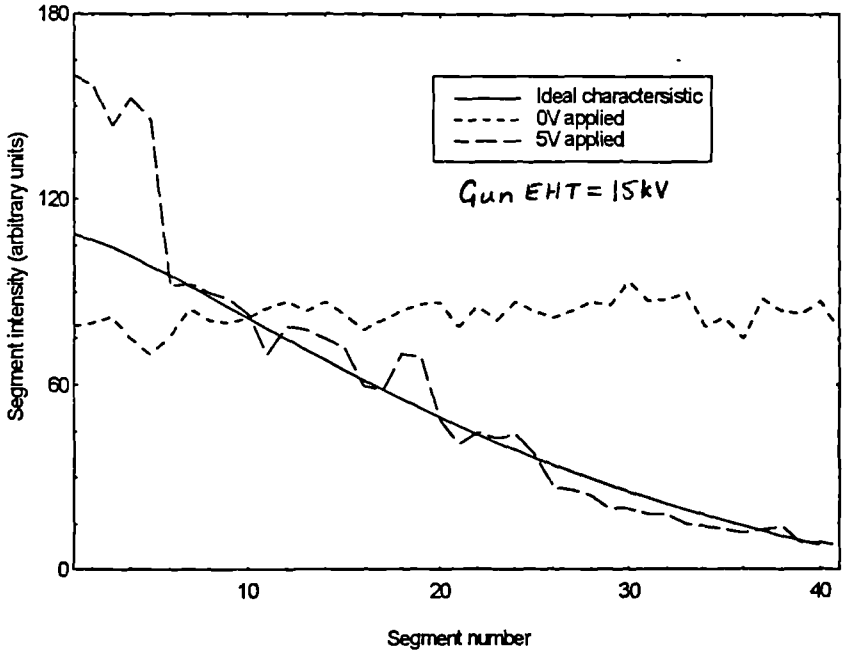


Figure 4.11 Intensity measurements along the contact chain with ideal characteristic superimposed

The image does not show any improvement in the random fluctuations of intensity along the chain. Furthermore, there is a large anomaly at the low voltage end of the track where the intensity drops very suddenly between segments. This was thought to be a high resistance contact causing a non-uniform variation of voltage along the chain. However, subsequent probing of the tracks indicated that this was not the case. Careful examination of the set of images indicated that one set of the four had a slightly higher overall intensity than

the other three. This was probably due to instability in the electron optics of the microscope and it accounts, to some extent, for the data points that have large errors although it does not reduce the low voltage anomaly. Subtraction of the topographic image using the procedure described previously did not improve the error overall and, in fact, it increased the error at the low voltage end of the chain.

Increasing the gun voltage did not appear to have the desired affect of reducing the error in measured intensity. It would seem that the local field effects have not been eliminated by this change and there is no advantage in using a high gun voltage other than to improve resolution generally.

4.3.2 Obtaining track voltages from intensity measurements

The voltage of a node on the IC can be obtained by comparison of the measured intensity with the ideal intensity curve and reading off the equivalent voltage. This process has been used to produce the graph shown in figure 4.12 which is a plot of segment voltage against position obtained from the 1.2 kV image. The value obtained by mechanical probing is shown for comparison. The graph shows good agreement between the two sets of measurements with a standard error of 6.32 and a maximum error of 0.55 V. It is felt that the main component of these errors results from the local field effect.

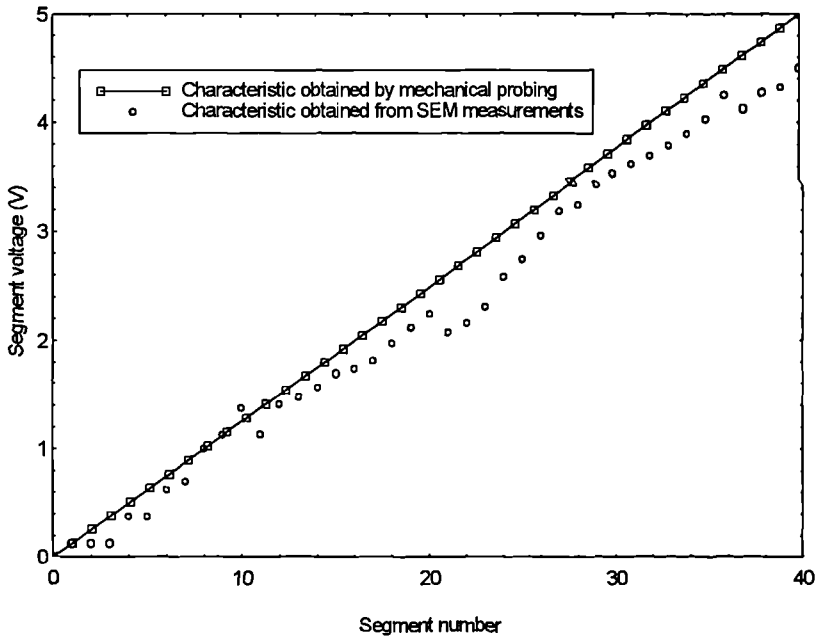


Figure 4.12 Voltage variation along the contact chain

4.3.3 Conclusions on the use of Intensity Analysis for Quantitative Voltage Contrast

The results presented in figure 4.12 show that voltage measurements can be obtained with a typical accuracy of ± 200 mV on unpassivated metal tracks using intensity analysis. The worst case error is 550 mV. This is quite adequate for detailed analysis in many circumstances and provides considerably more information than simple qualitative techniques which can only give information on the logic state of nodes. However, some caution must be expressed. As has been shown, care must be taken in deriving the ideal intensity characteristic, since the use of two reference points for its normalisation can result in error if they contain error components. It would generally be desirable to have several reference points to obtain the best fit characteristic.

The effect of local fields is significant and, in structures containing smaller densely packed features, this could introduce large errors. One solution to this problem suggested by Dinnis (and used in at least one commercial system) is to immerse the IC in the strong magnetic field produced by the SEM pole pieces. This causes the secondary electrons to spiral away from the sample in a direction normal to the surface helping to eliminate the interaction with other electric fields. A strong electrostatic extraction field can also be used for this purpose. Problems can also occur if there is instability in the electron optics, although this should not be a significant problem if a single image is used for the measurements of a node voltage as would normally be the case.

4.4 Voltage Measurements on Floating Gate Devices.

Having established that intensity based voltage contrast measurements can produce reasonably accurate indications of voltages on metal tracks it is necessary to extend the technique to the measurement of buried layers and hence floating gates. The principles behind measurements on buried nodes have already been described. As it was necessary not to disrupt the charge stored on that gate the *capacitively coupled* technique had to be used for the floating gate measurements. It was anticipated that further characterisation would be needed to account for the presence of the oxide layer as this was known to cause an apparent drop in the measured potential as illustrated by Wolfgang [4.3]. The drop in potential is confirmed by Ranasingh [4.13] but different values are given. This suggests that the voltage drop is dependent on device parameters, such as dielectric thickness for example, or on measurement system parameters such as the SEM gun voltage.

For characterisation, the gate voltage on a normal transistor was to be measured over the range 0 to 5V with 0 and 5V references on nearby gate electrodes. The transistors to be analysed had a channel length of 3 μm and width of 4 μm. The devices were examined with the SEM gun voltage set to 1.2 kV. Monitoring of the drain-source current while the device was in the SEM gave an indication of the effect that the electron beam was having on the

sample. No change was noticed in the drain-source current of the test device when the electron beam was switched on indicating that the primary electrons were not affecting the transistor operation.

Unfortunately, the image obtained at 1.2 kV was of very poor quality. It was to be expected that the polysilicon gate would not be visible on the image as the primary electrons were not penetrating the interlayer dielectric that covered the gate. The surface feature created by the presence of the gate should have been visible but due to the poor resolution of the image obtained, it could not be seen. Application of the gate voltage did not appear to affect the intensity of the image in the region of the gate. This lack of contrast is almost certainly due to the poor quality of the image. Attempts were made to improve the image by adjusting various SEM parameters but the image remained very poor.

Increasing the gun voltage resulted in charging of the oxide layer and subsequent loss of the image. At 15 kV the charging problem had ceased and a good quality image was obtained. The transistor gate could be clearly identified, and changes in the gate potential were apparent as contrast changes in the image. However, at this potential, the primary electrons would be interacting with the transistor, injecting charge into the gate oxide and channel regions. Consequently the SEM could not be used at this high gun voltage to investigate the floating gate potential without affecting the quantity to be measured.

The investigation of voltage contrast techniques for measuring the floating gate potential was stopped at this point. It was clear that the limitations of the microscope used would not allow the accurate measurements hoped for. However, the work has established that direct intensity analysis can be used to measure track voltages in the SEM. The accuracy of the technique seems to be limited by local field effects rather than topographic and material contrast as is generally assumed. This is a problem common with energy analysis techniques. A further limitation on the accuracy that can be obtained is due to errors in the readings used for the calculation of the characterisation curve. It is strongly recommended that several points are used for the normalisation of the characteristic to reduce this error.

The author has no personal experience of the use of energy analysis systems for voltage contrast but it is apparent from the literature that the technique would seem to be preferable to intensity analysis. However, the results of this investigation suggest that the technique can provide useful quantitative results when used with care.

4.5 Chapter Summary

The use of intensity analysis for quantitative voltage measurements on ICs has been demonstrated. The measured voltage-intensity characteristic was shown to be a reasonable fit to the ideal curve. An accuracy of approximately ± 200 mV can be obtained for voltage

measurements on unpassivated tracks. The technique is limited in its resolution primarily by local field effects from neighbouring conductors. This is also a problem with the more accepted technique used for quantitative voltage measurements in the SEM, i.e. secondary electron energy analysis. Despite these problems, the technique is particularly attractive for analysis of parametric failure modes of ICs as these often lead to intermediate voltages on gate outputs which cannot be analysed with simple qualitative methods.

Voltage contrast microscopy in the SEM did not provide a direct measure of floating gate potential as had been hoped. The fault does not lie in the technique itself, but simply in the limitations of the equipment used in this study. As alternative equipment was not available, it was necessary to use more conventional techniques for the analysis of the floating gate transistors and circuits. The resulting electrical analysis of the test circuits is described in Chapter 5.

Chapter 5
Experimental Analysis of Floating Gate Transistors

In previous chapters, the floating gate has been identified as an important form of fault which, it has been suggested, may not be revealed by standard testing techniques. Consequently, a set of test structures have been produced to allow the operation of the device to be investigated. In this chapter, we describe electrical measurements which reveal the effect of the fault on a range of devices and circuits. Measurements on individual transistors, inverters and simple logic gates are described and the results are analysed.

5.1 Analysis of Floating Gate Transistors

As described in Chapter3, several sets of logic gates containing a range of faults have been produced over four years. The final test chips produced included a set of twelve *n*- and twelve *p*-channel MOS transistors with floating gates. The purpose of the transistors was to provide data on the operation of floating gate transistors. This information was required to develop a model of the device which would, in turn, be used for the modelling of CMOS circuits with floating gate faults. In particular, three device characteristics were to be investigated: i) the floating gate potential with zero bias applied to the device, V_{FG0}; ii) the variation of V_{FG} with drain-source voltage; iii) the I_{DS} - V_{DS} characteristic for the floating gate transistor. The effect of device dimensions on the characteristics was to be investigated and so various length and width transistors were included. The dimensions of the devices are given in Table 5.1.

Transistors **a** - **d** cover a commonly used range of dimensions for a 3 μm technology and would provide data on the operation of standard devices. Transistors **e** - **h** were designed to investigate the significance of the ratio of the gate-drain overlap length to the drawn channel length by varying the channel length. Transistors **i** - **l** were intended to show the effect of capacitive coupling of signal lines to floating gates. Consequently, the floating gates were extended underneath the drain metal track by the lengths shown in the table.

Device	W x L (μm)	Comments	Device	W x L (μm)	Comments
a	4 x 3	Standard device	g	4 x 9	Long channel devices
b	8 x 3	"	h	4 x 12	" metal
c	12 x 3	"	i	4 x 3	Gate extended under drain by: 35 μm
d	16 x 3	"	j	4 x 3	70 μm
e	4 x 3	Long channel devices	k	4 x 3	105 μm
f	4 x 6	"	l	4 x 3	140 μm

Table 5.1 Floating gate test transistor dimensions

A layout diagram of the n-channel test devices is shown in figure 5.1. (The layout scheme used for the p-channel devices was very similar). The devices have common substrate and well connections and all sources are connected together. There is a separate probe pad for each drain terminal.

In addition to the faulty devices, a range of fault free transistors of various channel lengths and widths were fabricated. Measurements on these indicated that the devices operated as expected over the range of dimensions produced. The drain-source current varied as the square of V_{GS} as expected and the threshold voltage and transconductance of the devices were within the expected tolerances. A number of other parameters used in the SPICE models for these devices were checked and found to be as predicted by the process specification. The I_{DS} - V_{DS} characteristics for an *n*-channel transistor for a range of V_{GS} values are given in figure 5.4.

5.1.1 I_{DS} - V_{DS} Characteristics for Floating Gate Transistors

As the gate terminal is not accessible, the most useful measurement that can be made is the I_{DS} - V_{DS} characteristic. This was obtained for all devices using the circuit shown in figure 5.2. Figure 5.3 shows the I_{DS} - V_{DS} characteristics for the standard devices, **a - d**, as described in table 5.1. Measurements from four nominally identical samples are shown for each device. The transistors were all fabricated on the same process run but are on different die.

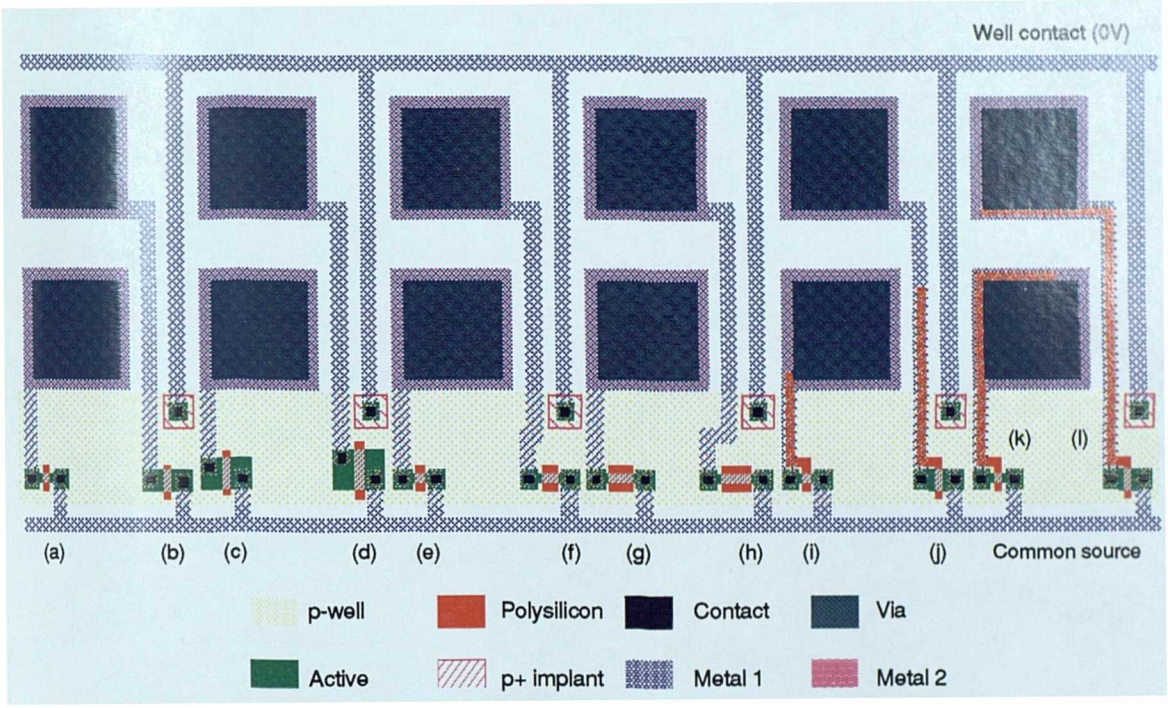


Figure 5.1 Floating gate n-channel MOS transistor test structure

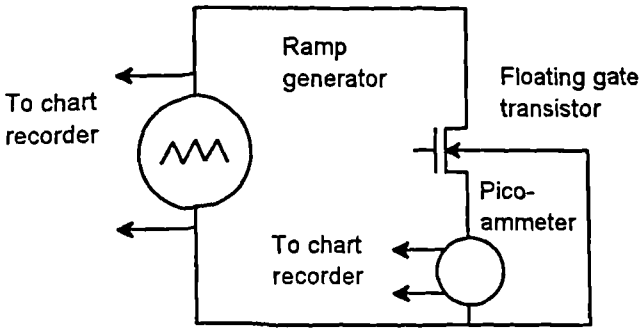
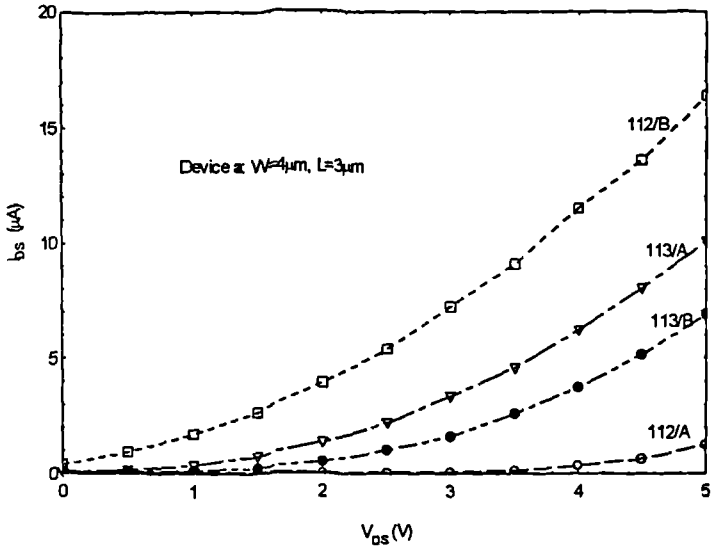
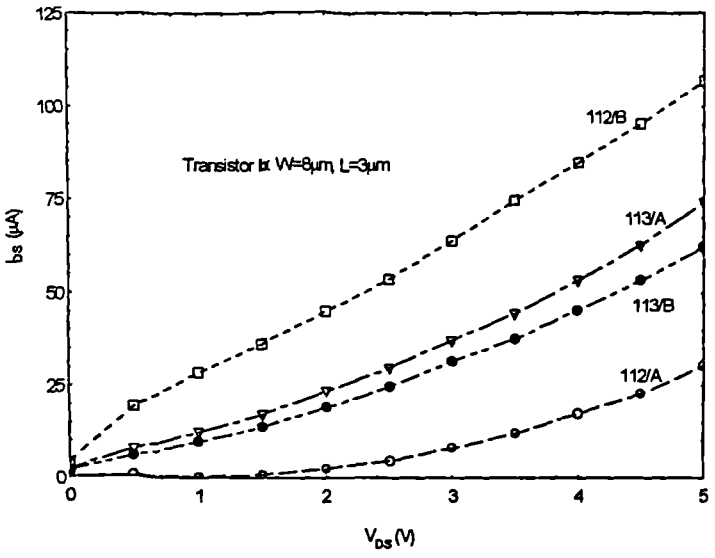


Figure 5.2 Test circuit used for the I_{DS} - V_{DS} characterisation of floating gate *n*-channel transistors



Device (a)



Device (b)

Figure 5.3(a) I_{DS} - V_{DS} for transistors a-b on four samples (Note different current scales)

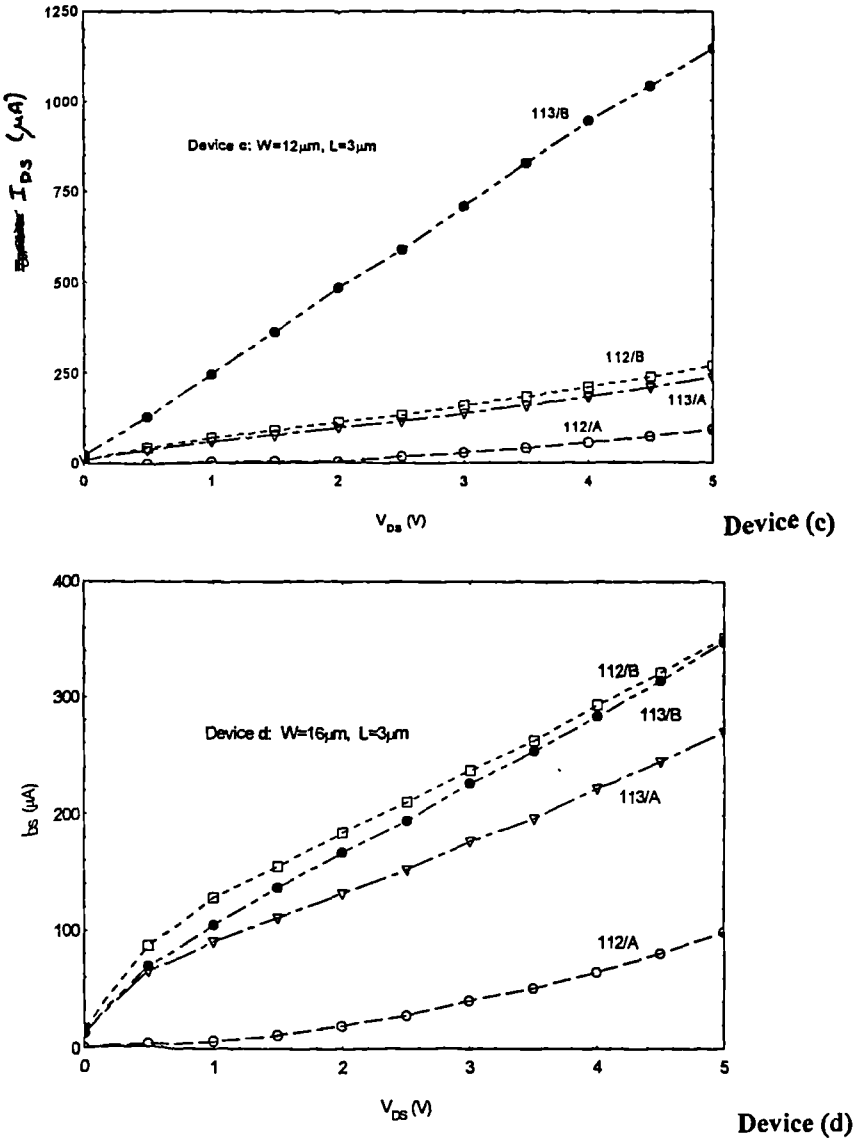


Figure 5.3 (b) I_{DS} - V_{DS} for transistors c - d on four samples (Note different current scales)

The characteristics are seen to vary considerably, even for devices of the same size. There are two quite distinct shapes to the curves. One of these looks similar to the standard drain-source current curve with two regions of different slope, the change occurring at low V_{DS}. Device **d** on sample 112/B (referred to as d-112/B) is of this type. The curves of other devices look similar to the I_{DS} - V_{GS} curve of a fault-free transistor with a constantly increasing gradient. Sample d-112/A is of this type. The other curves lie in between these two extremes. The characteristics are considerably different to that for a fault free device.

The most reasonable explanation for this range of characteristics is a variation in the gate potential between the devices. The theory developed in Chapter 6 shows that the I_{DS} - V_{DS} curve for a floating gate device is primarily determined by two parameters: i) the initial floating gate voltage, V_{FG0}; and ii) the capacitive coupling between the drain and source terminal and the gate. The second of these parameters explains why the shape of the I_{DS} curve is not the same as for fault free devices. The gate voltage actually varies with V_{DS} and

curves shown in figure 5.3 represent a drain-source current characteristic for increasing V_{DS} *and* V_{GS} rather than the fixed V_{GS} normally shown. The initial floating gate potential determines the conduction level of the device and effectively provides an additive offset to the variation of V_{GS} with V_{DS}.

5.1.2 The Floating Gate Potential and the V_{FG} - V_{DS} Relationship

The aim of the measurements presented was to determine the initial floating gate voltage, V_{FG0}, and the variation of V_{FG} with V_{DS}. These can be extracted from the graphs in figure 5.3 by the following method. The I_{DS} characteristic for a floating gate device is plotted with a set of I_{DS} curves for an identical fault free device obtained at various gate voltages. The intersection of the floating gate curve with each fault free characteristic gives the drain voltage required on the floating gate device to achieve the appropriate gate voltage. The set of V_{GS} and corresponding V_{DS} values give the V_{FG} - V_{DS} relationship for the floating gate device. This procedure is illustrated in figure 5.4.

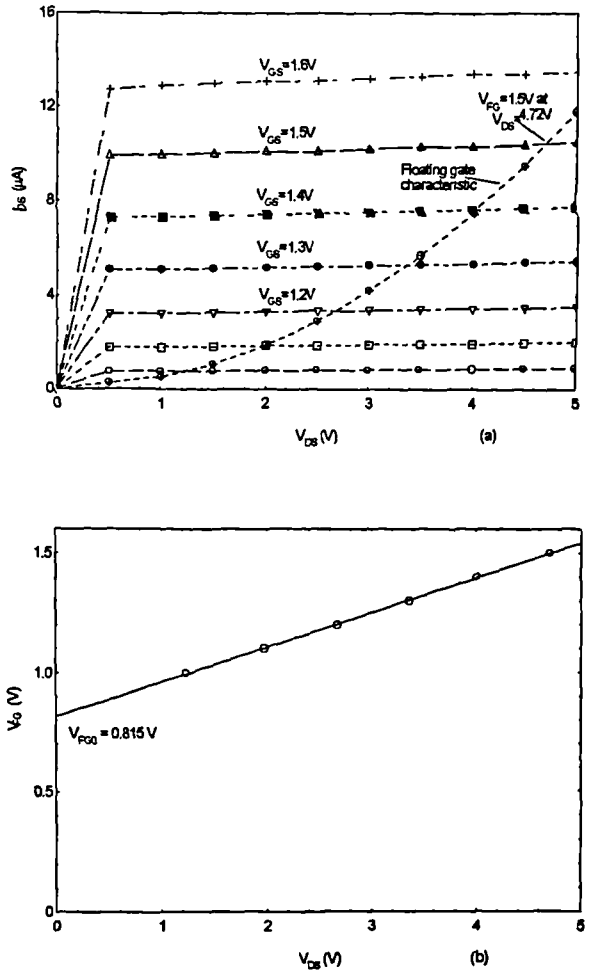
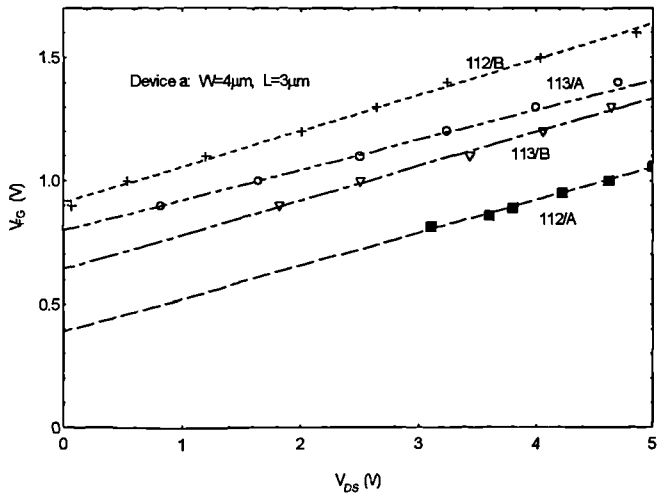
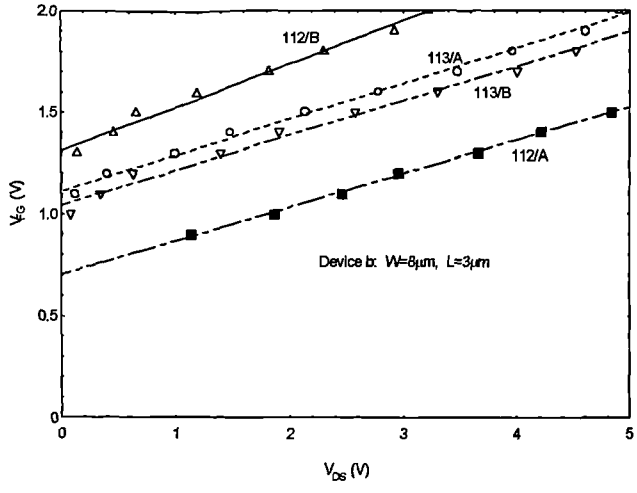


Figure 5.4 a) I_{DS} - V_{DS} curves for the floating gate and fault free devices; b) variation of V_{FG} with V_{DS} obtained from the intersection of faulty and fault free device curves in (a)

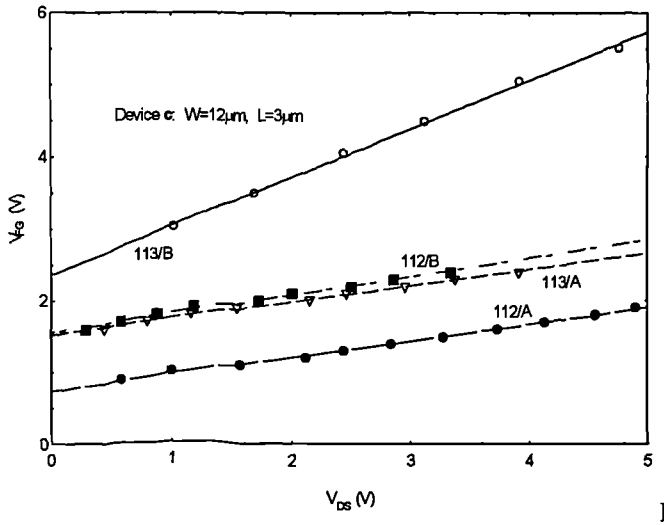
The V_{FG} - V_{DS} curve obtained can be extrapolated to give a measure of the initial floating gate voltage, V_{FG0}. This procedure has been used to obtain V_{FG} - V_{DS} curves for all sixteen devices in the set being considered. The curves are plotted in figure 5.5.



Device (a)

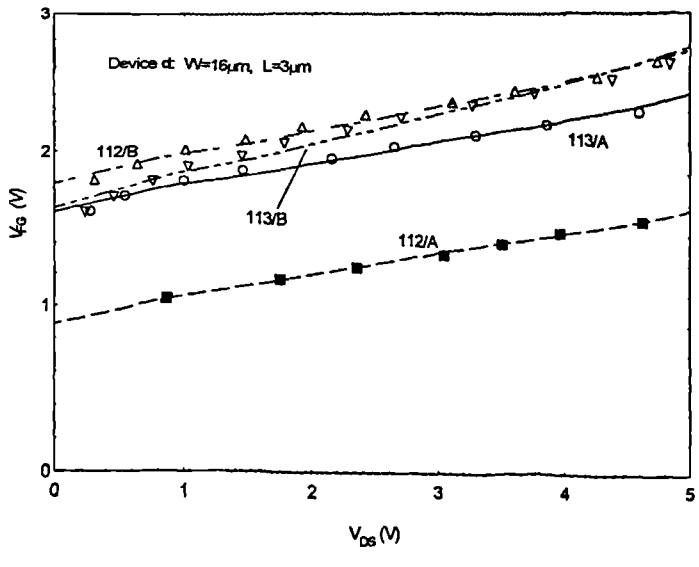


Device (b)



Device (c)

Figure 5.5 (a) V_{FG} -V_{DS} curves for four samples of *n*-channel devices a-c



Device (d)

Figure 5.5 (b) V_{FG}-V_{DS} curves for four samples of *n*-channel device d

The graphs reveal a range of values for both V_{FG0} and the amount of coupling of the drain potential to the gate. The initial floating gate voltage varies from 0.38 V to 2.30 V with a roughly uniform distribution between 0.6 and 1.7 V. This indicates that, although all devices have experienced identical process sequences, there is a range of values of the residual charge on the floating gates at the time at which measurements were taken. However, although the four samples are from the same batch of wafers, they may have been diced from different wafers.

It is therefore difficult to explain this variation in V_{FG0}. The chips have all experienced identical storage conditions and have not been exposed to extreme temperatures or other parametric variations (such as humidity). The circuits have not been powered up except for this testing and some very limited logic testing of other circuits present. As the precise source of the residual charge is not known, it is not very useful to speculate further at this point about variations in this parameter. It is, however, important to note that the range of V_{FG0} spans the threshold voltage for the devices (around 0.9 V) resulting in considerable variation in the transistor conductivities.

A slight trend does appear, however, if we plot V_{FG0} against device width as shown in figure 5.6. It would seem that V_{FG0} is generally higher for wider devices. This trend is difficult to explain for the reasons described above. It is to be expected that the net charge on a wide gate will be higher than that on a narrow gate because of the larger surface area. However, the resulting gate voltage should be the same for each device as the gate capacitance scales in the same proportion as the net charge. Therefore, an argument based on the increased surface area accumulating more charge during implantations cannot be used to explain the apparent increase. This is considered further in section 5.1.4.

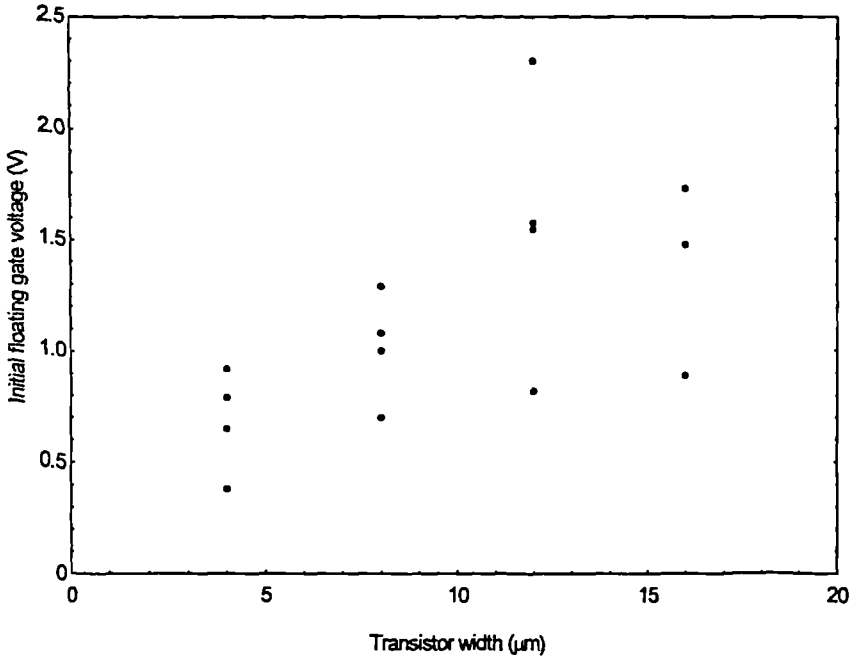


Figure 5.6 Variation of initial floating gate voltage with transistor width

Returning to figure 5.5, the relationship between V_{FG} and V_{DS} appears to be quite linear and straight lines provide a good fit to most of the data sets. The gradients of these lines are not all the same and the initial floating gate voltage is seen to vary. From these results we can see that, generally, the gate potential can be expressed as equation 5.1.

$$V_{FG} = V_{FG_0} + \alpha V_{DS} \tag{5.1}$$

where V_{FG0} is the gate voltage with zero drain or source bias and α is the slope of the V_{FG} - V_{DS} curve.

The proportionality constant, α, varies between 0.124 and 0.253 with one extreme point at 0.703. This latter point is considered to be due to a fault in the transistor which is in addition to the intended floating gate fault. This proposition is supported by the theory for the coupling of the drain voltage to the gate, which is developed in the next chapter. In this, it is shown that the maximum value for α is 0.5, indicating that the value of 0.703 is anomalous. The fault most likely to cause this effect is a gate oxide short. In such a fault, the channel (and, therefore, the drain) is resistively coupled to the gate. This could account for the large feedback that is observed from the drain to the gate.

While several of the data sets presented in figure 5.5 are a good fit to straight lines, some show distinct non-linearities. Most of the data sets are an equally good fit to a simple curve which has a monotonically increasing gradient. The gradient appears to vary by quite a large factor for some devices. The gradient of the curve for device **b** (width = 8 μm, length = 3 μm) on sample 112/B for example, varies between 0.3 and 0.16. Other curves have a similar variation. Generally, devices with higher V_{FG0} seem to have larger variations in the

gradient, α. This is confirmed by the graph shown in figure 5.7. In this we plot the average gradient of the V_{FG} - V_{DS} curves against V_{FG0}. A trend is apparent indicating that α increases with V_{FG0}.

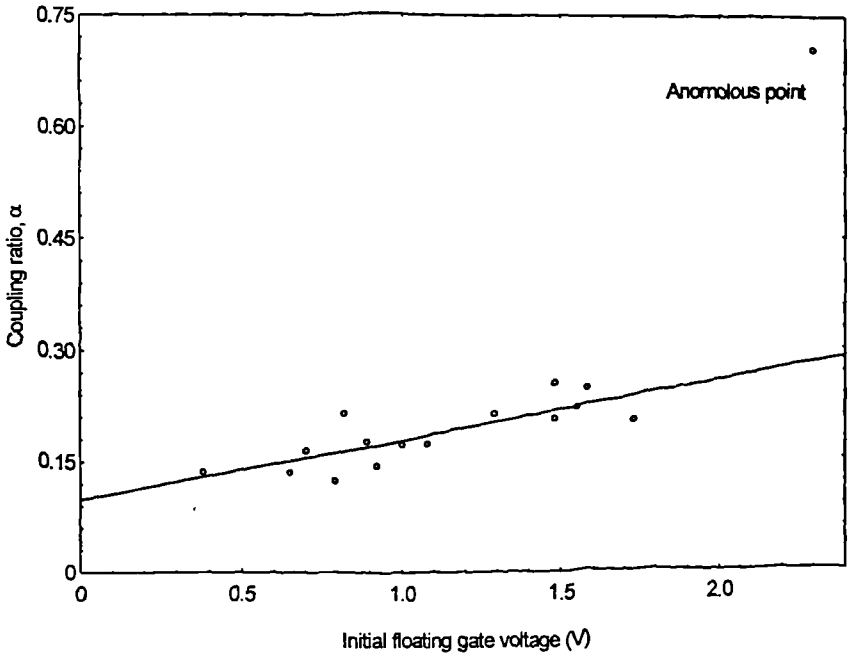


Figure 5.7 Variation of α with V_{FG0}.

Drain-source current characteristics were also recorded for floating gate devices **e - h** which were designed to show the variation of α with the ratio of the gate-drain overlap length to the channel length. This ratio was varied simply by changing the channel length as illustrated in figure 5.1. Devices **e - h** had channel lengths of 3, 6, 9 and 12 μm respectively. As the drain-gate overlap length, ΔL, is nominally fixed, the ratio ΔL / L decreases with increasing channel length.

The floating gate potential for these devices was found to lie in the same range as devices **a - d**, ranging from 0.53 V to 1.65 V with a tendency for longer devices to have higher values of V_{FG0}. This evidence, when combined with the results relating V_{FG0} to device width, suggests that V_{FG0} may be determined by the area of the gate electrode. It would seem that a higher gate area produces a higher floating gate voltage. Some caution must be expressed over this conclusion as initial floating potentials were only obtained for six long channel devices. However, for a total of 22 transistors, the initial floating gate potential seems to increase with gate area. As indicated previously, the relationship between the residual charge and the transistor dimensions is considered further in section 5.1.4.

The gradient of the V_{FG} - V_{DS} curve, for five out of the six devices measured, appeared to be constant. In addition, the proportionality constant, α, decreases with

increasing channel length as shown in figure 5.8. This relationship is explained in detail in Chapter 6.

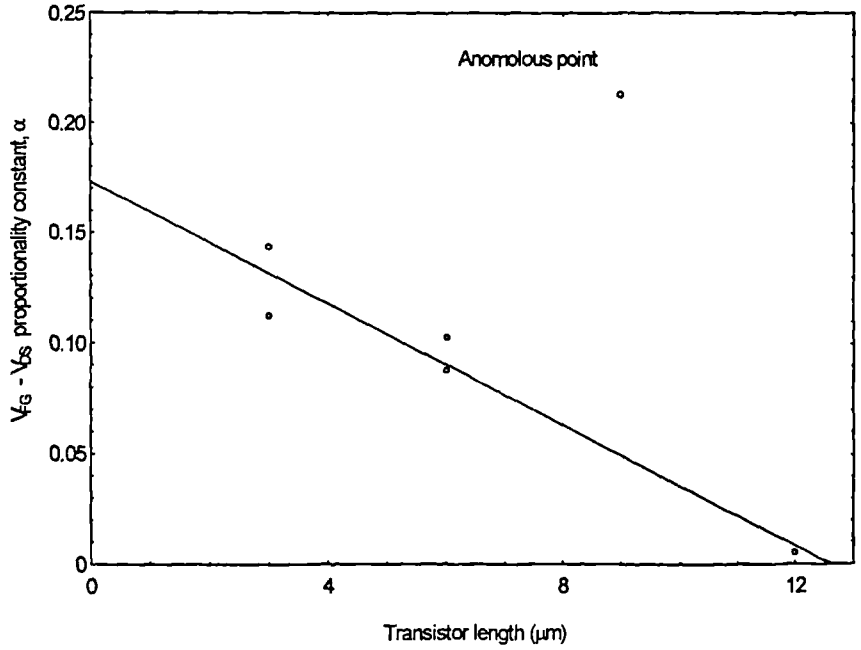


Figure 5.8 Variation of α with channel length

One sample of device **g** (length = 9 μm) in this set had a higher value of α than was expected ranging from 0.58 to 0.1 as the drain potential was increased. By comparison with devices **f** and **h**, the α value for the device should have been between 0.11 and 0.05. For this reason it was felt that this transistor may have a gate oxide short as previously described. While this appears to be the most likely cause of the high value of α, some concern must be expressed over this conclusion. Three gate oxide short defects in a total of 30 devices is a very high rate of occurrence for this failure. It is possible that transistors with floating gate faults are more susceptible to this form of defect than normal devices because very high charge densities may exist on some floating gates. Clearly the charge cannot escape, and so the gate oxide experiences a high field for an extended period, resulting in breakdown and a gate oxide short. This would not happen on a normal device as no residual charge exists on the gate. There may be other reasons for the high values of α but the author cannot suggest any alternatives.

The final set of floating gate transistors to be investigated (devices **i** - **l**), were all minimum size transistors. However, the polysilicon gate was extended beneath the drain metal by a different amount for each of these devices. This allowed the coupling of overlapping signal lines to the gate to be investigated. This increased coupling should give a high value of α for these devices. The devices are illustrated in figure 5.1.

From measurements on two samples, the coupling ratio was seen to increase with the overlap length as shown in figure 5.9. This suggests that floating gates can be significantly influenced by overlapping signal or power lines, provided that the overlap occurs for distances of several tens of microns.

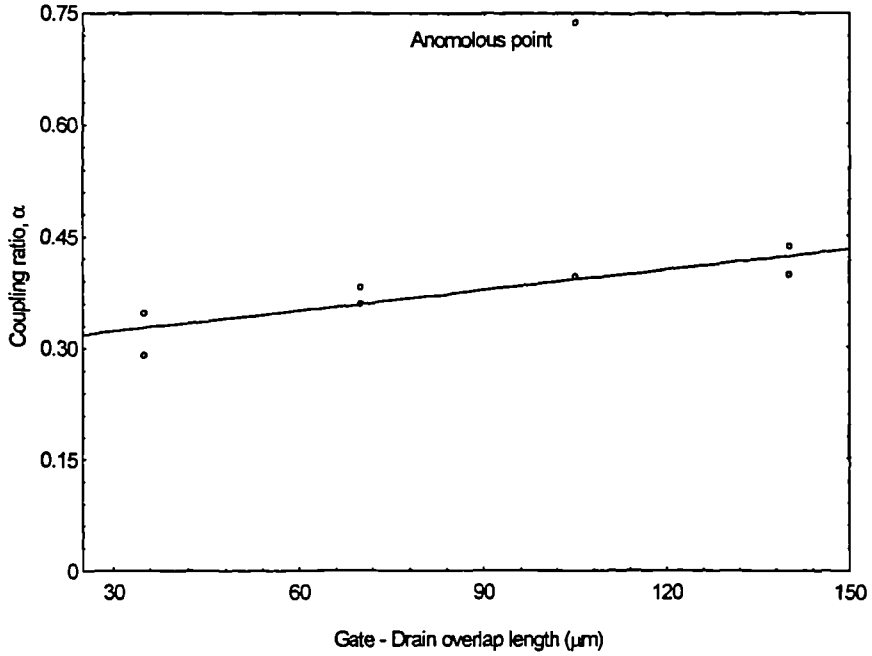


Figure 5.9 Variation of α with gate/drain overlap

The results obtained from the measurements discussed will be used for the evaluation of a theoretical model of the floating gate transistor, which is developed in chapter 5.

5.1.3 Measurements on *p*-channel Floating Gate Transistors

The drain-source current for two sets of twelve *p*-channel transistors was measured using a circuit similar to that used for the *n*-channel devices. The transistor dimensions are listed in table 5.1. For 20 out of the 24 devices probed, the current was too small to be measured. Four of the devices became conducting for $V_{DS} > -4$ V but the currents were very low (all less than 4 μA). The technique previously described was used to obtain the V_{FG} - V_{DS} relationship for the devices which passed a measurable current. The current in one of these devices was too small to provide sufficient points on the graph for extraction of the floating gate potential. The initial floating gate voltages and the α values for these devices are given in table 5.2.

The most important point to note about these devices is that the initial floating gate voltages are all positive as in the *n*-channel case. This explains why most of the devices could not be made to conduct. The floating gate voltage has to change from a positive value to become more negative than the *p*-channel threshold which was around 0.5 V for the

devices measured. The floating gate voltage is seen to change by typically 1 V for a 5 V change in V_{DS}. Consequently, the devices that have an initial floating gate voltage of greater than 0.5 V will not become conducting.

Device	V_{FG0}	α
j	0.24	0.158
k	0.18	0.196
l	0.09	0.288

Table 5.2 Initial floating gate potentials and α values for *p*-channel device j, k and l

The few values of α that could be measured for the devices were approximately as expected. They are a little higher than the *n*-channel values, but this may be due to measurement errors introduced by the low currents in the devices. The α values increase with increasing overlap length in the same way as the *n*-channel devices.

5.1.4 The Measured V_{FG0} Values and the Source of Residual Charge

Although very few results were obtained from the measurements on the *p*-channel devices, those that are available, and the assumptions made concerning the other devices, are important. The potential on a floating gate, with no additional bias applied, appears always to be positive and, for the devices measured, in the range 0.1 to 1.7 V. This implies that the residual charge on the gates is also always positive for both *p*- and *n*-channel devices. The amount of residual charge on the gate ranges from 3 to 170 fC.

The residual charge is most likely to result from the processing of the ICs. The devices had been stored unpowered for 21 months before testing. Consequently, charge is unlikely to have been added to the floating gates during this period. The only conclusion therefore, can be that the charge is deposited on the floating polysilicon islands during processing. Furthermore, as both types of device seem to have positive charge, it would appear that the origin of the charge is common to both types of device.

One possible source of the charge is the plasmas that are used in IC fabrication for the etching of various layers and resists. Consider, for example, the patterning of polysilicon by reactive ion etching. The ions at the surface of the sample are positive as the sample is placed on the cathode of the etching system. Interaction of the plasma with the polysilicon as it etches the layer will therefore result in the transfer of charge into the material. At this point the charge density will be roughly uniform within the sheet. Completion of the etch will

result in a set of isolated polysilicon tracks and gates which all contain positive charge. It is probable that the photoresist layer would then be removed by a further dry etching process during which the isolated polysilicon tracks are exposed to further positive charge. Subsequent exposure of the wafer to oxygen or nitrogen atmospheres will produce a silicon dioxide or silicon nitride coating on all polysilicon surfaces which will trap the residual charge. Later stages of processing will connect all fault free polysilicon tracks to metal tracks providing a conducting path for the tracks to be discharged. Polysilicon gates that remain isolated will retain their positive charge, and this is seen as the unbiased floating gate potential when devices are used.

The argument presented so far suggests that the charge density will be approximately uniform throughout the polysilicon tracks. From the measurements of the floating gate potential, this is not seen to be the case. It is felt that there are two possible causes for this variation.

The first stage of the charge deposition during the reactive ion etch of the polysilicon should result in a uniform charge distribution. The polysilicon layer is a single conducting sheet until the etch is complete, and so for most of the etch a uniform charge distribution will occur. The situation is different during the removal of the photoresist. For most of this process, each isolated polysilicon track is exposed to the plasma only along its perimeter. One might therefore assume that the amount of charge accumulating on each track should be determined by the length of the track perimeter. This would result in higher values of residual charge for longer or wider tracks or devices. This is indeed the case, as was described in section 5.1.2.

Figure 5.10 shows the variation of the floating gate charge with gate perimeter for all of the devices measured. Although there is a wide scatter at each of the perimeter values, it is clear that the residual charge increases with gate perimeter. It is less clear whether this relationship is linear. However, this data provides a better fit to a straight line than a graph of Q_{FG0} against gate area. This suggests that the argument presented above is more realistic than the previous discussion in which the residual charge was related to gate area. It is likely that the gate charge is produced by both plasma processing and ion implantation, and that the former is the dominant source resulting in a net positive charge on all floating gate devices.

The wide variation in Q_{FG} values for devices of the same dimensions probably arises from leakage of the charge through the gate oxide during storage. Consider, for example, devices with a perimeter of 38 μm. The difference between the maximum and minimum Q_{FG} values for these devices is 28 fC. The devices have been stored for 21 months and so this change of charge represents a leakage current of 5×10^{-22} A over this period. This is a very small current indeed and it represents the leakage of one electron every 100 seconds. Although the polysilicon is surrounded by a very good dielectric, this amount of leakage is

quite feasible, particularly since the gate oxide will almost certainly contain some defects. If the leakage is due to defects, the spread in the residual charge values would be expected to increase with larger gate areas. This is seen to be the case, as shown in figure 5.10.

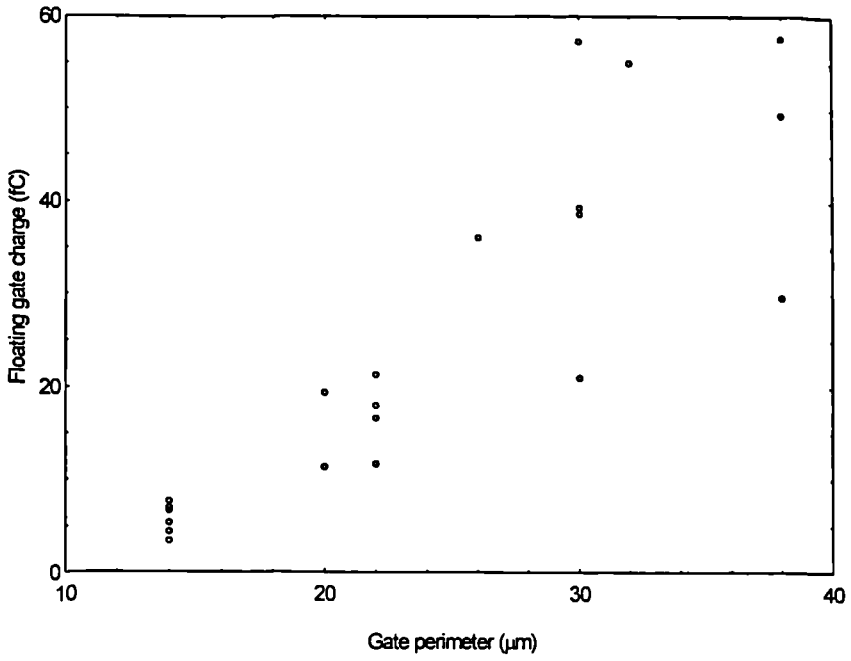


Figure 5.10 Variation of floating gate charge with gate perimeter for all measured devices

Based on the theory presented above, the devices with extensions of the polysilicon gate under the drain metal have lower charge values than expected. They are not included on the graph (Figure 5.10) because they would compress the perimeter axis by a large factor. The residual charge is typically a factor of 5 to 10 lower than expected when the large gate and track perimeter are taken into account. Such large changes would have raised the gate voltage by up to 6.5 V. It is possible that the high initial gate voltage is the reason for the gate voltage now being lower than expected. A high voltage is likely to increase the gate leakage current providing more rapid discharging of the gate. Without further detailed analysis of the charge deposition and leakage mechanisms, no firm conclusions can be drawn on these anomalies.

5.2 Operation of Inverters with Floating Gate Faults

A set of five inverters were fabricated on each test chip. Floating gate faults were introduced into four gates with one fault free circuit included for reference. The configuration of the inverters is described in Table 5.3.

Gate	Configuration
a	Fault free
b	*n*-channel driven, *p*-channel floating
c	*n*-channel floating, *p*-channel driven
d	*n*- and *p*-channel floating and connected together
e	Transmission gate with floating gates driving a fault free inverter

Table 5.3 Floating gate faults in the inverter test circuits

Plots of the test circuits are shown in figure 5.11. All *n*-channel devices are 7μm wide and all *p*-channel devices are 20 μm wide. Bond and probe pads are not shown in this diagram but each inverter had independent supply pads and a probe pad for the output. A common input pad supplied all fault free transistors.

The inverter outputs are unbuffered so that they could be directly monitored. However, because of this, dynamic testing of the gates was not possible. The purpose of the test circuits was to obtain the transfer characteristics for inverters containing one floating gate fault. It should be noted that only specific cases of floating gate faults are being analysed by this process. From the preceding discussion, it is clear that the floating gate can adopt a range of potentials that are quite randomly distributed for a device of a given size. Consequently, the floating gate voltage present on the inverter cannot be predicted. The measurements, therefore, only represent examples of inverter operation with floating gate faults; the characteristics are not definitive.

The circuits can provide useful experimental results and include a floating gate configuration not included in the single transistor sets i.e. the connected floating gate fault. The results should also provide useful references for comparison with simulation results obtained later in the work using models for the floating gate devices.

As the inverter outputs were unbuffered, care had to be taken to ensure that the devices were not influenced by the testing procedure. An electrometer was used to measure the output voltage of the inverter circuits. This presents a very high load impedance to the inverters and emulates the situation on chip to some extent.

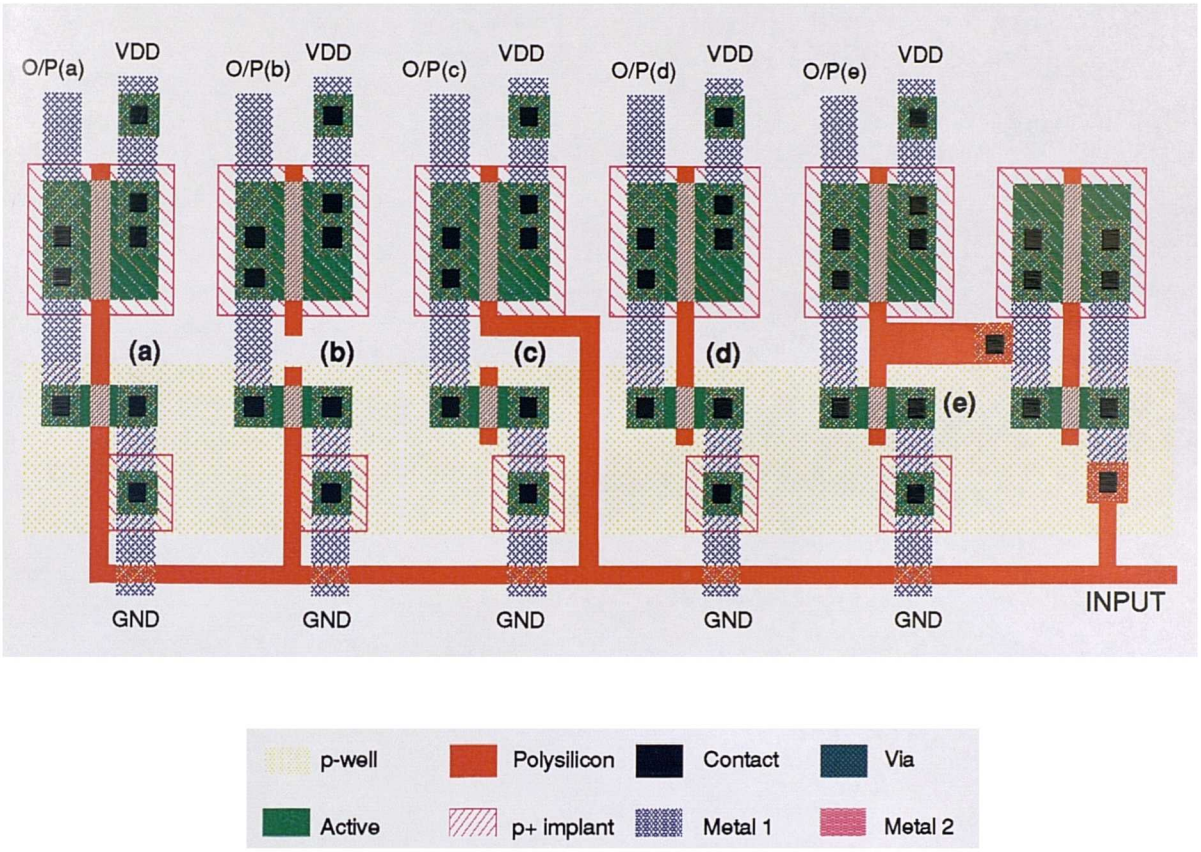


Figure 5.11 Layout diagram of floating gate inverter test circuits

5.2.1 Output Characteristics for the Inverter Circuits

The inverter output voltages for high and low logic states (V_{OH} and V_{OL}) were measured for two samples of each gate. The results are presented in Table 5.4.

	Gate	V_{OL}	V_{OH}
112/C	a	0.0	5.01
	b	0.03	4.74
	c	0.09	4.88
	d	0.11	0.11
	e	0.0	5.01
112/D	a	0.0	5.01
	b	0.0	4.51
	c	1.38	4.99
	d	0.46	0.46
	e	0.0	5.01

Table 5.4 Output voltage for floating gate inverters

The fault free inverter is seen to operate correctly. The faults introduced to inverters **b, c** and **e** allowed control of the circuit output through the fault free transistor. With one exception, the results for the single floating gate circuits show that the output voltages of the inverters are generally within one threshold voltage of the supply. This means that subsequent logic gates would interpret the logic level correctly and produce full logic swings at their outputs for seven out of the eight test circuits. This conclusion can only be made for static testing conditions. Gate **c** (floating gate *n*-channel) on sample 112/D has an output low voltage of 1.38 V. This is unlikely to produce a full 5 V output on a following logic gate. However, after passing the signal through further gates, full logic swings would almost certainly be produced. This will be investigated further in Chapter 7 in which a model of the floating gate device is used to predict circuit performance in the presence of floating gate faults.

It is important to note that both V_{OH} and V_{OL} may be affected by the single floating gate fault. This is explained as follows. For an inverter with a *p*-channel floating gate one

would expect the V_{OL} value to be raised. This is because the *p*-channel device may not be fully switched off when the *n*-channel is on. In fact, the voltage on the *p*-channel gate will be pulled towards 0 V by the inverter output. This will tend to cause the *p*-channel device to become more conducting as the output falls. The output low voltage will therefore be above 0 V.

It would seem reasonable however, for the output high voltage to be at the V_{DD} supply voltage. This is because the *n*-channel device can be fully switched off. This should allow the output to be pulled up to V_{DD} through the *p*-channel device, provided that it is conducting to some degree. In fact, the feedback from the *p*-channel drain, i.e. the inverter output, to the gate once again has an effect. The rising output voltage pulls the *p*-channel gate voltage towards V_{DD}. Consequently, the floating gate voltage may become less than V_{Tp} before the gate output reaches V_{DD}. The final gate output will be determined by the initial floating gate voltage. The same argument can be applied to the *n*-channel devices.

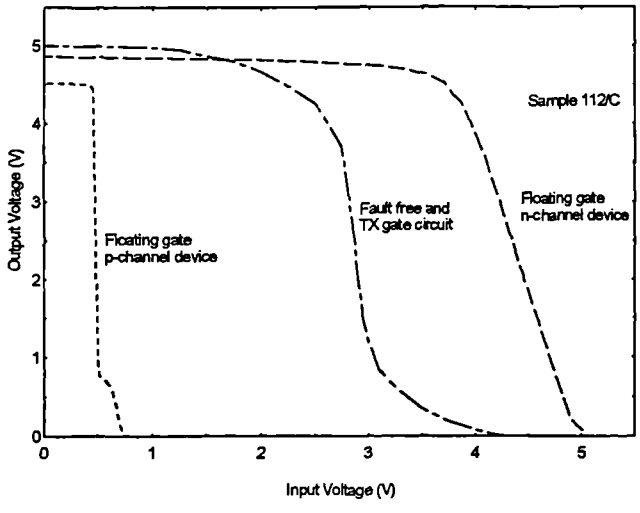
The output voltage for the inverters with interconnected floating gates were unaffected by the input voltage as expected. The low value of the output voltage suggests that the *n*-channel device is more conducting than the *p*-channel. The output is essentially stuck-at logic 0.

The inverter driven by the transmission gate with floating gates produces full logic levels at its output. This suggests that both devices in the transmission gate are conducting. If this was not the case, a threshold voltage drop would occur across the gate producing a measurable deviation from the fault free values for V_{OH} and V_{OL}.

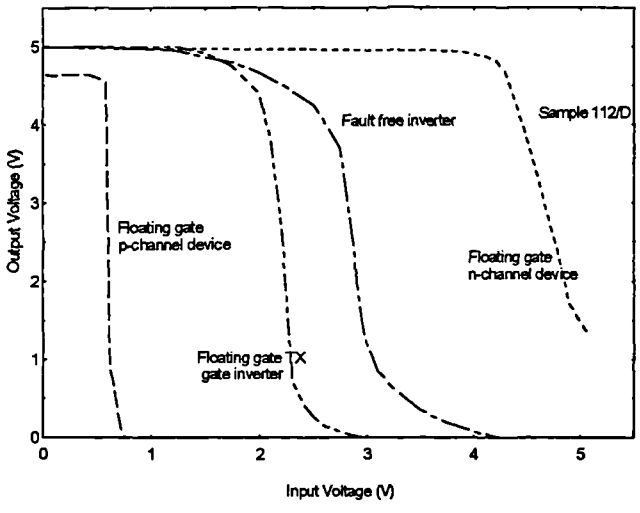
The transfer characteristics for the inverters are shown in figure 5.12. These show extreme shifts in the transfer curves and transition voltages for the floating gate inverters (**b** and **c**). Clearly this will affect the speed of operation of the inverters but it will not necessarily result in an increase in delay. The transition voltage for the floating gate *p*-channel inverter is very low. The input rising to output falling transition will therefore occur very rapidly, probably reducing the output falling propagation delay. The output rising delay will of course be increased because of this shift. For the floating gate *n*-channel circuit, the output rising delay is reduced and the output falling delay is increased. Dynamic testing could not be performed on the devices and so the change in these delays is not known. Delays will be considered in detail when the modelling of floating gate circuits is considered in Chapter 7.

There is an interesting difference in the transfer characteristic for the faulty transmission gate circuits. The output of this gate for sample 112/C is identical to the fault free inverter thus supporting the earlier conclusion that both floating gate transistors are conducting. The transition voltage for this circuit on sample 112/D is lowered from 2.8 V, for the fault free circuit, to 2.2 V. It is difficult to see why the floating gate fault on the

the inverter gate voltage is higher than the input voltage but this is not a sensible conclusion. A parametric fault in the inverter, such as a very short n-channel device, could produce the characteristic observed. Although there is no visual evidence for this hypothesis, it is felt that this is more likely to be the cause of the change than an effect produced by the floating gate fault.



Sample 112/C



Sample 112/D

Figure 5.12 Transfer characteristics of inverters with floating gate faults

5.2.2 Floating Gate Voltages in the Inverter Circuits

The individual floating gate transistors in the inverter test circuits were probed to obtain the I_{DS} - V_{DS} characteristics. The initial floating gate voltage and the α value for each device was then obtained using the method described in section 5.1.2. The results are presented in Table 5.5.

	Device	V_{FG0}	α
112/C	b	0.9	0.282
	c	0.83	0.302
	d	1.5	0.576
112/D	b	Device was non-conducting	
	c	0.74	0.153
	d	0.84	0.58

Table 5.4 Initial floating gate voltage and α values for floating gate devices in the inverter circuits

The main reason for making these measurements is to provide experimental data for comparison with the models discussed in Chapter 6. Using the results obtained and the model developed, it should be possible to correlate circuit behaviour with floating gate device parameters. The floating gate and α values obtained for these devices are roughly as expected. The high value of α for the circuit **d** devices (connected floating gate) may be due to the coupling of the drain voltage through two devices to the common gate. This will be investigated further in Chapter 7.

5.3 Operation of Logic Gates with Floating Gate Faults

Several sets of test circuits were included for the investigation of floating gate fault effects in logic gates. Despite serious problems with the circuits, some useful results have been obtained. In addition, a significant result concerning open circuit supply faults has arisen from the testing. As described in chapter 3, the logic gate test circuits consisted of sets of two-input NOR gates with faults introduced in various ways. The test gate was buffered on its inputs and output, and the output could also be directly probed.

5.3.1 Problems with the Electrical Testing of the Logic Gates

Four sets of test gates were fabricated over a period of four years. The reason for the repeated redesign and fabrication was that the circuits exhibited a very high static current consumption and a tendency to go into latch-up. The two problems were felt to be related and so the source of the excess static current flow was investigated.

The test circuit had been designed to allow each logic gate (plus its inverters) to be powered independently. This would provide information on the current consumption of faulty gates. To reduce the pin count for the test circuits, a common 0 V line was used for all logic gates but V_{DD} was supplied to each circuit through its own supply pad. The supply current for each of the test circuits (including the fault free gate) was found to be in the range 0.8 mA to 2.1 mA and this varied with the input signals that were applied. The current was always higher when both inputs were at logic 0.

Although the supply current was high, tests on the fault free gate in the first set of test chips showed that the gate appeared to operate correctly. The output voltage levels for the gate were 0 V and 5 V. The first set of test circuits however, were very susceptible to latch-up and this severely hampered testing. The two subsequent test circuits also had latch up problems despite a redesign of the substrate and well taps. The failure of the third set of test chips was probably due to poor processing.

The supply current for the fourth set of test circuits was in the same range as the previous samples, but the circuits did not appear to be so susceptible to latch-up. Latch-up could be induced by electrical interference (for example, switching mains electrical apparatus near the sample) but it did not occur under normal operating conditions. It was during the testing of these circuits that the reason for the high static supply current and latch-up problems became apparent.

The problem was occurring because, contrary to expectation, the separation of the V_{DD} supplies for each logic gate did not result in only a single gate being powered up. By powering a single test circuit, the substrate was, of course, taken to V_{DD}. All other test circuits have a common substrate with connection to their independent V_{DD} supply track and pad. Consequently, by applying 5 V to one test circuit, all other logic gates were being supplied through the substrate. As the current to the "unpowered" gates flowed through the significant resistance of the substrate, a voltage drop occured and the potential on the V_{DD} tracks of these gates was less than the 5 V supply. This was confirmed by probing the V_{DD} supply pads of unpowered logic gates which all exhibited a voltage of between 3.8 and 4.5 V. The voltage depended to some extent on the location of the gate with respect to the powered circuit.

A sufficient condition for latch up to occur in a bulk CMOS IC is for the substrate potential to fall approximately 0.6 V below a *p*-channel source or drain. This causes a lateral *pnp* transistor in the substrate to switch on which in turn switches an *npn* device on completing a low resistance path between the supplies. The result is a very high supply current limited only by the substrate and well resistances. The parasitic devices and resulting circuit are shown in figure 5.13.

The potential difference between *p*-channel source and substrate (or *n*-channel source and well) is often brought about by transient spikes on logic gate outputs caused by high

capacitance loads. The standard way to reduce latch up problems is to ensure that the substrate potential does not fall below the supply voltage and this is achieved by ensuring that substrate taps are placed liberally around the circuit. A typical design rule would require all active regions to have substrate or well taps within 100 μm of each other.

It is not obvious that the potential difference required to trigger latch-up will be developed in the test circuits being considered. The substrate current flow must be in such a direction that the substrate potential near the *p*-well is lower than that of the local *p*-channel sources. However, this is certain to happen in some of the test gates and latch-up will occur.

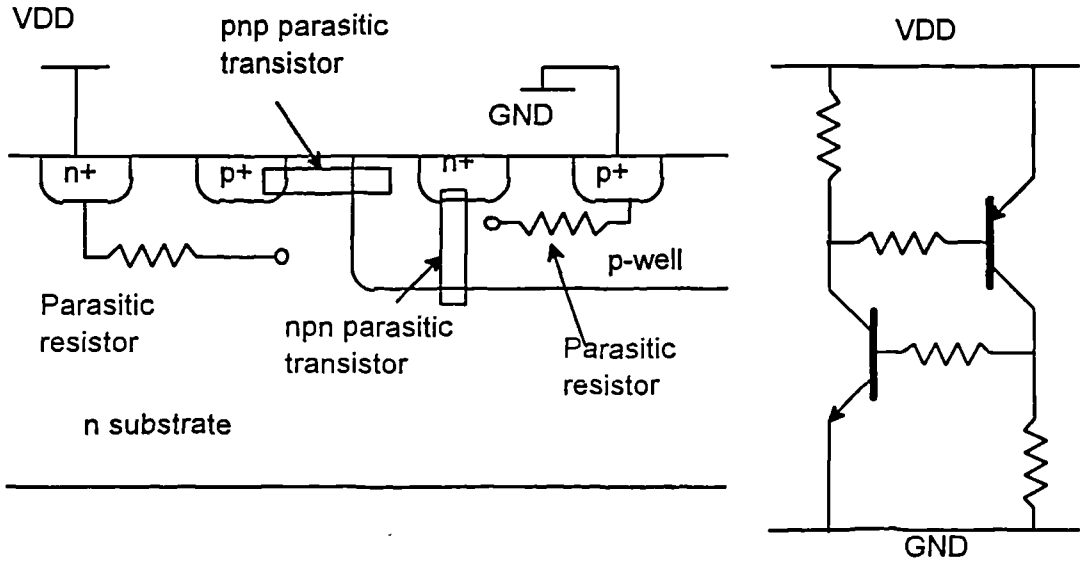


Figure 5.13 Latch up in a bulk CMOS circuit

The above description explains the latch-up behaviour of the fault free test circuits. It also suggests a possible failure mode for an IC which contains a V_{DD} supply stuck-open fault. It is generally assumed that this fault will lead to the outputs of all gates affected being stuck-at 0. This will not be the case provided that the faulty gates have a surviving connection to the substrate. This situation is very likely, since substrate taps are placed close to active regions for the reasons previously described. The faulty logic gates will be supplied through their substrate taps and will probably operate quite well. The resistance in the supply of such gates will typically be of the order of kilohms but this is unlikely to cause a serious degradation of the circuit performance and indeed, may not cause a failure to occur. One possible result is that the circuit will be susceptible to latch-up but this is difficult to predict. The effects of V_{DD} supply stuck-open faults will be quite variable, and in many cases they will not result in the catastrophic stuck-at 0 fault that is normally predicted.

The source of the latch up problem has been identified above. The misconception over power supply isolation also helps to explain the presence of high supply currents in the test circuits. Rather than supplying a single NOR gate and three inverters, all logic gates in the set

were being supplied. The high value of supply current occurs because of the faults introduced to the test circuits. Some of the faults lead to both *p*- and *n*-channel networks in the logic gate being active simultaneously. This results in high static supply currents for these gates. Other gates include short circuits which will have a similar effect on the supply currents. The GND line was cut on some of the samples so that the supply currents for individual gates could be measured and the supply current for the fault free gates was found to be very small (less than a few nA). Other gates that had floating gate faults or short circuits had higher supply currents some of which were several hundreds of microamps.

High static supply current has been suggested as an indicator of certain types of fault in CMOS ICs [2.56]. It is apparent from this discussion that many of the faults introduced result in a high static or *quiescent* current (I_{DDQ}). This will be discussed in detail in Chapter 7 when tests for floating gates are considered.

5.3.2 Logical Operation of Gates with Floating Gate Faults

The problems described in the previous section should not affect the operation of the logic circuits, provided that each is powered directly for testing. It is therefore a valid exercise to analyse the logical operation of the NOR gates despite the fault in the V_{DD} supply. The aim was to provide data on the output voltage levels of the faulty gates and to assess the ability to drive subsequent logic gates. The three forms of floating gate fault previously described had been included in the logic gates. Each test circuit was independently powered and the complete set of test vectors for the two input gate were applied. It should be noted that the inputs were buffered by a single inverting stage. Results were obtained for eight samples for each type of test circuit.

The output voltage for the floating *n*-channel test circuit (Gate H) ~~for all input combinations~~ is shown in Table 5.5. The output voltage is seen to differ from the expected value on several samples for two input combinations: (0,1) and (1,1) (logic levels) on Input 1 and Input 2 (IP1 and IP2) respectively. For the first of these combinations, the fault free output value would be 0 V. In the faulty gate, the output voltage varies from 0.1 to 3.57 V, indicating that the floating gate device is unable to pull the output to 0V. However, for most samples the output is sufficiently low to be interpreted as a logic 0 by the output inverter. The logic gate therefore appears to be fault free with the *n*-channel floating gate fault for seven out of the eight samples tested. This assumes static testing.

For the input combination (11), the fault free output would be 5 V. For the faulty circuits the output voltage varied between 3.86 and 4.95 V. In many cases, this is sufficiently high to be interpreted as a logic 1 by the output inverter and an output voltage of less than 0.1 V is produced. In some cases the output low voltage for the output inverter rises to around

0.2 V but this is sufficiently low to be interpreted as a logic 0 by subsequent gates. All samples appear to operate correctly for this input combination.

Sample	V_{OUT} (inputs = 01)	V_{OUT} (inputs = 11)
915 / 2B	-	4.75
915 / 4A	0.11	4.30
915 / 7A	0.47	4.62
915 / 7B	0.34	4.48
915 / 8A	3.57	4.95
915 / 8B	3.57	4.93

Table 5.5 Output voltages of NOR gates with *n*-channel floating gate faults

Measurements on the NOR gates with floating gate*p*-channel devices produced similar results. Although the output of the NOR gate was not at 5 V for the input combination (11), the output inverter interpreted the logic level correctly for six out of seven samples tested.

The output of the NOR gate with interconnected floating gates appeared to be stuck-at 0 for all input combinations. The output voltage did rise to around 0.2 V on some samples but this is clearly interpreted as a logic 0 by the output inverter.

5.3.3 Floating Gate Voltage Measurements

An alternative method to that previously described was used to ascertain the initial floating gate voltages for these devices. Fault free transistors had not been fabricated on these samples so that the comparison technique could not be used. Furthermore, the floating gate*p*-channel devices were in series with a second device which would alter the characteristics obtained.

The technique used involved the use of SPICE simulations of the logic gate with the floating gate voltage represented by an independent voltage source. The fault free inputs of the NOR gate were set to values which would reveal the fault condition (i.e. (01) or (11)). The floating gate voltage was then varied and the output voltage monitored. A graph of output voltage against V_{FG} is shown in figure 5.14. The graph can be used to find the potential of the floating gate by comparing the measured output voltage with the simulated characteristic.

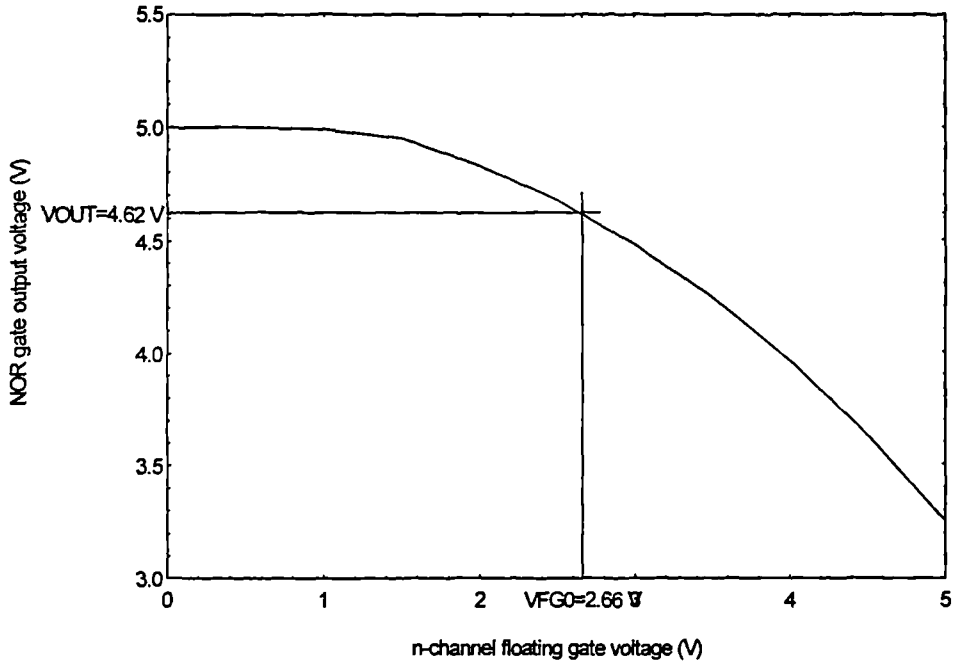


Figure 5.14 NOR gate output voltage against V_{FG}

The procedure was used to obtain V_{FG} for the floating gate *n*-channel and *p*-channel devices. Two values were obtained for the circuit with the fault on the *n*-channel device; one for each input condition that revealed the fault. The floating gate voltage for the *n*-channel devices is given in Table 5.6.

Sample	V_{FG} (inputs = 01)	V_{FG} (inputs = 11)
915 / 2B	-	2.25
915 / 4A	0.89	3.45
915 / 7A	0.88	2.65
915 / 7B	0.89	3.00
915 / 8A	0.77	1.45
915 / 8B	0.77	1.50

Table 5.6 V_{FG} for *n*-channel devices in NOR gates test circuits determined by simulation

The floating gate voltages obtained for the input combination (01) are very close to the values reported earlier in this chapter. It should be noted that the devices were obtained from two different fabrication runs at the EMF. These values almost correspond to the zero bias value of V_{FG} as the output voltage for most cases is close to 0 V. The V_{FG} values for the input combination (11) are higher and this is to be expected as the gate voltage is raised by coupling from the logic gate output. The coupling appears to be higher for these devices than

from the individual transistors discussed earlier in this chapter. For example, the α value for sample 915/4A is 0.609 which is two to three times higher than the ratio obtained from individual devices. It is also higher than the theoretical maximum. It is most probable that this high value of α is due to inaccuracies in the SPICE models used for the simulations. Although the models were provided by the EMF, they were not generated specifically for the samples being tested. Experience has shown that changes in certain model parameters can significantly affect simulation results.

The single measurements of the floating gate voltages of *p*-channel devices in the logic gates were consistent with the measurements on individual devices. The floating gate voltage varied from -0.68 to -0.81 V. With one exception, the drain-source voltage for these measurements was around -4.7 V. These voltages correspond to an initial floating gate voltage of around +0.2 V, if a value of 0.2 is assumed for the coupling ratio, α.

In general, the values of the floating gate voltages obtained for these samples are similar to those obtained for samples 112-3/A-D. This is an interesting point to note, as the samples were fabricated on different runs of the EMF process. Furthermore, the measurements on the logic gates were spread over a period of 15 months with the first measurements being made 12 months after fabrication. The small spread observed in the floating gate voltages could suggest that there will be little further change in the voltage with time and that the values currently observed may be the long term values of the floating gate voltage. This suggestion requires further investigation.[1]

The logic gates generally produce the correct logic output to within 0.5 V of the fault free value. These voltages are sufficient to drive subsequent logic gates correctly. This indicates that with static testing, the floating gate faults would not be revealed.

5.3.4 Dynamic Testing of Logic Gates with Floating Gate Faults

A set of NOR and NAND gates were included in the test set to enable high speed testing of the faulty gates. The circuits consisted of four NAND and four NOR gates containing the three types of floating gate fault previously described. A fault free gate was also included. The four logic gate outputs were multiplexed through transmission gates to a pad driver. This was necessary as there was insufficient space on the test chips for eight output pad driver circuits. Unfortunately, due to a mistake in the layout of these circuits, the multiplexers were incorrectly connected with the result that the circuits could not provide any useful test data on the dynamic operation of the logic gates.

However, the detailed data on floating gate device operation that was obtained from individual devices has assisted in the development of a simulation model for the floating gate

1. Measurements of V_{FG0} made on the devices on samples 112-3/A-D after storage for a further 6 months indicate a slight further decay of charge from the gates. Shifts in V_{FG0} of around -0.1V were measured for *n*-channel devices. This corresponds to a leakage rate of one electron per thousand seconds of storage. This suggests that the rate of charge leakage from the floating gate is indeed decreasing with time as proposed.

transistor. As a result, it has been possible to analyse the dynamic behaviour of circuits containing these faults. This analysis is described in detail in Chapter 7 of the thesis.

5.4 Chapter Summary

Electrical measurements on a range of floating gate devices and circuits have been described. It has been shown that floating gate devices, arising from design or photolithography defects, can have a range of values of residual gate charge. This charge generally appears to be positive for both *n*- and *p*-channel devices for the process used in this investigation. A possible source of the charge has been suggested with supporting evidence for the proposed theory. The charge is felt to arise from plasma processing stages and ion implantation. The amount of charge has been seen to vary roughly in proportion with the perimeter of the floating gate. A probable reason for the variation in gate charge for identical devices is slow leakage through imperfections in the surrounding dielectric layers. A spread in the amount of leakage between devices was observed.

The range of values of residual charge leads to a range of unbiased floating gate potentials. For the devices measured, these varied from 0.1 to 1.7 V for *n*- and *p*-channel devices. It is important to note that the gate potential may be less or greater than the typical *n*-channel threshold voltage of 0.9 V. Consequently, the conduction of *n*-channel floating gate devices is extremely variable. The positive unbiased floating gate potential found on *p*-channel devices results in most of these devices being non-conducting. Only a small number of devices with large gate to drain coupling could be made to conduct.

The gate voltage of a floating gate transistor varies significantly with the applied drain-source potential. The amount of coupling of the drain-source potential to the gate appears to vary between devices and is generally in the range 0.1 to 0.3. This phenomenon results in a wide range of currents for any given device. Transistors have been observed to switch from a subthreshold state to a strongly conducting state due to this coupling. The coupling ratio appears to be reasonably constant for some devices but varies significantly for other devices as the transistor becomes more conducting. The ratio can vary from 0.35 to 0.2 for some devices. There is some evidence to suggest that the coupling ratio is related to the initial gate voltage - higher V_{FG0} values, implying a high coupling ratio.

Static tests of inverter circuits with floating gate faults appear to indicate that the inverters may operate satisfactorily with floating gate faults present. This is to a large extent due to the variable conduction of the floating gate devices. Similar tests on two-input NOR gates with floating gate faults show that such gates can produce correct logic values at their outputs. Furthermore, the faults are even less apparent when the outputs of subsequent gates are observed. Dynamic testing of the circuits has not been performed.

An unexpected result concerning stuck-open faults in V_{DD} supply tracks has arisen from this investigation. It is generally assumed that a break in a V_{DD} supply track will result in all outputs on affected gates being stuck-at 0. This is not the case if the faulty circuits remain connected to the substrate. In this case, the circuits receive supply current through the substrate and they can operate satisfactorily. They will have reduced output voltage levels for logic 1 but this will normally be corrected by subsequent logic gates. The circuits will be more susceptible to latch-up when this form of failure occurs. However, latch-up will not always occur. The fault may therefore produce intermittent behaviour in logic circuits rather than the fatal fault effects normally expected.

The measurements presented in this chapter have shown that the floating gate transistor is not simply a faulty switch as is often assumed. The transistor can be fully off, weakly conducting or switched fully on. Furthermore, the conductivity of the transistor varies with applied drain and source potentials. The measurements have provided an indication of the way in which the device operates. However, in order to understand the effect of this fault on a wide range of circuits, a detailed understanding of the device operation is required. This understanding should ideally be expressed in a model that can be used to predict the performance of logic circuits with floating gate faults present.

In the next chapter we will develop a model of the floating gate transistor and describe its operation in detail. The results obtained in this chapter will be used to verify the model which can then be used to predict the operation of logic gates with floating gate faults introduced. This analysis will the allow testing techniques to be identified which will reveal the fault, resulting in better fault coverage for CMOS integrated circuits.

Chapter 6
A Model Of The Floating Gate MOS Transistor

It has been established in chapters 2 and 3 that the floating gate fault is a relatively common and significant fault in CMOS VLSI circuits. The experimental analysis presented in Chapter 5 has given a good indication of the characteristics of floating gate transistors. However, information on the operation of logic circuits containing floating gate faults is limited to a set of circuits with specific values of floating gate charge. In order to be able to predict the effects of this fault more generally, it is necessary to understand the operation of the floating gate MOS transistor in more detail. This will allow a model to be developed which can be used to predict the performance of any circuit containing floating gate faults.

In this chapter, we review the current understanding of the operation of the floating gate MOS transistor. The literature on this subject is limited and so we develop a detailed model of the device, and new equations for the floating gate potential and the drain-source current are derived. The model is evaluated numerically using a computer program and some results from the model are presented. The predictions of the model are compared with the characteristics obtained from measurements of floating gate devices described in Chapter 5.

6.1 Existing Models

The models of the floating gate transistor that are presented in the literature are generally based on experimental analysis of device and circuits. The earliest discussion of the device is presented by Kahng and Sze in 1967 [6.1] where it is postulated as a non-volatile storage device. More recent work has investigated the effect of the floating gate transistor on circuit operation, when it occurs as a fault. The models produced by this work will now be discussed.

6.1.1 The Floating Gate Avalanche MOS Transistor

The floating gate transistor occurs not only as a faulty device in CMOS IC's but has also been used as a non-volatile memory storage element for many years. The first practical device was demonstrated by Frohman-Bentchkowsky in a publication in 1971 [6.2]. In a subsequent publication [6.3] Frohman-Bentchkowsky named the device the "Floating Gate Avalanche MOS transistor", which gives rise to the acronym "FAMOS". It is worth considering this paper in some detail as it contains useful theory and experimental results.

The FAMOS device described in [6.3] is a simple *p*-channel MOS transistor with a polysilicon gate which is left unconnected. The conduction through ~~of~~ the transistor is determined by the amount of negative charge on the 'floating' gate which is varied by inducing avalanche injection from the depletion regions around the drain or source. A reverse junction voltage (greater than -30V in the paper) applied across either junction results in avalanche multiplication of electrons in the depletion region. Some of the electrons thus produced have sufficiently high energies to overcome the oxide-silicon potential barrier, and due to the electric field present, they will enter the floating gate depositing a net negative charge on the gate. This charge induces a positive charge in the surface of the semiconductor which, if sufficiently large, will invert the channel region causing the transistor to conduct in the conventional manner.

Frohman-Bentchkowsky gives the I_{DS} - V_{DS} characteristics for the FAMOS transistor in two charged states. These are shown in figure 6.1 which is copied from the paper.

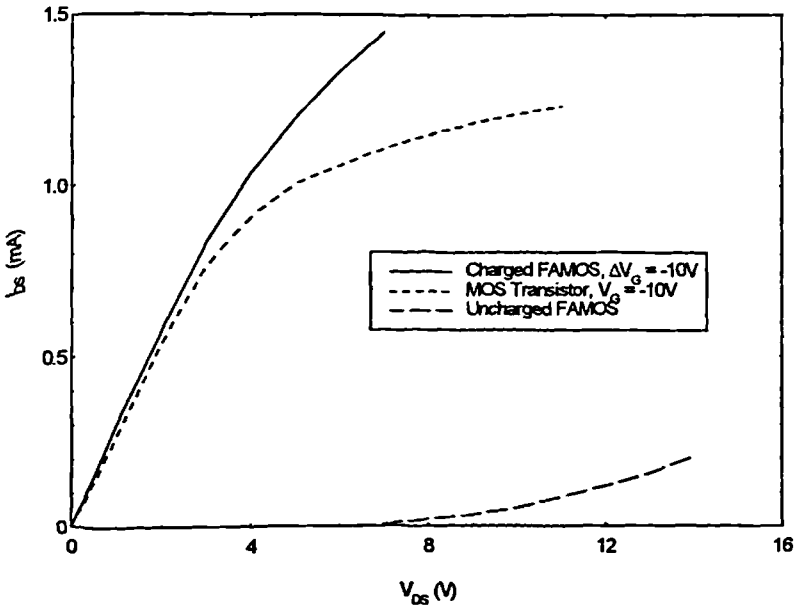


Figure 6.1 The I_{DS}-V_{DS} characteristics of 'charged' and 'uncharged' *p*-channel FAMOS transistors and a conventional MOST with V_{GS} = -10V. From [6.3]

The 'charged' FAMOS has a gate-substrate voltage of -10V. The gate voltage cannot, of course, be directly measured as there is no connection to the gate. However, the voltage can be estimated from comparison of the drain-source current in an identical MOS transistor with a gate connection. By plotting the I_{DS}-V_{DS} characteristic of this device, the gate voltage corresponding to the measured current in the FAMOS transistor can be obtained from the graph using the technique described in Chapter 5. The 'uncharged' FAMOS has no net charge on the gate.

As can be seen from Figure 6.1 the conductance of the charged and uncharged transistors varies with the drain - source voltage V_{DS}. The characteristic for a conventional *p*-channel MOS transistor with V_{GS} = -10V is shown for comparison. The uncharged FAMOS device conducts significantly for V_{DS} > -8V. This implies that the gate voltage, which was initially 0V, must have become negative, inducing a channel in the semiconductor and allowing current to flow. This occurs by coupling of the drain potential to the gate.

The coupling between the drain voltage and the gate is explained by Frohman-Bentchkowsky as arising from "*capacitive feedthrough from the drain terminal to the floating gate*". He models the capacitive coupling in the FAMOS device by the following equation.

$$V_{GS} = V_{DS} \cdot \frac{C_{GD}}{C_{GD} + C_{GS} + C_{GB}} \qquad \textbf{(6.1)}$$

where C_{GD} = gate-drain capacitance, C_{GS} = gate-source capacitance and C_{GB} = gate-substrate capacitance.

C_{GD} and C_{GS} consist of fixed overlap capacitances and contributions from the gate oxide capacitance. The latter is split between C_{GD} and C_{GS} according to the conductive state of the channel.

Frohman-Bentchkowsky presents some experimental results and shows that the capacitance ratio in Equation 6.1 is constant i.e. there is a linear relationship between V_{GS} and V_{DS} in the FAMOS device. Graphs of the change in V_{GS} against the change in V_{DS} is given in the paper for charged and uncharged devices. They are a reasonable fit to a straight line but the slopes of the two graphs are different. The reasons given for the apparent linearity of the relationship are different for the two cases.

For the charged FAMOS Frohman-Bentchkowsky suggests that as a "*continuous conductive path (exists) between the drain and source ... the ratio C_{GD}/C_{GS} is independent of V_{DS}*" This is in fact an oversimplification because the charge distribution in the channel will vary with V_{DS} causing the capacitance ratio to vary. Furthermore, the variation of the depletion region capacitance has been ignored in this argument. These two factors suggest that the total capacitance ratio of Equation 6.1 will vary with V_{DS}.

A different argument is given to explain the apparent linearity of the V_{GS}/V_{DS} plot for the uncharged device. Since V_{GS} is always less than V_{DS} in this case, the device will be saturated for at least some of its operation. The presence of the pinch-off region and the variation of the capacitance associated with it will clearly affect the C_{GD}/C_{GS} ratio. However, this variation is said to be balanced by the effect of the electric field in the oxide on the length of the pinch-off region [6.4] resulting in a linear relationship for this mode of operation. The variation of the capacitance of the bulk depletion region beneath the channel has, however, been ignored.

Frohman-Bentchkowsky's explanations for the apparent linearity of the V_{GS}/V_{DS} relationship are therefore incorrect as they do not account fully for all variations of the capacitances with V_{DS}. It will be shown in subsequent sections that the relationship is non-linear. Furthermore, the analysis presented in [6.2] only considers the two charge conditions previously described; for the operation of the floating gate transistor arising due to a fault, a continuous range of charge conditions must be considered.

Other aspects of the operation of the FAMOS device have been modelled. Most publications on the subject, (for example [6.4], [6.5]), consider the charging and discharging process which is not important in this context.

6.1.2 Experimental Analysis of Circuits Containing Floating Gate Faults.

The author has been able to find only three ~~four~~ papers that report experimental work on the characteristics of faulty floating gate transistors in CMOS circuits: Renovell et al [6.6, 6.7], Maly et al [2.62] and Henderson et al [2.23].

The earliest papers on the subject by Renovell, Faure and Cambon illustrate the layout dependency of floating gate fault effects. The faults are introduced by cutting tracks in a fault free circuit using methods that are not described. However, it should be noted (and is not by the authors) that the ends of cut tracks that are open to the atmosphere will be able to lose or gain charge. This would not be the case with, for example, a polysilicon track which is completely surrounded by oxide due to a fabrication fault. This is an important distinction because the value of the potential of the floating gate (which is determined by the residual charge) has a significant effect on the floating gate transistor characteristics. Three types of floating gate fault are defined by Renovell et al, and these are discussed below:

a) *The "Stuck-at" floating gate.* This is said to arise where the gate is not connected to its driving signal, but is connected to an active area in the circuit as shown in figure 6.2. The active area will be biased by its reverse biased diode connection to the substrate which causes the gate potential to follow the substrate (or well) potential, normally V_{DD} or 0V. The node is therefore assumed to be stuck-at 1 for a *p*-channel device and stuck-at 0 for *n*-channel. It will now be shown that this reasoning is incorrect as the gate cannot be

considered to be a 'stuck' node.

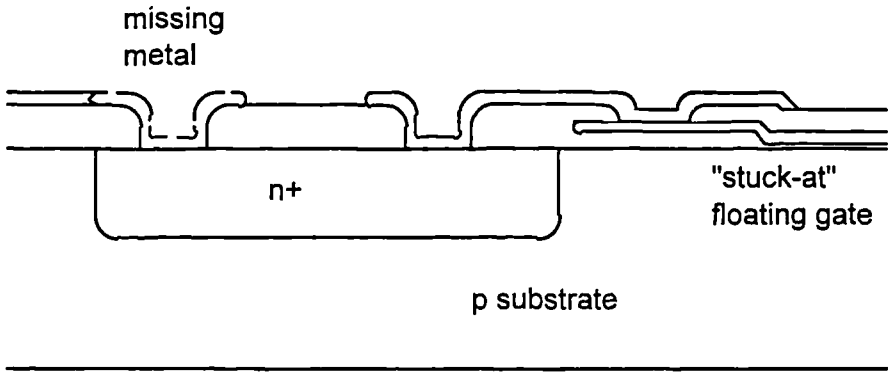


Figure 6.2 Example of the *stuck–at floating gate fault*

The rate at which the floating gate capacitance charges will be limited by the very small reverse saturation current of a drain or source to bulk junction which is typically of the order of femtoamps or less. Given that the gate capacitance is a few tens of femtofarads, the maximum rate of change of the gate voltage due to charging by the reverse saturation current (given by $\Delta V = I\Delta t/C$) is of the order of 0.1 $V.s^{-1}$. A change in the gate voltage due to capacitive coupling from the source or drain could therefore persist for several seconds until the reverse saturation current discharges the gate capacitance. Consequently the changes occurring in the drain or source voltages on the nanosecond time scales associated with IC operation, will affect the gate voltage as shown by the following example.

The two-input NAND gate in figure 6.3(a) has a *stuck-at floating gate fault* on transistor M1: the gate is connected to a "floating" active track as shown in figure 6.2. Given sufficient time, the gate of M1 will charge to V_{DD} switching the *p*-channel transistor off. Consider the inputs *a,b* to change from $a=V_{DD}$, b=0V to $a=V_{DD}$, $b=V_{DD}$. The output of the gate will change from V_{DD} to 0V. In doing so the capacitive coupling from the drain to the gate of M1 will cause the gate potential to drop as shown by simulation in figure 6.3(b). As stated above, the reverse saturation current will start to charge the gate towards V_{DD} taking several seconds to do so, allowing the *p*-channel device to switch on, thus pulling the node *f* towards V_{DD}. The rising voltage on *f* will then feed back to the gate raising its potential, and tend to switch the device off.

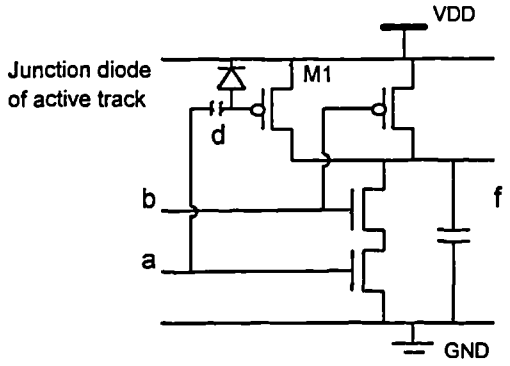


Figure 6.3 (a) NAND gate with *stuck-at floating gate fault*

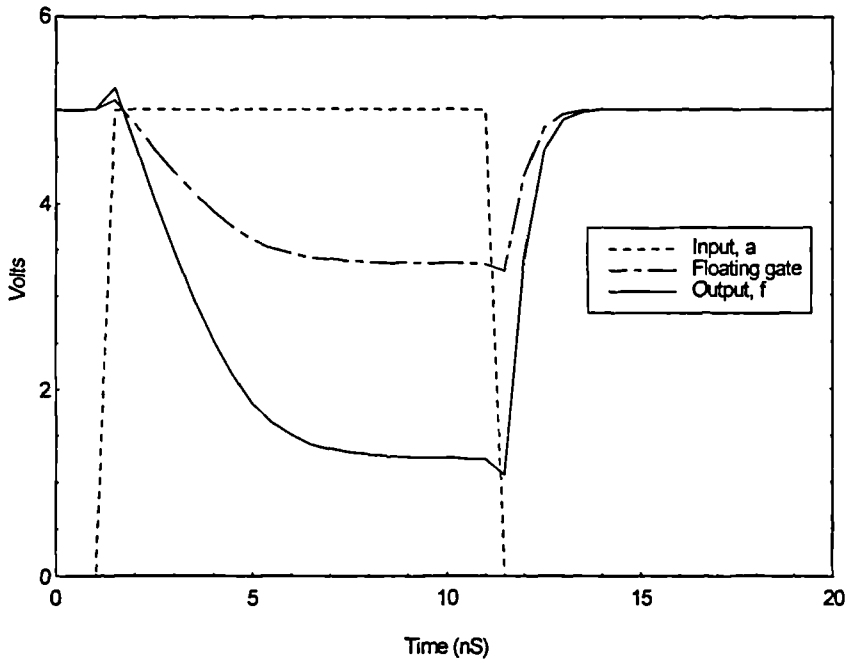


Figure 6.3 (b) Simulated circuit response to the test waveform

The precise effect of the output feedback depends on many circuit and device parameters, but it will generally result in the floating gate *p*-channel device conducting when the gate output is at logic 0. The floating gate is clearly not stuck-at 1 as suggested in [6.5] and if modelled as such may lead to incorrect prediction of circuit operation.

b) *The "Free" floating gate.* In this case the gate electrode is completely isolated with no capacitive coupling or DC biasing from external components or signals. A "free" floating gate is capacitively coupled to its drain and source electrodes.

c) *The "Influenced" floating gate.* In this case the potential of an isolated gate electrode is influenced by the surrounding signal tracks through capacitive coupling. This is a generalised case of the *free floating gate*. The relative strengths of the external and internal device couplings will determine the significance of the external influencing signals. This is particularly dependent on the layout topology and is very difficult to predict.

From the above discussion it is apparent that the *stuck-at floating gate* and the *influenced floating gate* are special cases of the simple *free floating gate*. Experiments are reported in the paper by Renovell et al which were designed to ascertain the relationship between V_{GS} and V_{DS}. A range of voltages were applied to a track which is *influencing* the floating gate and, the I_{DS} - V_{DS} characteristics of the floating gate device were measured. The gate voltage for each influencing potential is then determined by superimposing the measured characteristics on a set of fault free curves as described in Chapter 5 and from this the effect of V_{DS} and V_{MS} (the influencing potential) on the gate potential can be found. The conclusions drawn give a simple relationship between V_{DS} and V_{GS} as given in Equation 6.2.

$$V_{GS} = \alpha . V_{MS} + \beta . V_{DS} \qquad (6.2)$$

The constants α and β are technology and layout dependent and are given as $\alpha = 0.2$ and $\beta = 0.16$ for the test structures that are used. This expression is a reasonable estimate of the device operation. However, no theoretical justification is given and, as will be shown, a wide range of values for β can occur. The parameters also vary with applied bias and this has a significant effect on the device operation This must be accounted for in an accurate model of the floating gate MOS transistor. Furthermore, the Renovell model does not allow for any residual charge on the gate of the faulty devices.

Maly, Nag and Nigh [6.8] present a model for the floating gate transistor which is similar to that described by Frohman-Bentchkowsky for the FAMOS device. Although they consider four separate ranges of V_{DS}, they state, incorrectly, that V_{DS} must be greater than V_{GS}. This assumes that the initial floating gate voltage is 0V. This assumption has been shown to be incorrect in the preceding chapter, and the value of V_{FG0} has a significant effect on the device operation. Furthermore, the capacitance model used by Maly et al does not allow for the complex interaction of the gate voltage and device charges that will be described later in this chapter.

More recently, Henderson, Soden and Hawkins [6.8] have described measurements on floating gates caused by electromigration failures. It is demonstrated that tunnelling conduction occurs in such devices leading to a range of values for V_{FG0}. However, the model of the floating gate transistor that is used for analysis in this paper is based on simple fixed device capacitances.

The theoretical and experimental analyses presented in all of these papers use two invalid assumptions (in various combinations) of the faulty floating gate transistor operation. (In fairness, Frohman-Bentchkowsky 's paper does not address a faulty device). The two assumptions are:

a) that the relationship between V_{GS} and V_{DS} is a simple linear proportionality, and

b) that the floating gate is either completely discharged or "fully" charged.

Neither of these assumptions can be justified for a faulty floating gate transistor and they will be shown to be erroneous in subsequent sections. A full explanation of the operation of the floating gate MOS transistor is now presented.

6.2 Operation of the Faulty Floating Gate MOS Transistor

The operation of the faulty floating gate *n*-channel MOS transistor will now be considered. The *p*-channel floating gate device is assumed to operate in the same way as the *n*-channel device and will not be discussed in this chapter. The faulty floating gate transistor is assumed to have no electrical connection to the gate terminal. As shown in Chapter 2, this may arise due to a break in a signal track either at processing or during operation, or due to a

faulty contact again arising from a processing defect or from wearout mechanisms. The track or contact break is assumed to be completely open circuit. Furthermore, the conductance of the dielectric layers surrounding the gate is assumed to be negligibly small. Consequently, the total residual charge on the gate electrode is constant. The source to substrate potential, V_{SB} is assumed to be 0V. The floating gate transistor is illustrated in figure 6.4.

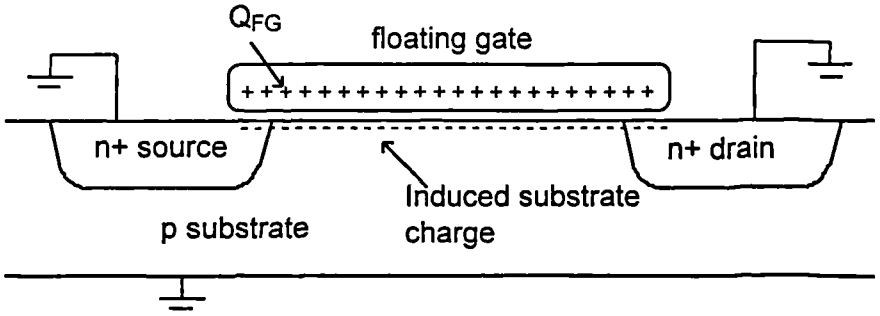


Figure 6.4 Floating gate *n*-channel MOS transistor with a residual charge Q_{FG} and V_{DS} = 0V

6.2.1 The Floating Gate Potential with Zero Drain Bias.

The potential of the floating gate may be expressed in terms of the residual charge on the gate, Q_{FG}, the semiconductor surface potential, ϕ_S, and the gate oxide capacitance, C_{OX}. The transistor gate forms one electrode of a parallel plate capacitor with the gate oxide as the dielectric and the semiconductor surface as the second electrode. The potential across the capacitor is $V_{FG} - \phi_S$ and hence we have:

$$V_{FG_0} = \frac{Q_{FG}}{C_{OX}} + \phi_s \tag{6.3}$$

Note. The zero subscript indicates that this is the gate potential for V_{DS} = 0V.

The value of the surface potential is determined by Q_{FG}. If the gate charge is insufficient to induce an inversion layer at the surface then all of the charge in the semiconductor due to the gate charge is contained within a depletion region with the surface potential maintaining it. Assuming that all acceptors in the depletion region are ionised and that free carrier concentrations are negligible, (the *depletion approximation* [6.8]) the width of the depletion region, W_D, is given by:

$$W_D = \frac{Q_{FG}}{qN_A WL} \tag{6.4}$$

where N_A is the acceptor doping concentration in the *p*-type substrate and W and L are the transistor width and length respectively.

The depletion region at the surface of the semiconductor can be treated as a single sided *p-n* junction. The width of such a region is normally represented by the following equation:

$$W_D = \sqrt{\frac{2\varepsilon_0\varepsilon_{Si}}{q}\phi_S\frac{1}{N_A}} \tag{6.5}$$

where the junction potential has been replaced by the surface potential. An approximation has been made in deriving this equation; that the edge of the depletion region is perfectly abrupt. This introduces a large error into the expression for small values of ϕ_S. If the correct form of the charge distribution for the depletion region is used in the derivation, the expression for the width of a single sided junction becomes:

$$W_D = \sqrt{\frac{2\varepsilon_0\varepsilon_{Si}}{q}\left(\phi_S - \frac{kT}{q}\right)\frac{1}{N_A}} \tag{6.6}$$

Justification for this is given by Sze [6.9] for the double sided junction in which the correction is $2kT/q$.

By combining equations (6.4) and (6.6) and re-arranging the resulting expression we can obtain equation (6.7). This is the value of the surface potential when Q_{FG} is insufficient to create an inversion layer.

$$\phi_S = \frac{Q_{FG}^{\ 2}}{2\varepsilon_0\varepsilon_{Si}qN_AWL} + \frac{kT}{q} \tag{6.7}$$

For the purposes of the current investigation, we wish to calculate the surface potential in terms of the floating gate potential. We therefore re-arrange equation 6.3 and substitute for Q_{FG} in 6.7 to give:

$$\phi_s = \frac{C_{ox}^{\ 2}(V_{FG} - \phi_s)^2}{2\varepsilon_0\varepsilon_{Si}qN_AWL} + \frac{kT}{q} \tag{6.8}$$

This equation can be re-arranged into quadratic form and solved for ϕ_s giving:

$$\phi_S = V_{GS} - \eta\left[\sqrt{1 + \frac{2}{\eta}\left(V_{GS} - \frac{2kT}{q}\right)} - 1\right] \tag{6.9}$$

where

$$\eta = \frac{2\varepsilon_0 \varepsilon_{Si} q N_A T_{ox}^{\ 2}}{\left(\varepsilon_0 \varepsilon_{SiO_2}\right)^2} \tag{6.10}$$

Equation 6.9 is similar to one derived by Sze [6.9] which is based on channel charge distribution analysis described by Brews [6.10]. This model of conduction processes in the MOS transistor assumes that the inversion layer charge is contained in a sheet of zero thickness at the semiconductor surface (*the charge sheet model*). This is a useful assumption as it removes the complication in the solution of Poisson's equation that the charge distribution is non-uniform in the semiconductor once the inversion layer has formed. The zero thickness charge sheet cannot have a potential drop across it, and so Poisson's equation simplifies to that for a uniform charge distribution with a single discontinuity. The solution of the relevant equations from Sze and Brews work results in an equation for the surface potential which is very similar to equation 6.9.

We will adopt the charge sheet approximation for our analysis, and so the equation for the surface potential derived above is valid for the range of gate voltages from 0V to the device threshold voltage, V_T.

The surface potential reaches a maximum value at 'strong inversion', the point at which the surface is as strongly *n*-type as the bulk is *p*-type. This occurs when

$$\phi_S = 2\phi_F \tag{6.11}$$

where

$$\phi_F = \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right) \tag{6.12}$$

It is reasonable to assume that ϕ_S will not increase significantly past $2\phi_F$ since at this point even a very small increase in the surface potential will result in a large increase in the surface charge density. Therefore, the depletion region width will not increase further and so ϕ_S must have reached a maximum. Substitution of the equations for the surface potential (equations 6.7 to 6.11) into equation 6.3 allows the prediction of the floating gate voltage for a given floating gate charge with zero drain bias, so that the quiescent operating point for the device can be determined.

6.2.2 The Floating Gate Potential with Applied Drain Bias.

In Section 6.2.1 we considered the relationship between the gate charge and the gate-substrate potential with zero drain bias. When a drain-substrate potential is applied, the gate-drain capacitance will couple changes in the drain potential to the gate causing the floating

gate potential to vary. We will now consider this variation in detail.

Figure 6.6(a) shows the main elements of capacitive coupling in the MOS transistor with a simple electrical model given in figure 6.6(b). Consider the charge distributions within the device. The total charge on the floating gate of the device is fixed, but the distribution of this charge on the electrodes of the various capacitances it serves is variable. A positive increase in the drain potential will cause a change in the voltage across the gate-drain overlap capacitance and negative charge will accumulate on the gate electrode in the overlap region. As the total gate charge is fixed there must be a corresponding decrease in the negative charge density elsewhere on the gate. This is equivalent to an increase in the positive charge density over the source overlap and channel regions. An increase in the positive charge on a fixed capacitor (i.e. the gate-source overlap capacitance) must cause an increase in the voltage across the capacitor (as V=Q/C) so that the gate potential will increase as the drain potential is increased.

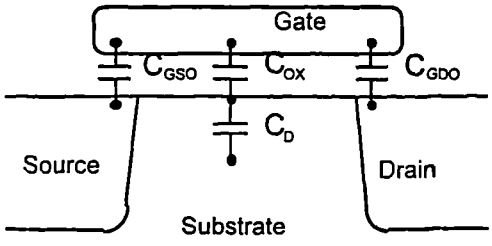


Figure 6.6 (a) Capacitances in the MOS transistor.

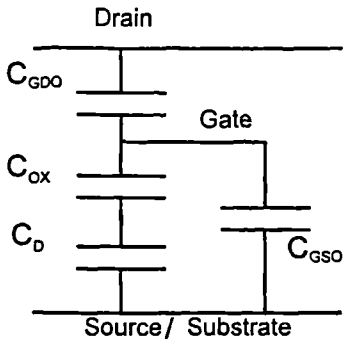


Figure 6.6 (b) Electrical model of capacitance in the MOS transistor.

A further change also occurs in the device when V_{DS} is increased. As noted, there is an increase in the positive charge over the channel region. This will induce a negative charge in the channel which will increase the conduction in the device, which is consistent with the increase in the gate potential already described. The above argument also holds for a decrease in the drain potential which results in a decrease in the gate voltage. Hence, any changes in the drain potential are coupled to the gate causing variations in the gate potential. The transistor will then react as if a changing potential was being applied directly to the gate.

It is interesting to note that the changes in the gate potential can cause a device which

was initially non-conducting to enter the conducting state i.e. the device can be switched on with *no net charge* on the gate. This occurs because the free carriers within the gate are separated by the electric field distributions in the structure leading to localised concentrations of positive or negative charge on the various capacitor plates. The total net charge of course does not change.

The change in gate potential can be quantified by considering the circuit given in figure 6.6(b). Considering this as a simple capacitive divider network with the drain potential as the input and the gate potential as the output we can obtain:

$$V_{GS} = V_{DS} \frac{C_{GDO}}{C_{GDO} + C_{GSO} + \dfrac{C_{OX}C_D}{C_{OX} + C_D}} \tag{6.13}$$

where C_{GDO} and C_{GSO} are the overlap capacitances, C_{OX} is the gate oxide capacitance over the channel and C_D is the capacitance of the bulk depletion region beneath the channel.

The equation was derived independently by the author before locating [6.3] in which Frohman-Bentchkowsky gives a similar equation, except that C_{GB} is replaced by the series combination of C_{OX} and C_D and only the overlap capacitances have been included (i.e. the capacitance contribution from the channel itself has been ignored). Although equation 6.13 still has several shortcomings as an accurate model of the transistor's behaviour, it does give the basic form of the variation of V_{GS} with V_{DS}.

Figure 6.7 shows graphs of I_{DS} against V_{DS} for charged and uncharged floating gate transistors. The gate voltage is calculated for each value of V_{DS} using equation 6.13, and is used in the simple first order equation given in equations 6.14 (non-saturated operation) and 6.15 (saturated operation) to calculate I_{DS}

$$I_{DS} = \frac{\mu C_{OX} W}{L}\left[\left(V_{GS} - V_{T_0}\right)V_{DS} - \frac{V_{DS}^{\,2}}{2}\right] \tag{6.14}$$

Non-saturated operation

$$I_{DS} = \frac{\mu C_{OX} W}{L}\frac{\left(V_{GS} - V_{T_0}\right)^2}{2} \tag{6.15}$$

Saturated operation

The first order current equations are adequate to show the form of the graphs since the simplistic analysis used is inadequate in several ways. The analysis does not account for

several factors including the coupling of the channel potential to the gate, its variation with V_{DS} or the variation of the depletion capacitance, C_D with and V_{DS}. In the next section we will consider these factors and include them into the equations for the device operation.

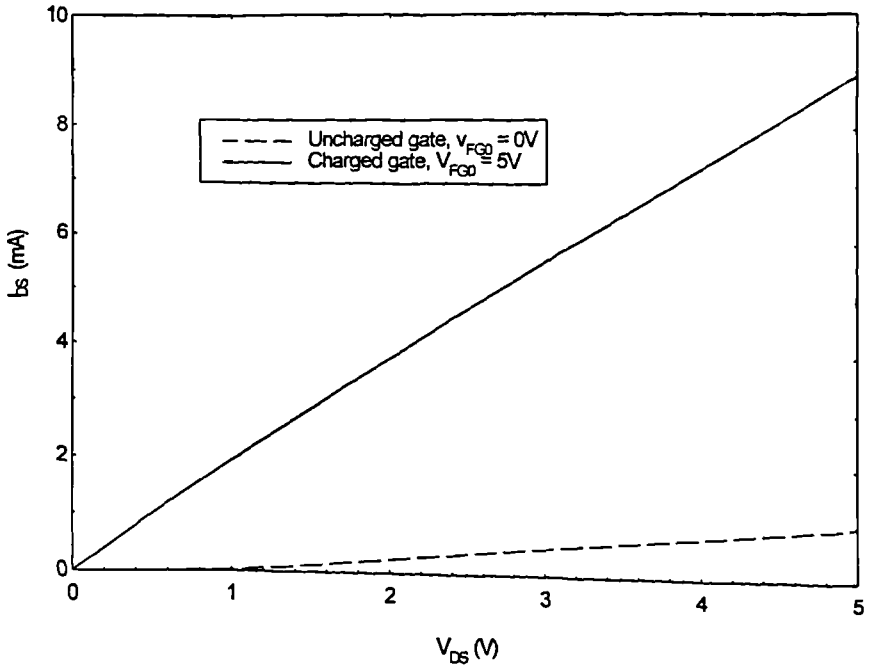


Figure 6.7 I_{DS} - V_{DS} characteristics for charged and uncharged floating gate MOS transistors calculated using equations 6.13, 6.14, 6.15.

6.3 Refinements of the Model

To understand the refinement of the model it is first necessary to consider the distribution of charge in the MOS transistor during its operation. We will consider a floating gate transistor with a total gate charge Q_{FG} and initial V_{DS} = 0V. The source will be connected to the substrate. The charge distributions to be considered are as follows.

i) Charge associated with the gate-drain overlap capacitance, C_{GDO}.

ii) Charge associated with the gate-source overlap capacitance, C_{GSO}.

iii) Mobile charges on the gate over the channel region. This distribution varies along the channel.

iv) Charge within the surface inversion layer, which varies along the channel.

v) Depletion region charge beneath the channel.

It is the precise configuration of the charges that gives rise to the conduction and other parameters of the transistor.

The total charge on the floating gate of the transistor is fixed, but its distribution is variable. The distribution of the charge, $Q_{FG,}$ with V_{DS}=0V, is shown in figure 6.8.

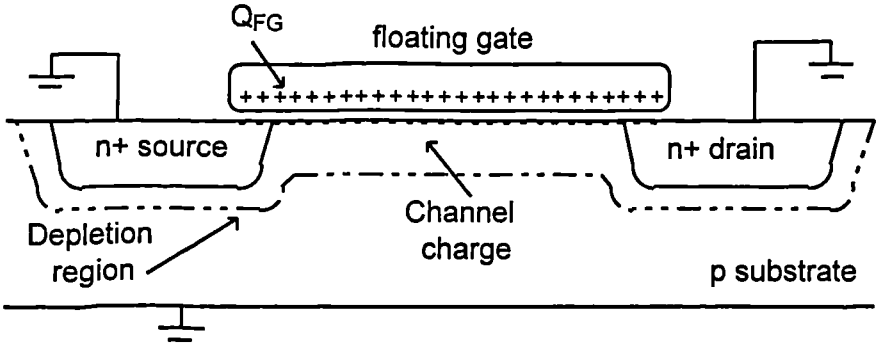


Figure 6.8 Initial charge distributions

For simplicity, Q_{FG} is assumed initially to be sufficient to induce a strongly inverted channel region in the *p*-type substrate. With V_{DS} = 0, the charge distribution in the substrate is uniform along the channel and overlap regions.

Figure 6.9 illustrates the effect of increasing V_{DS} on the transistor charge distributions. (For clarity the charge associated with the depletion region capacitance is not shown). The increasing drain to gate potential causes an increase in the positive charge on the drain terminal of $C_{GDO.}$ An equal negative charge is induced on the gate terminal of this capacitor and, since the total gate charge is fixed, this results in a reduction in the negative charge over the channel region and on the gate terminal of C_{GSO}. The subsequent increase in the negative charge in the channel region is provided by mobile charge from the drain and source regions resulting in increased conduction through the device. However, the increased conduction itself increases the channel potential decreasing the negative charge at the drain end of the channel and cancelling, to some extent, the effect of the increase in the drain voltage.

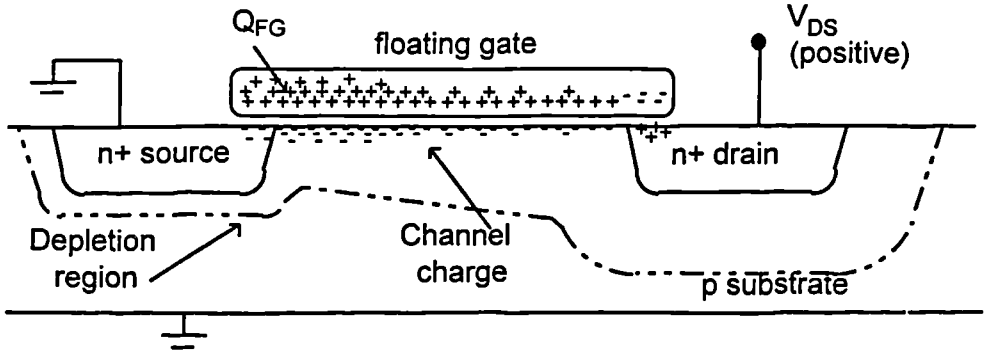


Figure 6.9 Charge movements in the floating gate MOS transistor

Another effect of the drain potential on the gate, that must also be included for a full

model of the device operation comes from the coupling of ~~channel region~~, the channel potential through the gate oxide capacitance. The potential varies along the channel and the distribution is determined by the channel current and V_{GS}. The variation of the channel potential also affects the depletion region beneath the channel, increasing its depth to enclose the extra charge that is required.

The interactions of the charges and potentials within the device are complex. All the charge movements described above affect the gate potential which, in turn, affects the charge distributions. The latter are also affected by V_{DS} and the channel potential which is itself dependent on V_{DS} and the gate potential V_{FG}. The interdependence of the variables in the floating gate device is summarised in Table 6.1.

Variable	Depends on
V_{FG}, the gate potential	V_{DS}, Q_{FG0}, C_D
C_D, the depletion charge	V_C, V_{FG}
$V_C(x)$, the channel potential	x, V_{DS}, V_{FG}
$Q_B(x)$, the depletion charge	V_C
$Q_C(x)$, the channel charge	x, V_{DS}, V_{FG}
$Q_G(x)$, the gate charge	V_{DS}, V_{FG}, V_C

Table 6.1 Variable dependency in the floating gate MOS transistor

The complex interactions of the charges and potentials within the floating gate transistor are expressed by device equations that cannot be solved analytically. Numerical techniques have therefore been used to solve the current and voltage equations for the floating gate MOS transistor enabling the gate voltage to be predicted as a function of the total gate charge and the drain-source potential.

6.4 Analysis of the Floating Gate MOS Transistor Operation

From the description in Section 6.3 it is apparent that to calculate the floating gate potential for a given value of V_{DS} it is necessary to know the variation of the potential along the device channel. This is determined by the conductance of the transistor and the drain-source potential. The conductance is dependent on the amount of coupling of the drain and channel potentials to the gate which is in turn determined by the capacitance values of the

device. The depletion region capacitance is determined by the channel potential. Clearly the channel potential is an important starting point for the calculation of device conductance.

In order to describe the operation of the floating gate transistor, the equations describing the charge distributions within the device will be derived. This will enable the potential of floating gate to be determined and hence the conductance for given Q_{FG} and V_{DS} values can be calculated and device performance predicted.

Fundamentally the transistor works in the same way as a fault free device and many of the equations are similar to those used in the normal second-order model of the MOST. However, the essential difference between the floating gate and the connected gate device is that it is the *gate charge* and not the *gate potential* that remains fixed during operation. It will be assumed that the device has no gate-source or gate-drain overlap for the initial analysis.

6.4.1 Floating Gate Transistor Charge Distributions.

The starting point for the analysis of the floating gate transistor operation is to establish equations for the charge distributions in the device. These are fundamental quantities which can be used to calculate many other parameters. The three charge distributions that are significant, Q_B in the bulk depletion region, Q_S in the surface channel region and Q_G at the gate-oxide interface are shown in figure 6.10. All of the charges vary along the channel and expressions will now be derived for each of them.

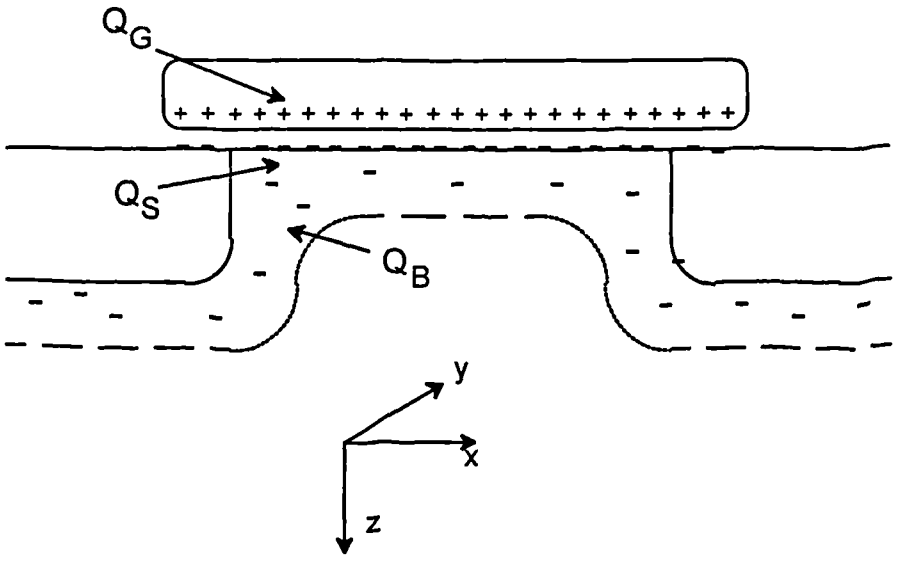


Figure 6.10 Charges in the MOS transistor

i) Bulk depletion region charge, Q_B.

The distribution of the charge contained within the bulk depletion region is determined by the channel potential distribution, V_x. We assume a rectangular distribution for the depletion charge into the semiconductor (i.e. $Q_B(z)$ = const. for $0 < z < W_d$). As suggested in section 6.2.1, the width of the depletion region, and hence the charge distribution, can be calculated using the single sided *p-n* junction approximation. For the depletion region width at position *x* along the channel we have:

$$W_D(x) = \sqrt{\frac{2\varepsilon_0\varepsilon_{Si}}{qN_A}(\phi_S + V_X)} \tag{6.16}$$

The charge contained within a volume $dx.W.W_D$ (where dx is an incremental length along the channel, W is the device width) is given by:

$$Q_B = qN_A dxWW_D(x) \tag{6.17}$$

giving

$$Q_B(x) = \sqrt{2\varepsilon_0\varepsilon_{Si}qN_A(\phi_S + V_X)} \tag{6.18}$$

The area factor, $dx.W$, has been dropped for clarity in this and subsequent equations. Q_B, Q_S and Q_G are therefore strictly surface charge densities.

ii) Gate-oxide interface charge, Q_G.

The charge on the gate can be determined by considering the potential across the simple parallel plate capacitor consisting of the gate electrode and the semiconductor surface. The potential across this capacitor is V_{GS} - (ϕ_S+V_X) and so the charge on this capacitor is given by :

$$Q_G(x) = C_{OX}\left(V_{GS} - (\phi_S + V_X)\right) \tag{6.19}$$

iii) Surface channel charge, Q_S.

The charge at the channel surface is simply the difference between the gate charge and the bulk charge at any point along the channel. The channel charge at *x* is therefore given by:

$$Q_S(x) = Q_G(x) - Q_B(x) \tag{6.20}$$

Expanding we have:

$$Q_S(x) = C_{OX}\left(V_{GS} - (\phi_S + V_x)\right) - \sqrt{2\varepsilon_0\varepsilon_{Si}qN_A(\phi_S + V_x)} \tag{6.21}$$

which can be rewritten as:

$$Q_S(x) = C_{OX}\left(V_{GS} - (\phi_S + V_x) - \gamma(\phi_S + V_x)^{\frac{1}{2}}\right) \tag{6.22}$$

where

$$\gamma = \frac{(2\varepsilon_0 \varepsilon_{Si} q N_A)^{\frac{1}{2}}}{C_{OX}} \tag{6.23}$$

The distributions described by equations 6.18, 6.19 and 6.22 are valid for a normal MOS transistor or for a floating gate device. However, for the floating gate device, we have the constraint that the total gate charge is fixed as shown in equation 6.24.

$$\int_0^L Q_G(x)\,\mathrm{dx} = Q_{FG} \tag{6.24}$$

By substituting the expression for $Q_G(x)$ from equation 6.19 we obtain:

$$C_{OX}\int_0^L \left(V_{FG} - (\phi_S + V_x)\right)\mathrm{d}\,x = Q_{FG} \tag{6.25}$$

From this equation (in which V_{GS} has now been called V_{FG}) we can seen that as Q_{FG}, C_{OX} and ϕ_S are known quantities (or can be calculated from device parameters), the floating gate potential can be found if we are able to determine the distribution of the channel potential, V_x.

6.4.2 Channel Potential Distribution.

It has been shown in equation 6.25 that the floating gate potential is dependent on the channel potential distribution and that if this can be obtained, V_{FG} can be calculated. It is therefore necessary to obtain an expression for the variation of the potential along the channel. A method of achieving this, similar to the normal MOS current derivation, is now described.

Conduction in the channel of an MOS transistor operating out of saturation, occurs by the drift of carriers in the electric field along the channel. If the mobility of the electrons in the channel is μ, the carrier velocity in the field E_x is $\mu.E_x$. The electric field strength is given by the gradient of the potential, and so for a channel of width W, the conventional current at any point in the channel is given by:

$$I_{DS} = \mu W Q_S \frac{dV_x}{dx} \tag{6.26}$$

This equation can be used to derive the full second order I_{DS} equation. Hence we obtain:

$$I_{DS} = \mu C_{OX} \frac{W}{L}\left[V_{DS}\left(V_{GS} - V_{T_0} + \gamma\phi_S^{\frac{1}{2}}\right) - \frac{2}{3}\gamma\left(\left(\phi_S + V_{DS}\right)^{\frac{3}{2}} - \phi_S^{\frac{3}{2}}\right) - \frac{V_{DS}^{\ 2}}{2}\right] \tag{6.27}$$

By equating equations 6.26 and 6.27 and rearranging the resulting expression we can obtain an equation for V_x. Hence:

$$\mu C_{OX} \frac{W}{L}\left[V_{DS}\left(V_{GS} - V_{T_0} + \gamma\phi_S^{\frac{1}{2}}\right) - \frac{2}{3}\gamma\left(\left(\phi_S + V_{DS}\right)^{\frac{3}{2}} - \phi_S^{\frac{3}{2}}\right) - \frac{V_{DS}^{\ 2}}{2}\right] = \mu W Q_S \frac{\mathrm{dV_x}}{\mathrm{dx}} \tag{6.28}$$

giving:

$$\frac{C_{OX}}{L}\left[V_{DS}\left(V_{GS} - V_{T_0} + \gamma\phi_S^{\frac{1}{2}}\right) - \frac{2}{3}\gamma\left(\left(\phi_S + V_{DS}\right)^{\frac{3}{2}} - \phi_S^{\frac{3}{2}}\right) - \frac{V_{DS}^{\ 2}}{2}\right]\mathrm{dx} = Q_S\,\mathrm{dV_x} \tag{6.29}$$

If we now substitute the expression for Q_S given in equation 6.22 into the right hand side, the resulting equation can be integrated over V_x:

$$\int Q_S(x)\,\mathrm{dV_x} = C_{OX}\int_0^{V_{DS}}\left(V_{GS} - \left(\phi_S + V_x\right) + \gamma\left(\phi_S + V_x\right)^{\frac{1}{2}}\right)dV_x \tag{6.30}$$

Performing the integration the right hand side of 6.29 becomes:

$$\mathrm{RHS} = C_{OX}\left[V_x\left(V_{GS} - \phi_S\right) - \frac{V_x^{\ 2}}{2} - \frac{2}{3}\gamma\left[\left(\phi_S + V_x\right)^{\frac{3}{2}} - \phi_S^{\frac{3}{2}}\right]\right] \tag{6.31}$$

Similarly the left hand side can be integrated over x to give:

$$\begin{aligned}\mathrm{LHS} &= \int \frac{C_{OX}}{L}\left[V_{DS}\left(V_{GS} - V_{T_0} + \gamma\phi_S^{\frac{1}{2}}\right) - \frac{2}{3}\gamma\left(\left(\phi_S + V_{DS}\right)^{\frac{3}{2}} - \phi_S^{\frac{3}{2}}\right) - \frac{V_{DS}^{\ 2}}{2}\right]\mathrm{dx} \\ &= C_{OX}\frac{x}{L}\left[V_{DS}\left(V_{GS} - V_{T_0} + \gamma\phi_S^{\frac{1}{2}}\right) - \frac{2}{3}\gamma\left(\left(\phi_S + V_{DS}\right)^{\frac{3}{2}} - \phi_S^{\frac{3}{2}}\right) - \frac{V_{DS}^{\ 2}}{2}\right]\end{aligned} \tag{6.32}$$

By equating 6.31 and 6.32 and rearranging the expression we can obtain an equation in V_x with known quantities, equation 6.33, so that the V_x distribution can be found.

$$\frac{V_x^2}{2}+\frac{2}{3}\gamma\left[(\phi_S+V_x)^{\frac{3}{2}}-\phi_S^{\frac{3}{2}}\right]-V_x(V_{GS}-\phi_S) + \frac{x}{L}\left[V_{DS}\left(V_{GS}-V_{T_0}+\gamma\phi_S^{\frac{1}{2}}\right)-\frac{2}{3}\gamma\left((\phi_S+V_{DS})^{\frac{3}{2}}-\phi_S^{\frac{3}{2}}\right)-\frac{V_{DS}^2}{2}\right]=0 \quad (6.33)$$

Equation 6.33 applies to a normal MOS transistor and to a floating gate device. For the case of the normal transistor all quantities in the expression are known and so it should be possible to find the distribution of V_x. In fact, there does not appear to be an analytic solution to this equation because of the term $(\phi_S + V_x)^{3/2}$. The situation is even more complicated for the floating gate transistor since the floating gate potential is not known: it is the quantity that we are trying to determine. The lack of a simple solution to equation 6.33 means that we cannot substitute an expression for V_x into equation 6.25 as intended. The solution to this problem lies in using a numeric technique to solve equation 6.33 for V_x, and this is described later.

If it is assumed that the channel potential can be calculated from the drain (or more precisely, the pinch-off point) to the source, then all other dependent variables can be calculated and V_{GS} can be found. The method that has been used will now be described.

6.5 An Algorithm for the Calculation of the Floating Gate Voltage

The aim of this algorithm is to use the equations derived in the previous section to find the gate-source potential of a floating gate MOS transistor, given the total gate charge and the drain-source potential. The algorithm is illustrated in the flow diagram, figure 6.11. (The roman numerals on this diagram refers to the following sub-sections). As can be seen from the diagram, an iterative method is used for the calculation of V_{GS}. This is necessary because of the complex nature of the interaction of the charges and potential around the device described in the previous section. The main stages of the algorithm used for the calculation of the gate-source voltage of a floating gate MOS transistor will now be described.

i. Initial Estimate of the Floating Gate Voltage

The algorithm uses an estimator function for V_{GS} which is based on the simple capacitve coupling model presented in section 6.2. The function used is

$$V_{GS} = V_{FG_0} + \alpha V_{DS} \quad (6.34)$$

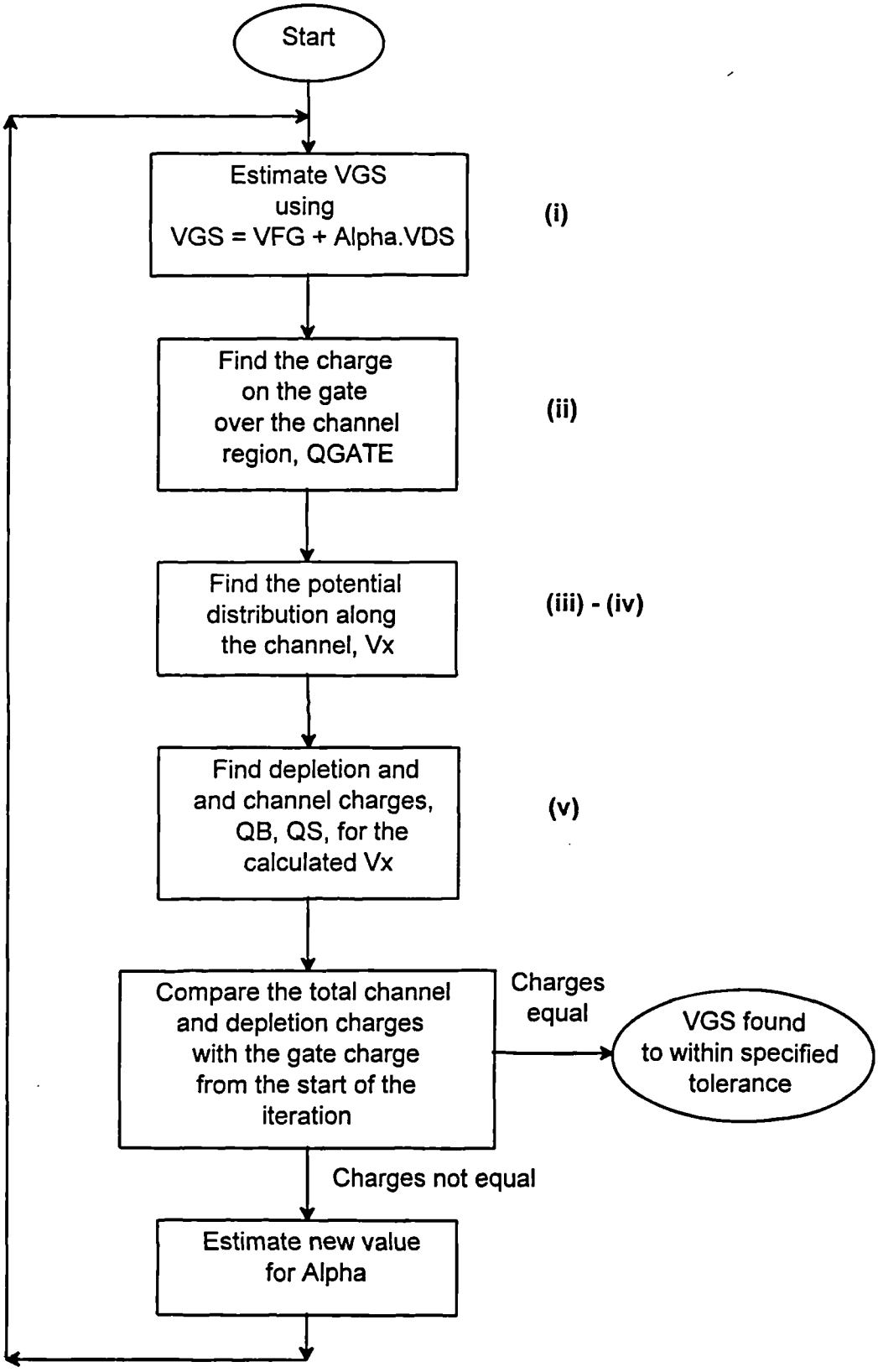


Figure 6.11 Flow diagram of the V_{GS} algorithm

The parameter α is related to the capacitive coupling of the drain and channel potentials to the gate. It is important to note that α (which is equivalent to β in Renovell's equation) is not a constant as suggested by Renovell and Henderson, but varies with V_{DS}. This is because of the variation of the depletion capacitance below the channel region, which varies with the channel potential.

As stated previously, α is related to the capacitive coupling of the drain and channel potentials to the gate. The initial value of α is taken as 0.5 and V_{GS} is set to $V_{FG0} + \alpha.V_{DS}$ according to equation 6.31.

ii. Find the charge on the gate over the channel region

The model derived in the previous section did not account for the drain and source overlap regions, as it was not necessary to do so. However, for calculations involving real devices we must account for the charge stored in these capacitances. The charge on the gate over the channel region for this gate potential is therefore calculated by subtracting the charges associated with the drain and source overlap capacitances from the total gate charge. The remaining gate charge is used at the end of the iteration, to compare with the sum of the depletion and channel charges for the calculated V_x distribution resulting from the estimate of V_{GS}. If the two charges are within a specified tolerance, the gate voltage has been determined and the iteration is complete.

iii. Calculation of the channel potential in the inverted region

The channel potential distribution can now be calculated based on the estimate for V_{GS}. For the calculation of charges and potentials in the device, the channel is segmented into a number of short elements over which the changes in channel potential and charges are assumed to be insignificant. This is shown in figure 6.12.

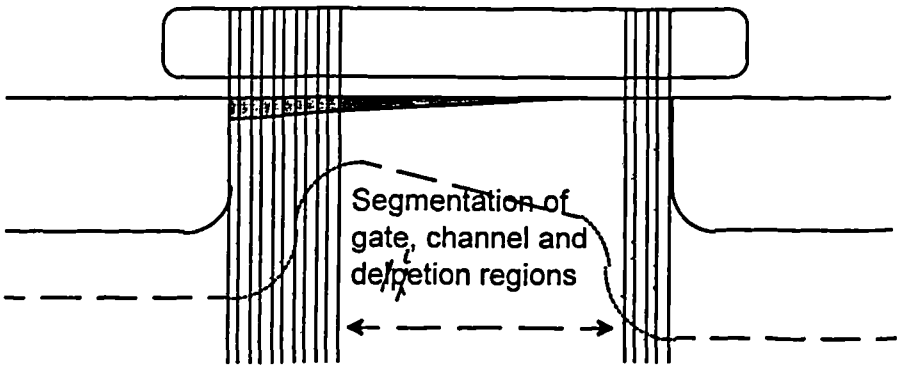


Figure 6.12 Segmentation of the channel for voltage and charge calculation

A set of discrete values for V_x is then calculated using equation 6.33. Once again an

iteration process is used to find each value of V_x, solving equation 6.33 for iterated values of V_x until the resulting error is less than a specified relative tolerance. Interval halving is used for this iteration process.

iv. The channel potential in the pinch-off region

Equation 6.33 is not used for the calculation of V_x in the pinched off region of the channel. To consider the form of the V_x distribution in this region, several parameters are required. The channel potential at the pinch off point, V_{DSsat}, can be determined by considering the channel charge. At pinch off the channel charge is zero and so we have from equation 6.22:

$$Q_S(x) = C_{OX}\left(V_{GS} - V_{FB} - (\phi_S + V_x) - \gamma(\phi_S + V_x)^{\frac{1}{2}}\right) = 0 \qquad (6.35)$$

At the pinch off point x_p, V_x=V_{DSsat}. This can be substituted into the above equation to give:

$$C_{OX}\left(V_{GS} - V_{FB} - (\phi_S + V_{DS_{Sat}}) - \gamma(\phi_S + V_{DS_{Sat}})^{\frac{1}{2}}\right) = 0 \qquad (6.36)$$

If we now rearrange the equation and square both sides the resulting quadratic can be solved for V_{DSsat}. Hence,

$$V_{DS_{Sat}}{}^2 + (2\phi_S - 2(V_{GS} - V_{FB}) - \gamma^2)V_{DS_{Sat}} + \left((V_{GS} - V_{FB})^2 - 2\phi_S V_{GS} + \phi_S{}^2 - \gamma^2\phi_S\right) = 0 \qquad (6.37)$$

Now using the equation for finding the roots of a quadratic and simplifying we have:

$$V_{DS_{Sat}} = V_{GS} - V_{FB} - \phi_S + \frac{\gamma^2}{2}\left(1 \pm \sqrt{1 + \frac{4(V_{GS} - V_{FB})}{\gamma^2}}\right) \qquad (6.38)$$

(Note that the solution which adds the final square root term does not produce sensible values for V_{GS}.)

Equation 6.38 agrees with an expression derived by Grove [6.8] but it is more accurate than the simple expression that is normally used, i.e. V_{DSsat}=V_{GS} - V_T.

We now need to determine the length of the pinch off region, L_p. A simple way to calculate this is to consider the region as a one sided pn junction depletion region. This approximation was first proposed by Reddi and Sah [6.12]. The potential maintaining the depletion region is V_{DS} - V_{DSsat} and by substituting this into the equation for the depletion

region width we can obtain:

$$W_D = \sqrt{\frac{2\varepsilon_0\varepsilon_{Si}}{qN_A}\left(V_{DS} - V_{DS_{Sat}}\right)} = L_P \tag{6.39}$$

A more accurate derivation of this parameter is given by Frohman-Bentchkowsky and Grove [6.11]. In this paper, the authors consider not only the electric field due to the potential difference between the drain and the pinch-off point, but also the transverse components of the electric fields due to the gate potential.

The equation resulting from this detailed consideration is:

$$\frac{1}{L_P} = \frac{1}{\sqrt{\frac{2\varepsilon_0\varepsilon_{Si}}{qN_A}\left(V_{DS} - V_{DS_{Sat}}\right)}} + \frac{\varepsilon_0\varepsilon_{SiO_2}}{\varepsilon_0\varepsilon_{Si}T_{OX}}\left(\frac{a\left(V_{DS} - V_{GS}\right) + b\left(V_{GS} - V_{DS_{Sat}}\right)}{V_{DS} - V_{DS_{Sat}}}\right) \tag{6.40}$$

The parameters a and b are called field fringing factors and represent the extent to which the gate oxide fields contribute to the transverse field. Values of $a = 0.2$ and $b = 0.6$ are given for a wide range of device parameters. Equation 6.40 was used in the program.

Having established the saturation voltage and length it is now necessary to determine the potential distribution in the pinch off region. The complex form of the electric field in the region implies a complex charge distribution. However, the complexity is confined to the surface region and, for most of the volume occupied by the space charge, the field has a simple structure implying a uniform charge distribution. Hence it is reasonable to ignore the complex nature of the electric field at the surface and assume a uniform electric field distribution throughout the pinch off region. It is therefore assumed that there is a simple linear variation of potential from the pinch off point to the drain junction.

v. *Calculation of channel and depletion region charges*

The result of the calculation of the V_x distribution in the inverted and pinch-off regions of the channel, as described above, is then used to calculate the charge distributions in the channel and depletion regions. Equations 6.18 and 6.19 are used to calculate Q_B and Q_S for each element of the channel. These charge elements are then summed to give the total depletion and channel charges for the estimated V_{GS} value. The sum of these two charge components should equal the gate charge calculated at the start of the iteration and so a comparison of these two values will indicate the accuracy of the estimate of V_{GS}. If the difference is less than a specified tolerance then the solution for V_{GS} for a given V_{DS} has been found. If not, a new value of α is chosen using the interval halving method, and the process

is repeated until a solution is found.

6.6 The Floating Gate MOS Transistor Model Program, FLGMOD.

The algorithm described in section 6.5 has been implemented in a computer program written in Pascal. The program allows the calculation of V_{FG} for a range of V_{DS} values and also allows the calculation of I_{DS}. Various parameters such as the device charges and capacitance ratio, α, can also be investigated. The transistor that is modelled is an *n*-channel floating gate MOS device with both source and substrate connected to 0V.

Many of the process and device parameters used in the calculations can be set by the user. The default values correspond to the process used for the fabrication of the test circuits described in Chapter 3. The default values of all of the parameters are given in Table 6.2.

Parameter	Program name	Value
Process:		
Substrate doping	Na	1.65×10^{22} m^3
Flat band voltage	Vfb	-0.7 V
Electron mobility	mu	0.0558 $m^2V^{-1}S^{-1}$
Gate oxide thickness	Tox	50×10^{-9} m
Device:		
Channel length	Length	3×10^{-6} m
Channel shortening	dL	0.5×10^{-6} m
Channel width	Width	10×10^{-6} m
Channel narrowing	dW	0.636×10^{-6} m
Floating gate voltage (at $V_{DS} = 0V$)	VFG	2.0 V, default

Table 6.2 Default process and device parameters used in the FLGMOD program

A second order model is used for the calculation of I_{DS}. The equations in this model account for the variation of the depletion charge along the channel and with V_{DS}. This variation is an important factor in the floating gate model described in this chapter.

A listing of the program is given in Appendix A.

6.7 Some Basic Results of the Model

The model described in the previous sections and coded into the FLGMOD program predicts the gate potential for specified initial floating gate potential (or charge) and the drain-source potential. From this the drain-source current can be calculated. One of the basic predictions of the theory described in this chapter is that changes in the drain voltage will be coupled to the floating gate. This is indeed the case as shown in figure 6.13 which shows the variation of V_{FG} with V_{DS} for a typical device: W = 8μm, L = 3μm, V_{FG0} = 1.3V. These parameters correspond to the device **b**-112/B in the test set and the measured floating gate voltage for this device is also given on the graph. There is good agreement between the model and the measured data.

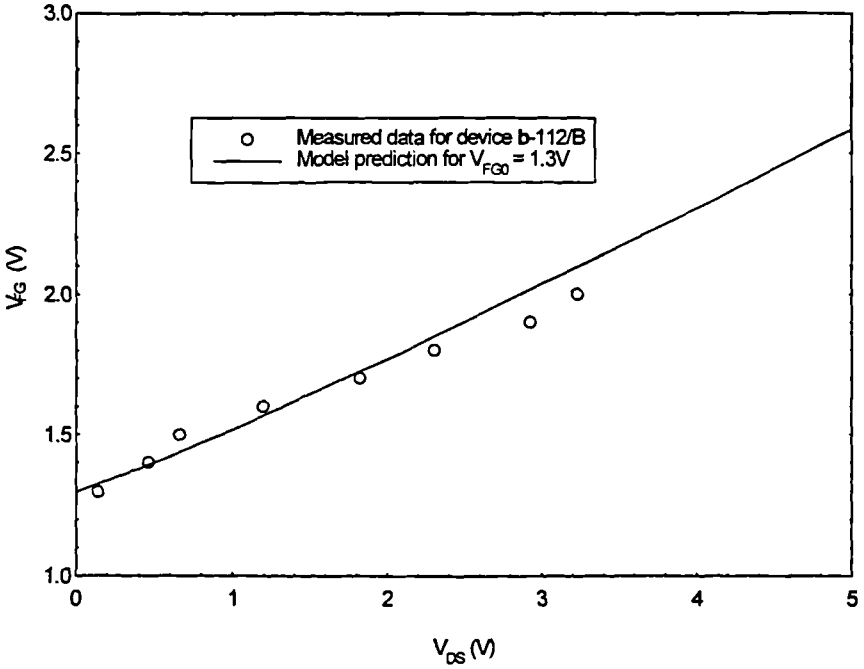


Figure 6.13 Variation of V_{FG} with V_{DS}

Other examples do not provide quite such a good correlation between ~~of~~ experimental data and the predictions of the model. This can be attributed to the variations observed in the coupling ratio, α, for devices that are nominally very similar. A possible source of this variation is the gate-drain overlap length. This parameter is seen to have a significant effect on α when it is changed in the program.

The theoretical value of α for the example given is approximately 0.25 which means that a change in V_{DS} from 0 to 5V will increase the gate voltage by around 1.25V. Consequently, even if the gate has no residual charge, the transistor can become conducting as demonstrated by Frohman-Bentchkowsky and similar cases observed in the measurements described in Chapter 5.

Careful examination of the theoretical curve in figure 6.13 reveals that the slope, and therefore α, is not constant. This was predicted early in this chapter and is due to changing capacitances in the device as the charge distributions change. The principle change that is predicted by the theory is that as the channel potential rises, the depletion capacitance decreases. This causes the coupling ratio to increase as it is inversely proportional to the depletion capacitance. This variation is shown by the lower curve on the graph in figure 6.14 which gives the theoretical value ofα for device **b**-112/B for which V_{FG0} = 1.3V.

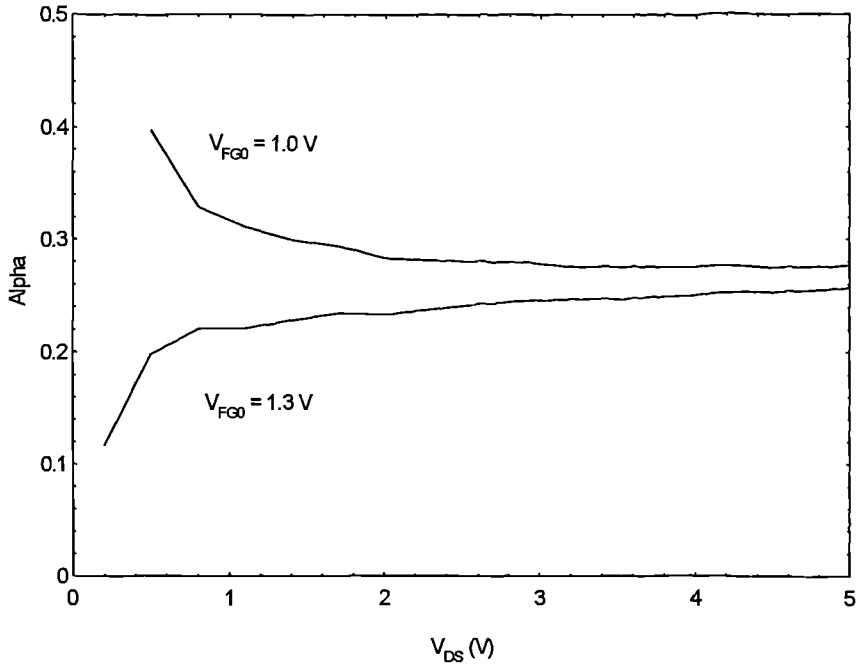


Figure 6.14 Theoretical prediction of the variation ofα with V_{DS} for device b-112/B with V_{FG0} = 1.3 and 1.0 V

The upper curve on this graph shows the predicted value ofα for device **b**-112/B with V_{FG0} = 1.0 V. Although this curve seems to contradict the theory outlined above, it follwos the form of the variation ofα that was observed in experiment for several devices: i.i. a high value for low V_{DS} which decreases as V_{DS} increases. Higher values of V_{FG0} give less variation inα.

Although the correlation between measurements and theory is not as good as might be hoped for this parameter, the model does seem to reflect the variability ofα quite well. The description of the capacitances in the MOS transistor previously presented is a simplification of a complex situation, and changes in the device capacitances and fields, particularly around pinch-off, are complex. Despite this, the model appears to predict the device performance quite well. Generally the value ofα predicted by the model is in good agreement with the measured values and some of the variations inα predicted by the model are observed in actual devices.

The drain-source current for the example device is shown in figure 6.15 with the measured characteristic superimposed. There is reasonable agreement between the shape of the two curves. The most likely cause for the lower current predicted by the model is incorrect device parameters for the device used in the example. Very slight changes in the gate-drain overlap length or the intial floating gate voltage cause significant changes in the drina-source current. This is illustrated in the two graphs shon in figure 6.16 which gives the I_{DS} - V_{DS} characteristics for the example device (**b**-112/B) for various V_{FG0} and gate-drain overlap lengths, ΔL. The variation in I_{DS} caused by ΔL is quite significant and this supports the suggestion made earlier that the variation in α observed for nominally identical devices may be due to variations in ΔL between these.

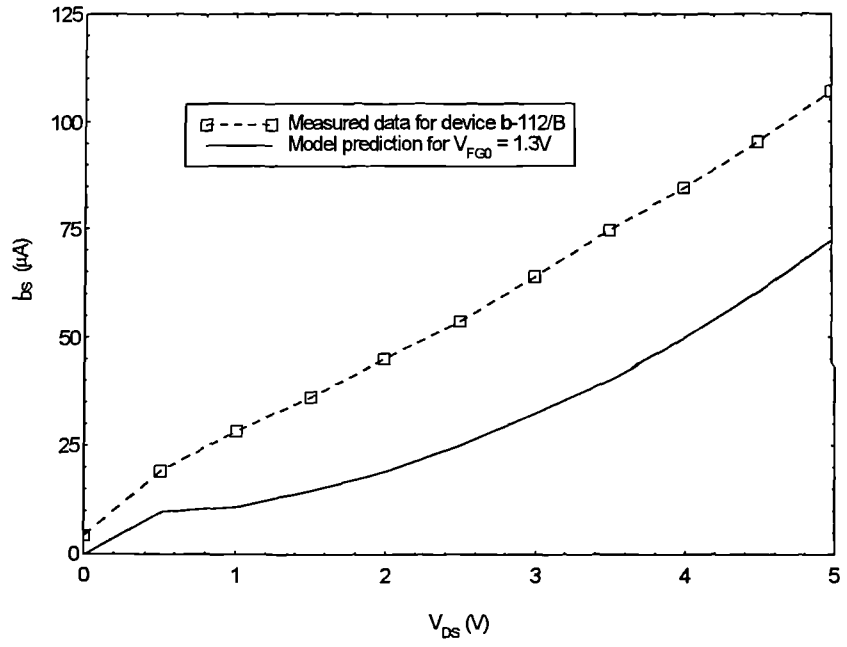


Figure 6.15 I_{DS} - V_{DS} characteristics for floating gate MOS transistor from measured data and model prediction

Generally the model appears to represent the operation of the floating gate MOS transistor quite well and there is reasonable agreement with many of the measured characteristics.

6.8 Chapter Summary

In this chapter we have considered the operation of an MOS transistor with a floating gate. Existing models have been described and found to be inadequate for accurate prediction of the floating gate potential. The operation of such a device has been considered and equations derived for the charge and potential distributions in the device. A numerical technique for the solution of the resulting equations has been presented and some example

results of calculations using the model have been given.

The floating gate transistor has been shown to have a wide range of current values determined by the quiescent gate potential (with zero drain bias) and the drain-source potential. The effect of this variation on circuit performance will be investigated in the next chapter.

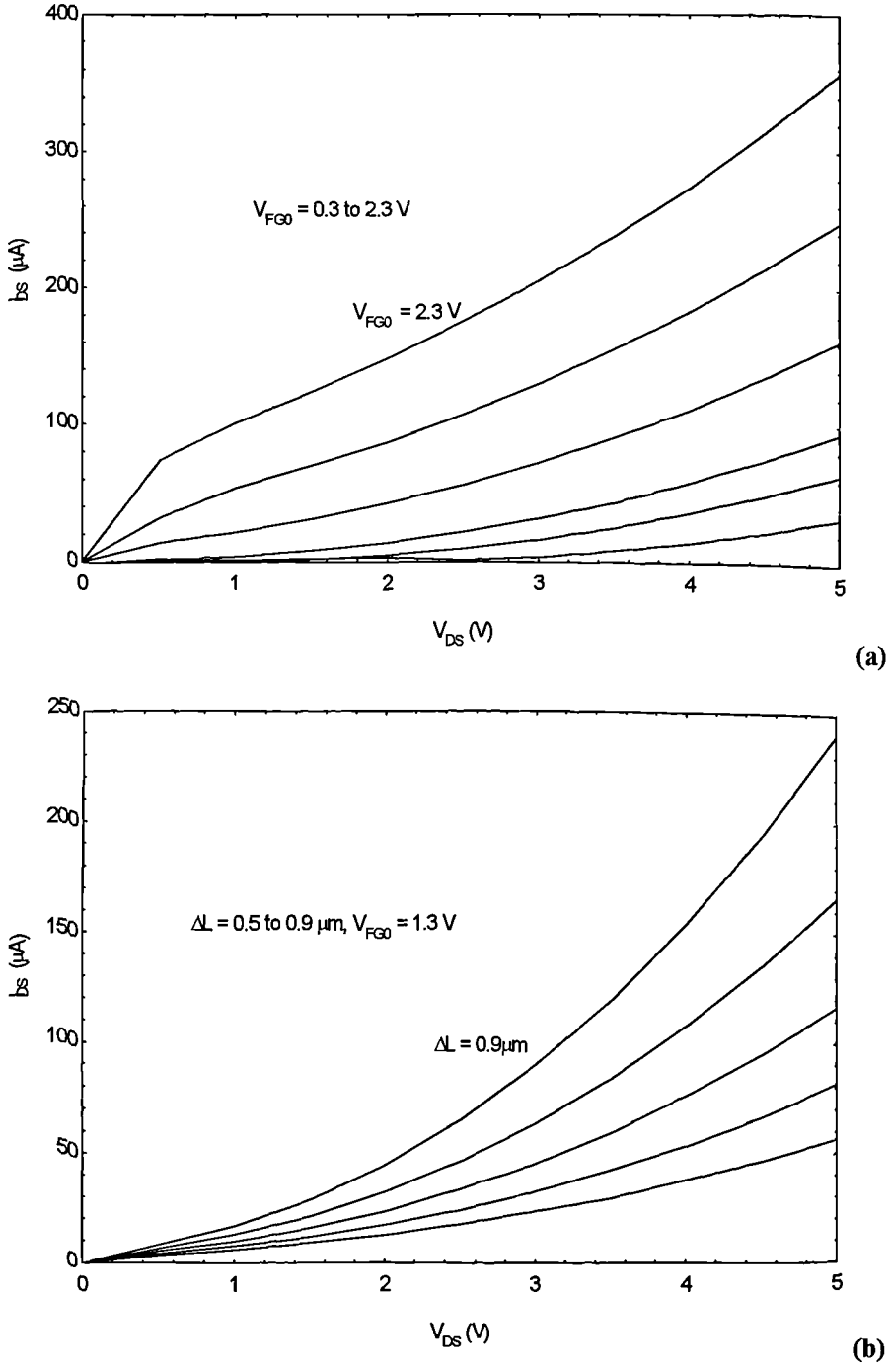


Figure 6.16 I_{DS} - V_{DS} characteristics for floating gate device b-112/B with (a) varying V_{FG0}, (b) varying ΔL

Chapter 7

CMOS Circuit Performance and Testing for Floating Gate Faults

In the previous chapters of this thesis, we have established the characteristics of the floating gate MOS transistor. The importance of the fault was established in Chapter 3. It was shown that a large proportion of photolithographic defects can lead to isolated gate nodes. The experimental analysis of faulty transistors, reported in Chapter 5, has provided data on the parameters and characteristics of the device. This has given some insight into the operation of floating gate transistors. Theoretical analysis of the operation of MOS transistors with fixed gate charge has produced a model for the operation of the device. This model has been verified by comparison with experimental results and it allows the prediction of floating gate transistor (FGT) characteristics for devices with parameters, such as gate charge, which are not available through experimental analysis.

To complete the investigation into the effects of the floating gate fault on CMOS IC's, the data and models generated must now be applied to CMOS logic circuits. In this chapter, we will investigate the operation of logic circuits containing floating gate faults. The circuit analysis has been aided by the use of circuit simulation. An accurate SPICE model of the floating gate fault is described and its application to a range of circuits is discussed. From the analysis and simulations presented, we are able to draw some general conclusions on the effects of floating gate faults in CMOS IC's.

The results of the simulations allow testing methods to be investigated. The adequacy of standard testing techniques for the detection of floating gate faults is considered and improved testing methodologies are suggested.

7.1 Circuit Simulation Models for the Floating Gate Transistor

The model for the floating gate transistor, developed in Chapter 6, provides a useful tool for investigating individual devices. However, for practical reasons, it is difficult to

incorporate this model into higher level models of circuit operation. For this reason, a model, compatible with the circuit simulator SPICE has been developed. This allows the effects of floating gate faults on a wide range of circuits to be investigated.

The model should emulate the operation of the floating gate transistor as indicated by the device measurements and theory previously discussed. The coupling of the drain and source potentials to the gate should be correctly represented. It should also be possible to set the floating gate charge (i.e. the initial gate voltage) to any value.

The particular version of SPICE that was used for this work was PSPICE Version 5.2 (MicroSim Corporation). Experience with SPICE over many years has shown that the level 3 MOS transistor model is the most reliable of the three basic models available in SPICE. The level 3 model is based on the Shichman-Hodges model for MOS operation [7.1]. The model differs from level 2 in that the I_{DS} equations have been simplified by the use of empirically derived formulae [7.2]. This speeds up calculations and improves the convergence performance of the model. Level 3 model parameters for the 3 μm process used in this work were provided by the Edinburgh Microfabrication Facility (EMF). Comparison of simulation results with measured characteristics indicated that the parameters used in the level 3 model provide an accurate simulation of the 3 μm devices used in this work.

The FGT is essentially the same as the normal MOSFET but with the constraint that the total gate charge is constant. One of the most important aspects of the device operation for accurate modelling of the FGT is the representation of the gate, channel and substrate charges. Levels 2 and 3 allow a choice between Meyer's model [7.3] and the Ward-Dutton model [7.4] of charge in the MOS transistor. Meyer's model represents the charge in the MOST by three non-linear capacitors in parallel: C_{GB}, C_{GS}, C_{GD}. These vary with V_{GS} and V_{DS} and provide an accurate representation of the charges throughout the device operation. The Ward-Dutton model is a simplification of Meyer's model which may provide better convergence in some circumstances. Both of these models provide a reasonably accurate representation of the gate charge and should be adequate for simulations of the FGT.

The use of the normal SPICE level 3 MOSFET model for the FGT simplifies the task of finding a model for this device. The problem is to find a way of using the existing SPICE MOS transistor model with the constraint of constant gate charge.

Several methods were investigated for the modelling of the floating gate transistor. One possibility was to simply leave the gate terminal unconnected. A SPICE directive could then be used to set the unconnected node to the required floating gate voltage. However, the SPICE program does not allow nodes to remain floating during a simulation so that it is not possible to leave the gate unconnected.

A second possibility that was investigated was to use an MOS transistor as a switch to isolate the floating gate after the simulation had started. The intention was that a DC voltage source could be used to set the gate voltage through the MOS transistor at the start of the simulation. The device would then be switched off immediately after the start and the gate charge would remain isolated, and therefore constant, for the remainder of the simulation. The difficulty with this method was that the gate charge did not remain at the required value as the transistor connected to the gate was switched off because some of the charge in the channel was expelled onto the source of the device, and hence the floating gate, as the channel collapsed. The gate voltage was found to rise by around 0.5 V due to this effect, but the increase varied with the gate voltage and so it was difficult to set the floating gate voltage with the desired accuracy.

A possible solution to this problem is to replace the enabling transistor with a high resistance value. This resistance effectively emulates the resistance to GND seen by the floating gate. This is very large in the real device. In practice, the largest value that could be used in SPICE simulations was around 10^{12} Ω which is sufficiently high for the purposes of this work. A diagram of this subcircuit is shown in figure 7.1.

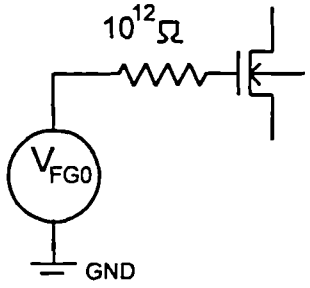


Figure 7.1 A model for the floating gate transistor for use in SPICE simulations

This model cannot be used for DC analysis in SPICE. In such simulations, the DC source, V_{FG0}, determines the gate voltage. Consequently, capacitive coupling from the drain or source is not modelled. This problem can be overcome by using a voltage ramp in transient analysis to provide the varying voltage for the analysis. The speed of the ramp must be high enough to prevent the V_{FG0} source from discharging any voltage changes that occur on the floating gate. This speed can be estimated as follows. The RC time constant of the floating gate source resistance and the gate capacitance is around 10 mS. To ensure that the floating gate source has no effect on the gate charge during the transient analysis, voltage ramps or pulses must be considerably shorter than 10 mS. This condition is easily met. Most simulations of interest would use pulses or ramps of around 10 nS and signal frequencies in the range of MHz. At these speeds the charge on the floating gate will effectively be fixed at the initial value set by the floating gate voltage source.

The I_{DS} - V_{DS} characteristics for the floating gate transistor determined using the model shown in figure 7.1 are shown in figure 7.2. (Note that the range of V_{DS} has been restricted to make comparison of the curve simpler.) Characteristics are given for two values of V_{FG0} which correspond to two typical devices from the test chips. The measured characteristics are also shown. The simulated characteristics can be seen to emulate the real devices well in terms of the shape of the curves produced. However, there is a significant difference in the size of the currents of the simulated and real devices: the simulations produced higher currents than were recorded on the real devices. This also occurred for all other devices that were tested.

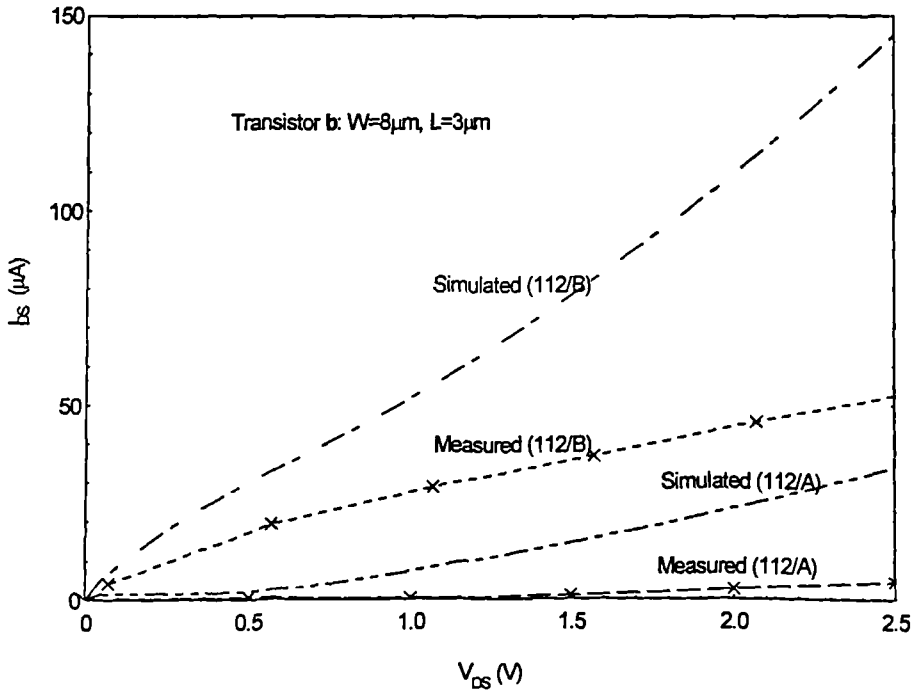


Figure 7.2 I_{DS} - V_{DS} Characteristics for simulated and real floating gate transistors

The source of the difference in the current magnitudes must be identified. The level 3 model parameters are known to be accurate from a comparison with fault free devices, indicating that the discrepancy does not arise from this possible source of error. Examination of the floating gate voltage revealed that this increased by approximately 0.5 V per volt increase in V_{DS}, i.e. the α value was around 0.5. This is much higher than the measured values, and those predicted by the author's model, which ranged from 0.15 to 0.3. The value of α was constant for all V_{FG} for a particular V_{FG0} but it was dependent on the initial gate voltage. For low values of V_{FG0}, α was approximately 0.35 which is considerably higher than the measured values.

The most probable cause for this discrepancy lies in the level 3 model equations themselves. These equations were derived from the level 2 equations to achieve faster

simulations and improved convergence. It is possible that in doing this some accuracy in the modelling of the gate oxide capacitance has been sacrificed. This suggestion is supported by simulations using the floating gate subcircuit with a level 2 MOST. The α value is more realistic with this model and it varies with V_{GS} as seen in both the measured characteristics and the model presented in the previous chapter. The value of α is, however, still higher than expected by a factor of two.

The inaccuracies in the floating gate model would produce misleading results in circuit simulations. The conductance of floating gate transistors would be too high resulting in higher currents and faster circuits than should be expected. This effect would be quite significant as the MOS transistor current is proportional to V_{GS}^2. Errors would therefore be multiplied.

For reliable circuit analysis, a more accurate model of floating gate transistor operation is required. The solution adopted was to use a voltage controlled voltage source (VCVS) to determine the floating gate voltage. This is a SPICE primitive device which allows the value of a voltage source to be set by other node voltages in a circuit. The output voltage can be determined to be a sum of weighted node voltages as shown in equation 7.1.

$$V_{SOURCE} = A_0 V_0 + A_1 . V_1 + A_2 . V_2 + \ldots \qquad (7.1)$$

where $A_n V_n$ are gain and voltage pairs for the controlling nodes. The VCVS can therefore be used to set the floating gate voltage by specifying an initial gate voltage as $V_0 = V_{FG0}$ (A_0=1) and a gain value and circuit node voltages for the drain and source terminals of the floating gate device. The gain value corresponds to the α parameter previously discussed. It is necessary to consider the case of the source-substrate voltage being non-zero for logic circuits, as transistor sources are not always connected to GND. The drain voltage must also be referenced to the substrate for the same reason.

This model has some disadvantages when compared to the previous suggestion. The value of α must be fixed for a simulation which eliminates the variation of α that has been observed. In real devices this variation was seen to be quite significant. However, a wide range of α values was also observed for identical devices. This "error" is as significant as the observed variation. Using a fixed value of α is therefore not a serious limitation of the model.

A second disadvantage is that the value of α must be determined before the simulation, i.e. it is not calculated by the simulation. This is not a problem for the work of this thesis as the devices have been characterised and the range of α is known for the process that is being used. However, this could present problems if process or device parameters were changed. Equation 7.2 gives an empirical relationship for α. This was obtained from the measurements of the parameter for a range of 3 μm devices.

$$\alpha = 0.1 + 0.076 \cdot V_{FG0} \tag{7.2}$$

As indicated above, there is a large spread of values of the parameter for a given device which can be considered to be an error on this prediction. The error in the α values is approximately ±0.05. The need to set α, rather than having it calculated, may be useful for investigation of the effects of the variation of this parameter. The model has a significant advantage over the previous model. The VCVS allows a large number of voltages to be coupled to the source. It is therefore possible to use this model to consider the effect of capacitive coupling from adjacent signal or power lines or from the substrate or well. The latter is considered to be significant by Henderson et al. [2.26].

The I_{DS} - V_{DS} characteristic for this model is shown in figure 7.3. The model is seen to fit the measured data well. Similar comparisons with a wide range of devices indicates that the model is an accurate representation of the floating gate transistor. The VCVS model for the FGT was therefore used for the simulations described in subsequent sections.

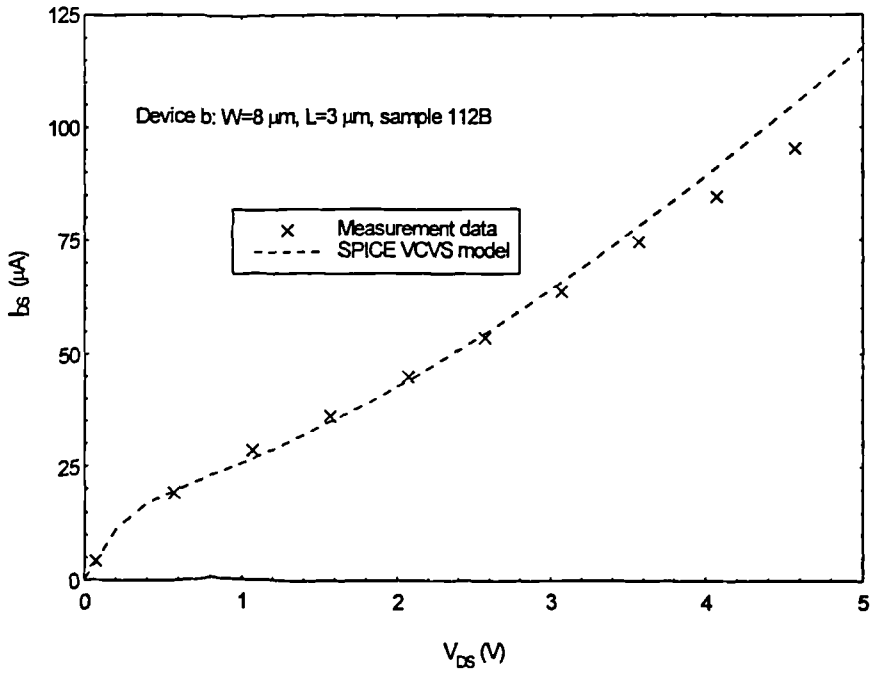


Figure 7.3 I_{DS} - V_{DS} measured and simulated characteristics for a floating gate transistor

7.2 The Operation of Logic Gates with Floating Gate Faults

The results of the analysis of individual FGTs indicate that the devices have variable levels of conduction which can be from subthreshold to saturation for a particular device. In general, most *n*-channel devices will be switched on at some value of drain potential. Very few *n*-channel devices remain fully off (or subthreshold) for V_{DS} = 5V. The variation in the device conductivion means that the FGT cannot be treated as either stuck-open or stuck-

closed. A device which is fully off with V_{DS} = 0V may be switched to saturation by the application of 5 V to the drain. This property means that simple analysis which uses a fixed conduction for FGTs will not predict circuit operation correctly. For this reason, the variation in conduction and the initial gate voltage must be taken into account in simulating logic circuits. With this point in mind, we will now investigate the static and transient operation of logic gates which contain floating gate faults. Where appropriate, the SPICE model described in the previous section will be used to verify the descrpitive analysis of the circuits.

7.2.1 Static Operation of a CMOS Inverter

For this analysis, we will consider a minimum sized inverter which has been designed for symmetric rising and falling propagation delays. The *n*-channel transistor width is 4 μm, the *p*-channel width is 7 μm and the channel length for both devices is 3 μm. The floating gate fault will be introduced onto the *n*-channel, *p*-channel and both devices.

*a) **Floating gate n-channel device.*** A floating gate fault on the *n*-channel transistor (and fault free *p*-channel) is most likely to affect the output voltage level for the logic 1 output state. If the faulty transistor is switched into conduction, either initially, or by coupling from the transistor drain, then the output voltage will be determined by the relative conductance's of the *n*- and *p*-channel devices. This will produce a weak value of V_{OH} which may, in turn, produce a weak output low voltage in subsequent gates. The quiescent supply current consumed by the gate will not be negligible as expected of a CMOS gate. It is possible that this could result in indeterminate logic levels and incorrect logical operation of a circuit.

The output voltage for a logic 0 output condition also depends on the initial value of the floating gate voltage. If this is greater than the *n*-channel threshold voltage V_{Tn}, the output low voltage should be 0V. For $V_{FG0} < V_{Tn}$ the output may be expected to be V_{DD} as the FGT is switched off. However, as explained in Chapter 5, the feedback from the logic gate output to the floating gate can switch on such a device, allowing the output to discharge to a voltage close to 0V. Although the output voltage is not at GND potential, the *n*-channel device is switched off and the quiescent supply current for this state is negligible. The raised voltage has the possibility of producing weak V_{OH} levels in subsequent gates which may result in incorrect logical operation of the circuit.

The output voltage for an inverter can be calculated for a given input by equating the drain to source currents of the *p*- and *n*-channel devices and then solving the resulting equation for V_{OUT}. This technique can also be used for the floating gate inverter but the procedure is complicated by the second variable, V_{FG}. As the floating gate potential and the output voltage are interdependent, further equations or useful boundary conditions are needed to solve the resulting equations for V_{OUT}. This problem was solved using the method described in section 7.1 allowing SPICE simulations to be used to find the V_{IN} - V_{OUT} transfer

curves for inverters with a range of V_{FG0} values from 0 to 5V. These characteristics are shown in figure 7.4.

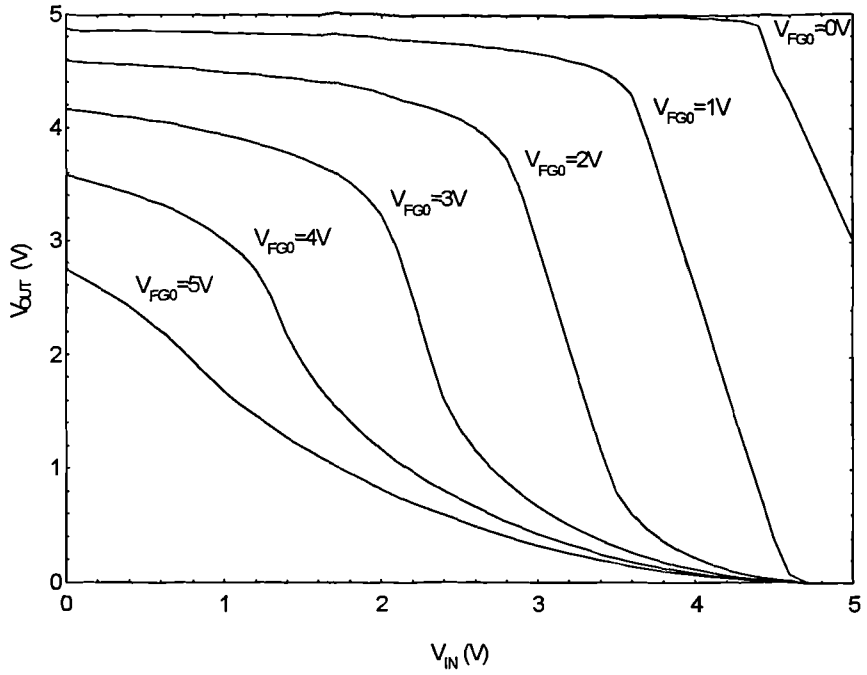


Figure 7.4 Transfer characteristics for an inverter with a floating gate *n*-channel device

It is now possible to assess more accurately the effect of the *n*-channel floating gate fault on the inverter characteristics. The lowest value of V_{OH} is 2.74V and this occurs at V_{FG0} = 5V. This is sufficiently low to switch on the *p*-channel devices of subsequent logic gates and produce a weak V_{OL} in these. For V_{FG0} > 4.7V, the output voltage is sufficiently high to give a clear logical error. However, for V_{FG0} < 4.7V the output voltage for the next inverter is less than V_{Tn} and so no logical error occurs. The gate therefore appears to be fault free for most values of V_{FG0}. The chance of detecting the fault decreases further when the observed range for the initial gate voltage is considered. For the maximum measured V_{FG0} of 1.7 V, the output high voltage for the faulty inverter is 4.7 V. therefore, for all observed V_{FG0} values of V_{FG0}, the gate would appear to be fault free.

The supply current for this condition is considerably higher than for the fault free circuit. For the case of V_{FG0} = 5V the inverter supply current is 320 μA when the output is at logic 1. This falls to 60 μA for V_{FG0} = 1.7 V. The supply current is a significant parameter which may be used as a fault indicator. This will be discussed in a later section.

The highest value of V_{OL} that is observed for the inverter is 3.0V. This occurs when V_{FG0} = 0V. The V_{OL} voltage falls rapidly with increasing V_{FG0} and is significantly below the *n*-channel threshold voltage for all V_{FG0} > 0.6V Consequently, for most values of V_{FG0} the

gate appears to be fault free for static testing. the inverter supply current for this condition is negligible.

This analysis has confirmed the conclusions drawn in Chapter 5 on the static operation of an inverter with an *n*-channel floating gate fault. The inverter appears to operate correctly for most values of the initial gate voltage, V_{FG0}. There are some deviations from normal behaviour for V_{FG0} around 5V but since this voltage has not been observed in the samples analysed, this is not considered to be a significant case. The fault would appear to be undetectable by static logic testing in most cases. Detection of floating gate faults will be analysed in detail in a later section.

b) Floating gate p-channel device. For this discussion, we will refer to the floating gate potential with the inverter power applied. This raises the substrate of the *p*-channel device to V_{DD} which in turn raises the floating gate potential by the same amount. The term V_{FG0} for the *p*-channel transistor in circuits, is therefore used to refer to the floating gate potential with power applied and no coupling from the source or drain. The drain and source potential are referenced to the substrate (i.e. V_{DD}).

The floating gate on the *p*-channel transistor of an inverter will cause the output low voltage to be greater than 0V for some values of V_{FG0}. For example, with V_{FG0} = 5V, as the voltage on the logic gate output, and hence the drain of the floating gate device, decreases to 0V, the floating gate voltage will become less than 5V. If V_{FG} becomes less than V_{DD} - V_{Tp} then the transistor will switch on and prevent the output from reaching 0V. The output low voltage will increase as V_{FG0} decreases. As for the *n*-channel device, a weak output voltage may result in incorrect operation of subsequent levels of gating and high supply currents.

The output high voltage for the inverter will be degraded by the same mechanism as that described for V_{OL} in the floating gate *n*-channel fault. The output high voltage is expected to be less than V_{DD} for V_{FG0} values around V_{DD} due to cut-off of the device channel.

Transfer characteristics for an inverter with a *p*-channel floating gate fault are shown in figure 7.5. The output high voltage is V_{DD} for all values of V_{FG0} from 0V to V_{DD}. The gate appears to be fault free for this output state. The cut-off observed for the *n*-channel fault is not apparent for the *p*-channel case. The most probable reason for this is that the *p*-channel device has a low threshold voltage ($\approx$ -0.5V) compared to the *n*-channel device ($\approx$ 0.9V). Consequently, the cut-off effect will occur for lower values of V_{GS}. This effect is enhanced by a high subthreshold current (or weak inversion current) for the *p*-channel device. The weak inversion current is generally thought to be exponentially related to V_{GS} and other parameters, including the surface state density. This parameter is generally higher for *p*-channel devices and this is the case for the models used in this simulation. Simulations show the *p*-channel subthreshold current to be greater than the *n*-channel current and in fact it

appears to be non-zero for $V_{GS} = V_{DD}$ on the *p*-channel devices. Cut-off effect will therefore not occur until V_{FG} is greater than V_{DD}.

The output low voltage is considerably higher than 0V for a wide range of V_{FG0} values. V_{OL} becomes greater than V_{Tn} for $V_{FG0} <$ 2.8V. This will cause the output high voltage of the gates driven by this signal to be poor. In practise, the V_{OH} of a subsequent gate would only become a problem for $V_{FG0} <$ 1.7V. The supply current is high for all values of V_{FG0}, rising from 10 to 265 μA for V_{FG0} = 5 to 0V.

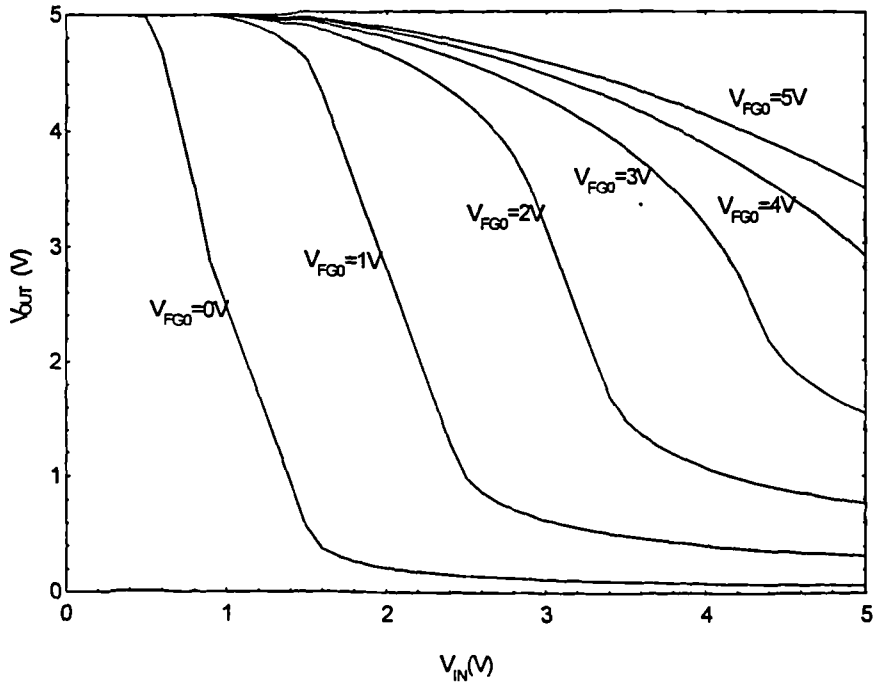


Figure 7.5 Transfer characteristics for an inverter with a floating gate *p*-channel device

The range of V_{FG0} values that were simulated correspond to a negative residual gate charge on the floating gate. In practise, the charge was found to be positive for most devices. with only three transistors having negative gate charge. To emulate the observed range of charges, a floating gate voltage of between 4.5 and 7.5V was simulated. The output low voltage for V_{FG0} between 4.5 and 6.5V is well below V_{Tn} and will not cause logical errors in the circuit. The cut-off phenomenon is observed for $V_{FG0} >$ 5V (V_{DD}), as expected, and so V_{OH} is low in these cases. The output high voltage is sufficiently low to potentially cause logical errors for $V_{FG0} >$ 6V.

The *p*-channel floating gate fault on an inverter appears to produce logically incorrect outputs only for $V_{FG0} <$ 1.7V and $V_{FG0} >$ 6V. For the process used for this work, the majority of *p*-channel devices are expected to have a V_{FG0} value between 4.5 and 7.5V. It is therefore clear that the logic gate appears to operate correctly for most values of V_{FG0} that occur in practice.

c) ***Floating gate n- and p-channel devices.*** In this case, the break in the polysilicon track that produced the fault has resulted in both *n*- and *p*-channel gates floating, but still connected together. The input voltage therefore, has no effect on the circuit and the output will be fixed at a potential determined by V_{FG0} at any value between 0V and V_{DD}. The supply current for this fault will depend on the initial floating gate potential. For the symmetric inverter being considered in this analysis, the maximum supply current would be 70 μA. This fault will certainly cause a logical error in the circuit.

The logical operation of the inverter does not appear to be significantly affected by a single floating gate fault when simple logical tests (i.e. not dynamic) are used for the analysis. The faults are only apparent for extreme values of V_{FG0} which tend not to occur in the samples that have been analysed. The fault would have been detected for only one out of the 35 samples that have been tested.

Clearly logic circuits are generally used in dynamic operating conditions, i.e. with repetitive, high speed signals applied. It is therefore necessary to analyse the dynamic behaviour of the inverter with floating gate faults introduced.

7.2.2 Dynamic Operation of a CMOS Inverter

a) ***Floating gate n-channel device.*** The floating gate *n*-channel transistor is expected to have a conductance that is generally lower than that of a fault free device with V_{DD} applied to the gate. We therefore expect the output falling propagation delay, T_F, to be increased for floating gate devices. The effect will be most significant for low values of V_{FG0} when the *n*-channel device will have very low conductivity. The increase in delay may be very significant (microseconds or more) for V_{FG0} = 0V, but can be expected to fall rapidly once the floating gate voltage becomes greater than V_{Tn}. In some cases, where $V_{OH} \neq V_{DD}$, T_F will be reduced as the output is discharging from a lower voltage.

For high values of V_{FG0}, i.e. $V_{FG0} > 4$V, we might expect T_F to decrease compared to the fault free value. This is because the floating gate potential is raised above V_{DD} (the value for the fault free case) and so the conductance of the faulty device is raised. This will increase the current that is discharging the load capacitance and hence speed up the transition.

The output rising delay should not be significantly affected by a floating gate *n*-channel fault. However, as the floating gate device will probably be switched on for at least part of the output rising transition, the current charging the output load will be reduced. This will cause the output rising delay T_R to increase. This effect will not be as significant as the increase in T_F.

To quantify the effects described, simulations were performed using the VCVS floating gate model with a range of values of V_{FG0}. The circuit used for simulations is shown

in figure 7.7. The inverter dimensions are: W_n = 4μm, W_p=7μm, $L_{n,p}$=3μm. A second, fault free, inverter is included to assess the effects of faults on subsequent logic gates. The final load capacitance represents a typical fan out of three or four logic gates.

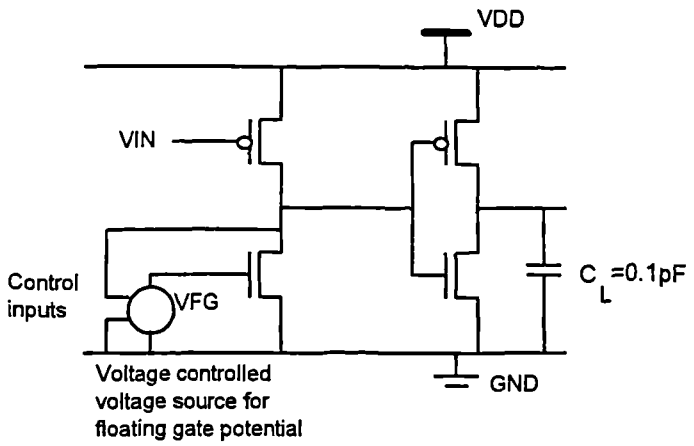


Figure 7.6 Inverter circuit used for transient analysis of floating gate fault effects

Transient simulations were run for V_{FG0} = 0 to 5V. The input waveform used is a simple pulse with rising and falling delays matched to the fault free inverter. Graphs of the inverter response to a 15ns pulse are shown in figure 7.7. It is worth noting that 15ns was chosen to simplify the illustration of both rising and falling output transitions. For a 15ns pulse to occur in a synchronous circuit, the clock frequency would be around 30MHz which is a high speed for a 3μm technology.

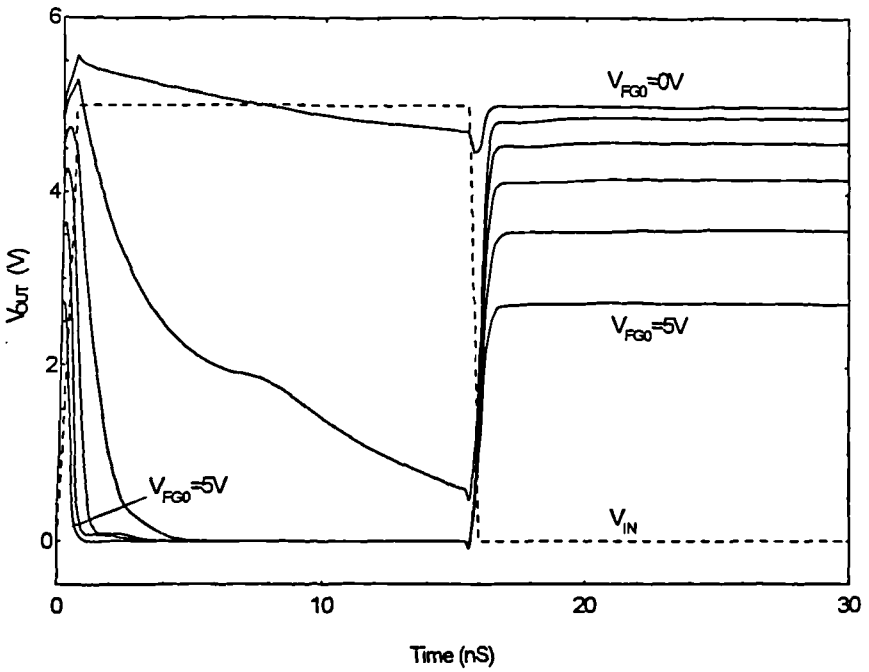


Figure 7.7 Dynamic response of an inverter with an *n*-channel floating gate fault

The increase in falling propagation delay predicted above is clearly seen in the graph. However, it is important to note that for many values of V_{FG0}, the increase in T_F from the fault free value of 0.24ns is only a few nanoseconds. (Propagation delays have been measured from the 50% input voltage to 50% output voltage). Whilst this is a large increase of ten or fifteen times the fault free value, the absolute increase is probably not very significant in most cases. The change in T_F as V_{FG0} varies is shown in figure 7.8. The shape of the curve is as predicted in the previous discussion with a rapid rise in T_F for $V_{FG0} < 1.0$V. Also, the propagation delay is decreased for $V_{FG0} > 3.5$V.

The point at which the increase in T_F becomes significant is not clear. Consider, for example, a large CMOS synchronous ASIC which is designed to operate at 25MHz which is a reasonably demanding speed requirement for a 3μm technology. The speed of the circuit is limited by the maximum propagation delay that exists on the chip between two bistables. this delay (including the bistable delays) must be less than 40ns for the circuit to operate at 25MHz but for a typical design, it is probable that only a small percentage of signal paths will have a total delay close to 40ns. The majority will be 10 or 20ns less than this. In such signal paths, the propagation delay for any one gate may increase by 10 or 20ns before the circuit fails to operate correctly. Clearly then, in many cases the *n*-channel floating gate fault will not cause the circuit to fail, even when it is operating at its full design speed.

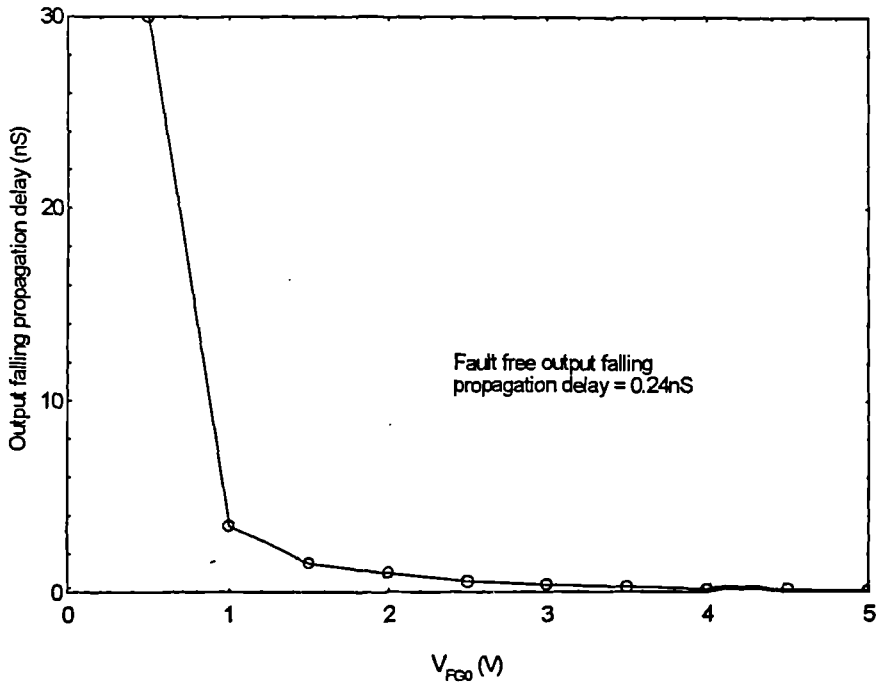


Figure 7.8 Variation of the falling propagation delay with V_{FG0} for an inverter with *n*-channel floating gate fault

The increase in the output rising delay, T_R, is measurable but not significant. The fault free delay of 0.23ns increases to 0.52ns for $V_{FG0} = 5$V, which is the worst case. The increases

inverter are increased by a slightly larger amount than those for the faulty gate for low values of V_{FG0}. This occurs because of the decreased slopes of the input transitions. The propagation delay does not degrade significantly past the next level of gating.

b) Floating gate p-channel device Introduction of the floating gate fault into the *p*-channel device affects the output rising delay, T_R. The delay increases rapidly for values of $V_{FG} > 4.5$V, reaching 17ns at $V_{FG0} = 5$V. Above this gate potential, delays are very significant, with for example, $T_R \approx 1000$ns when $V_{FG0} = 5.5$V. When $V_{FG0} > 5.8$V the output high voltage is less than the inverter transition voltage, i.e. the output is stuck-at 0. The graph of rising output delay against V_{FG0} for the *p*-channel floating gate fault is shown in figure 7.9. The increase in the output falling delay for this fault is not significant.

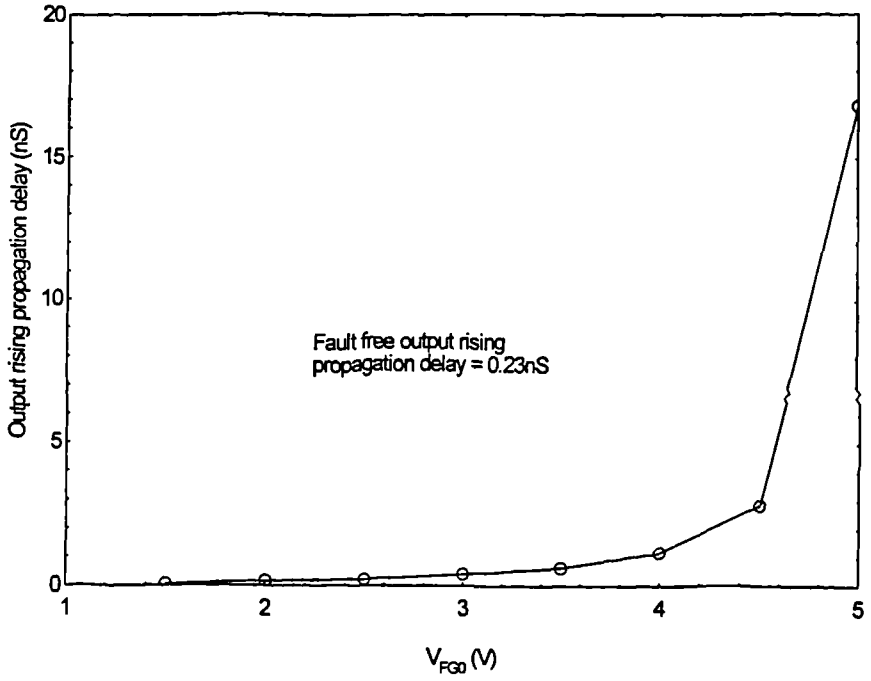


Figure 7.9 Variation of the rising propagation delay with V_{FG0} for an inverter with *p*-channel floating gate fault

From the preceding analysis, it is clear that floating gate faults do not necessarily degrade circuit performance sufficiently to cause logical or timing failures. If, for example, we consider the measured range of V_{FG0} values for *n*-channel devices and assume a uniform distribution in this range, we can estimate the number of floating gate faults that would cause failure of the circuit. We will assume that an increase in signal propagation delay of up to 10ns can be tolerated by 90% of the signal paths on a chip. In such circumstances, only 32% of the floating gate faults would cause a failure. This figure is possibly slightly optimistic as some of the devices with low V_{FG0} have an *n*-channel width of 8, 12 or 16μm. Clearly these gates will operate faster than the minimum sized inverter and the increase in delay will not be as significant. In fact, for $V_{FG0} = 0.6$V and $W_n = 12$μm, the output falling delay is only 4.5ns. This increase in delay will cause a logical error in very few circuits. If we consider higher

values of V_{FG0} which may occur in other processes, the chance of a logical error occurring decreases further.

The effect of *p*-channel floating gate faults on inverter operation is more significant for the observed values of floating gate charge. This is because the net residual gate charge seems to be positive for both *n*- and *p*-channel devices. Consequently, the floating gate voltage on the *p*-channel device when it is in a powered inverter circuit, is greater than the supply potential. The feedback to the gate from the drain and source nodes must overcome the initial floating gate potential and the *p*-channel threshold voltage before the transistor can switch on. It seems most likely that all *p*-channel floating gate faults on the test circuits would have caused logical errors if they had occurred in a logic circuit. Although the static characteristic would not have revealed the faults, dynamic testing of the gates would have produced logical failures due to the increased rising delay associated with these faults.

If we allow the *p*-channel floating gate charge to become negative, then the fault becomes undetectable by static and dynamic logic testing for a wide range of values. Obvious sources of negative charge do not arise in CMOS IC fabrication. The ions implanted at various stages in the process are always positively charged. A technique has recently been introduced specifically to stop charge build up during implantation using an electron wind that is passed over the wafer during implantation to compensate for the positive ionic charge accumulation. A net negative charge could then occur on isolated polysilicon gates if the charges are not correctly balanced. As shown previously, the amount of residual charge required to generate a gate potential of a few volts is very small (typically 100 fC) and so only a slight imbalance is needed to produce a *p*-channel floating gate voltage which is less than V_{DD}.

In conclusion, it seems that *n*-channel floating gate faults generally have little effect on inverter performance. In contrast, floating gate faults on *p*-channel devices will generally cause logical errors in inverters. We must note, however, that the magnitude of the effect of the fault varies greatly for each type of device. The most reliable conclusion is, in fact, rather inconclusive: *the effect of floating gate faults on CMOS inverters can be wide ranging, producing clear logical errors in some cases and allowing fault free operation in others.*

7.2.3 Multiple Input Logic Gates

Multiple input logic gates act in a similar way to inverters when floating gate faults are introduced. The output high and output low voltages may be weakened and propagation delays increased. However, the effects are only significant for a limited range of values for the initial floating gate potential, as for the inverter. The additional inputs on these gates may result in the effects of the fault being masked for certain input combinations.

As an example, we will consider the effect of floating gate faults on a two input NAND gate. In order to broaden the scope of the conclusions that can be drawn, we will consider a gate with symmetric layout, i.e. both *p*- and *n*-channel devices have the same dimensions. This is not uncommon in semicustom circuits (and indeed in full custom) although it does result in asymmetric propagation delays in the logic gates. In the gate to be considered, all devices have a channel width of 10μm and a length of 3μm. The circuit to be analysed is shown in figure 7.10 with a floating gate fault introduced onto one of the *n*-channel transistors.

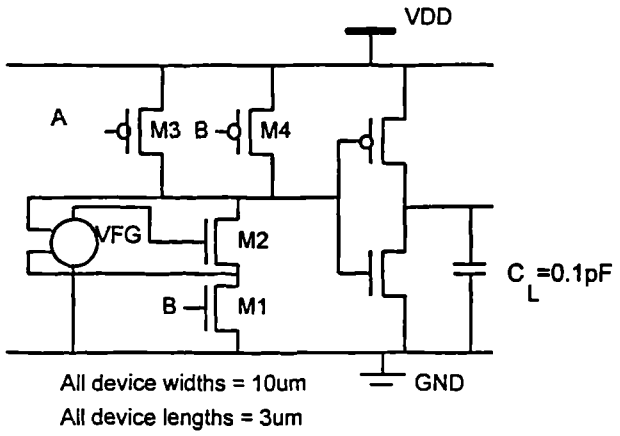


Figure 7.10 Two input NAND gate used for floating gate analysis

Four distinct forms of floating gate fault can be introduced into the two input NAND gate.

a) ***Floating gate p-channel device.*** The effect of a single floating gate *p*-channel device will be the same as that observed for the inverter circuit. (The same convention will be used for V_{FG0} in this section as that used in the discussion on inverter, i.e. a net zero residual charge will give $V_{FG0} = 5V$). The output low voltage will be greater than 0V for $V_{FG0} < V_{DD}$. The output low voltage for an inverter with a *p*-channel floating gate fault is weakened for $V_{FG0} < 1.7V$. The electrical asymmetry of the NAND gate will produce a higher V_{OL} value for a given floating gate potential when compared to the inverter. The voltage at which V_{OL} causes a failure will therefore be higher than that of the inverter. The fault is therefore more likely to have a noticeable effect on the circuit operation.

The floating gate *p*-channel fault may also cause a poor V_{OH} value due to the device cut-off phenomenon previously described. The effect is not determined by the pull-up and pull-down network resistance ratio, and so is independent of the number of inputs to a logic gate. The output rising propagation delay will be increased for high values of V_{FG0}. The degradation will be similar to that found for the inverter but the effect will be less significant due to the wider devices used in this gate. The increase in T_R can only be expected to be a problem for $V_{FG0} > V_{DD}$.

The effects described are summarised in Table 7.1. It should be noted that no fault effects are apparent if the fault free input is at logic 1.

A	B	OUTPUT
0	0	1
0	1	F: V_{OH} poor, T_R large
1	0	1
1	1	F: V_{OL} poor

Table 7.1 Truth table showing fault effects due to a floating gate *p*-channel fault in a two input NAND gate

*b) **Floating gate n-channel device.*** The effects of a floating gate *n*-channel fault on the two input NAND is similar to that found for the inverter although there are some differences in the dynamic response. There are two forms for this fault as the floating gate may occur on either of the two *n*-channel transistors in the gate. The effects may be different for the two cases since the drain and source potentials are not always the same for the two transistors.

The output high voltage for the gate is degraded if V_{FG0} is sufficiently high. The electrical asymmetry will make the fault effect less significant for a given potential in comparison with the inverter. The output low voltage may be raised above 0V by cut-off of the transistor for low values of V_{FG0}. The effect would be very similar to that observed for the inverter circuit. These effects on the static characteristics of the gate will be very similar for the two *n*-channel device faults.

The output falling delay will be increased for low values of V_{FG0}. The effect will be slightly different for the two cases, i.e. the floating gate on the top or bottom device in the *n*-channel series stick. The reason for the difference is that the voltages coupled to the floating gate are different for the two devices. If the fault occurs on the bottom transistor, the effect is similar to that for the inverter circuit. The source potential is fixed at 0V and so the gate potential cannot be raised by coupling from the source. The drain potential is coupled to the gate but the maximum drain voltage is V_{DD} - V_{Tn} due to the V_{Tn} drop across the top *n*-channel device. The voltage coupled to the floating gate is therefore slightly less than for the inverter. If the fault occurs on the top transistor, the full output voltage is coupled to the gate. Also, the source voltage of the top device is not 0V during transients and so the floating gate potential will be raised significantly higher for short periods during transitions.

A further point to be noted on the dynamic response of the gate is that the pull-down resistance is lower for the NAND gate being considered, compared to the inverter previously

discussed. We would therefore expect T_F to be smaller. In fact the output falling delay is increased compared to the inverter delay because the load driven by the NAND is larger due to the wider devices used in the following inverter. Because of this, the increase in T_F due to the floating gate fault should be similar to that seen for the inverter.

This conclusion is verified by SPICE simulations of the circuit. The output falling delay for the fault free gate is 0.45ns, which is twice the inverter delay. The increase in T_F, for the case of a floating gate fault on the lower *n*-channel transistor, is approximately 20% greater than that for the inverter. The increase for the floating gate fault on the upper device is approximately 15% *less*. These differences are not very significant. The simulations also verified the conclusions on the effect of the fault on the output voltage of the gate. The effects are very similar to those observed for the inverter.

c) ***Floating gate n- and p-channel devices.*** The fourth form of floating gate fault is the connected floating gate in which one pair of *n*- and *p*-channel devices have gates which are connected together but floating. The arguments about the increased significance of the coupling for the upper *n*-channel device also apply to this fault. However, these are less important than those observed for the single transistor fault and can be ignored.

The most significant difference between the connected and single floating gate faults is the initial floating gate voltage. If we ignore the drain voltage, the gate voltage can be determined by the ratio of the *n*- and *p*-channel gate capacitances. The gate is essentially the centre node of a capacitive potential divider circuit consisting of the two gate capacitances C_{Gp} and C_{Gn}. With the *p*-channel source and substrate taken to V_{DD} and for zero gate charge, the initial gate potential is given by equation 7.3.

$$V_{FG0} = \frac{C_{Gp}}{C_{Gp} + C_{Gn}} \cdot V_{DD} \tag{7.3}$$

For the example gate being considered, $C_{Gn} = C_{Gp}$ and so $V_{FG0} = V_{DD} / 2$. Residual gate charge will simply add a voltage, V_R, given by equations 7.4.

$$V_R = \frac{Q_{FG}}{C_{Gn} + C_{Gp}} \tag{7.4}$$

We therefore obtain the initial gate potential without drain coupling (or with the drain potential set to 0V) by adding equations 7.3 and 7.4:

$$V_{FG} = \frac{C_{Gp}}{C_{Gp} + C_{Gn}} \cdot V_{DD} + \frac{Q_{FG}}{C_{Gn} + C_{Gp}} \tag{7.5}$$

The drain (or source) voltage can then be added after scaling it by the α value for the transistor. Although there are two overlap capacitances contributing to the coupling, the other capacitors in the divider circuit, i.e. the gate and source capacitors, are also doubled. The ratio should therefore be very similar to the single device case. The variation of α with device width which was apparent from measurements may complicate the determination of α for geometrically asymmetric circuits (i.e. where $W_p \neq W_n$). However, this effect is not likely to be very significant and can be ignored for this analysis. It should be noted that source coupling should be included for any devices for which the source is not at 0V.

The consequence of capacitive coupling is that the initial gate voltage will be greater than V_{DD} / 2 for devices with a positive residual gate charge. The voltage is increased further by any geometric asymmetry which would cause C_{Gp} to be greater than C_{Gn}. This must be taken into account when considering the expected range of the potential of floating gates for circuits with connected gate faults.

Clearly if we try to change the output by changing the faulty input on a logic gate with a connected floating gate fault, the output will appear to be stuck at 0 or 1, depending on the logic value of the fault free input. The fault, however, may not have such a significant effect if we attempt to toggle the output via the fault free input. The response will depend on the gate voltage of the faulty input and, to a lesser extent, on the pull-up and pull-down resistance ratio.

Consider the case of the gates of transistors M2 and M3 in figure 7.10 connected together but otherwise floating, and with an initial floating gate voltage at 0V. If the fault free input is at logic 0, transistor M4 will be on, M1 will be off and the output will be at logic 1 irrespective of the value of V_{FG0}. However, when the fault free input is switched to logic 1, the NAND gate output will depend on the floating gate input. With the floating gate input at 0V, the faulty *n*-channel device, M2, will be off and M3 will be on. This will cause the output to rise towards V_{DD}. Since the output voltage is coupled to the floating gate, the gate potential will rise by approximately 1V. This will switch on the *n*-channel device causing the output to fall which will, in turn, will lower the gate potential until the *n*-channel device switches off. An equilibrium will be established in which the *n*-channel device is just on and the gate output is close to V_{DD}. The output low voltage is therefore well above the transition voltage for the gate and the floating input will appear to be a logic 0.

If we set V_{FG0} to V_{DD} the faulty *n*-channel device is switched on and the *p*-channel is off. The output will be at 0V and so the floating gate potential will stay at V_{DD}. The output low voltage for $V_{FG0} = V_{DD}$ is therefore 0V and the faulty input will therefore appear to be a logic 1. In between these two values of V_{FG0}, the gate output voltage will be between V_{DD} and 0V. Whether or not the fault is apparent depends on the driven value of the faulty input. Clearly, if the input is set to logic 1 and V_{FG0} is less than approximately 2.5V, then the gate

will operate incorrectly. Similarly, if the input is set to logic 0 and V_{FG0} is greater than 2.5V, the fault will be revealed. If, however, either of the converse situations occurs the fault will not be apparent.

The increase in propagation delay which might be expected to occur for the faulty gate is very small. At $V_{FG0} = 2.45$V, which is the worst case, T_F is only 0.9ns which is twice the fault free delay. From this we can conclude that if the floating gate voltage conforms with the set value of the faulty input, the gate will appear to operate correctly, even for high speed operation.

The gate charge was found to be positive for all samples measured on the test chips. Therefore, the floating gate voltage for a coupled gate can be expected to be greater than V_{DD} / 2. For most cases, we can therefore conclude that the floating gate fault would only be apparent for the input condition $AB = 01$. In this situation, the output of the faulty gate would be logic 0 rather than logic 1 as expected. The operation of the NAND gate with connected floating gate fault is summarised in Table 7.2.

A	B	OUTPUT for $V_{FG0} < 2.5$V	OUTPUT for $V_{FG0} > 2.5$V
0	0	1	1
0	1	1	V_{OH} too low
1	0	1	1
1	1	V_{OL} too high	0

Table 7.2 Truth table showing fault effects due to a connected floating gate fault in a two input NAND gate

We have now analysed the response of a two input NAND gate to all forms of floating gate fault. It is clear that in many cases the fault has little effect on the logical operation of the gate. This is true for both dynamic and static operation. Analysis of a two input NOR gate reveals similar operation when floating gate faults are introduced. However, for a gate constructed using geometrically symmetric devices, the effects are notably different in some circumstances. We will now consider a two input NOR gate in which all devices have a width of 10μm and length of 3μm. This produces asymmetric electrical performance with T_R = 0.65ns and T_F = 0.35ns.

The effect of a single *n*-channel floating gate fault on such a gate is significant and is revealed when the fault free input is at logic 0. For $V_{FG0} > 2$V, the output high voltage is very weak and the gate fails to drive the next level of logic gates. For $V_{FG0} < 0.5$V, the output is either stuck at logic 1 or T_F is large. For $V_{FG0} = 0.5$V, $T_F = 17$ns and this increases

as V_{FG0} becomes smaller. There is therefore a narrow range of values of V_{FG0} for which the gate will operate correctly: V_{FG0} must be in the range 0.5V to 2.0V for fault free operation. This range contains all of the values of V_{FG0} observed on the test devices. We can therefore conclude that NOR gates containing the floating gate devices on the test circuits would have operated satisfactorily. If V_{FG0} is outside of this range the gate will fail.

As for the case of the *n*-channel floating gate fault in the NAND gate, there are two distinct floating gate *p*-channel faults on the NOR gate. However, the differences between the two are quite slight and so only the worst case fault will be considered. This occurs on the lower device in the two transistor *p*-channel stick. For this fault, the gate operates correctly for $V_{FG0} < 4.5$V. Above this voltage, the output rising propagation delay is significantly increased. At $V_{FG0} = 4.5$V, $T_R = 12.5$ns. For the expected range of V_{FG0}, from approximately 5.0V to 7.5V, the increase in T_R is very large (> 100ns). The fault is therefore likely to cause incorrect dynamic operation of the circuit.

The response of the NOR gate for the case of the connected floating gate fault is very similar to that observed for the NAND gate. The fault may be revealed when the fault free input is set to logic 0. If $V_{FG0} > 2.5$V the gate output is at (or close to) 0V. With $V_{FG0} <$ 2.5V the output is close to V_{DD}. As the floating gate voltage will tend to be greater than $V_{DD}/2$ in most cases, the fault will be revealed if the faulty input is set to logic 0.

One further factor that should be considered for multiple input logic gates is the effect of increasing the number of inputs on the gate. The only change in the output response that occurs as the number of inputs is increased is related to the change in the pull-up to pull-down resistance ratio. As the electrical asymmetry of the gate increases, the range of V_{FG0} for which the gate will operate correctly is reduced. However, this decrease is quite small so that conclusions drawn for the two-input gate are valid but with a minor reduction in the actual values of V_{FG0} for which the gate will operate correctly. ~~However, this decrease is quite small and the conclusions drawn for the two-input gate are essentially valid, with minor modifications to the critical values of V_{FG0}.~~ Complex gate structures also have a very similar response to floating gate faults, the only differences again being introduced by the network resistance changes.

7.2.4 Summary of Logic Gate Operation with Floating Gate Faults

The effects of floating gate faults on the operation of static logic gates are summarised in Table 7.3. The output voltage of the gate is affected for most floating gate faults. It should be noted however, that in many cases, the effect is only significant for a limited range of floating gate voltages. Consequently, the output voltage would be quite normal in many instances. The propagation delays of gates are also affected but again the range of V_{FG0} for which this is important is limited. This will result in many gates toggling at full operating

speed when floating gate faults are present. The parameter which is most frequently affected is the quiescent power supply current, I_{DDQ}, of the faulty gate. This is raised by anything from a few hundreds of nanoamps to several hundred microamps. For this reason, it is the most reliable indicator of the presence of floating gate faults. However, it cannot be relied on as a testing method completely since the supply current is quite small in many cases and may it not be measurable above the normal supply current of a large VLSI circuit.

Fault	V_{OH}	V_{OL}	T_R	T_F	I_{DDQ}
INVERTER					
n-channel floating	Poor for V_{FG0} > 4.7V	Poor for V_{FG0} < 0.6V	Not affected	Increase sig. for V_{FG0} < 1.0V	Raised for V_{FG0} > 0.6V
p-channel floating	Poor for V_{FG0} > V_{DD}	Poor for V_{FG0} < 1.7V	Increase sig. for V_{FG0} > 4.5V	Not affected	Raised for V_{FG0} < 5.5V
Connected floating gates	Output stuck at voltage between 0 and V_{DD}				Raised for all V_{FG0}
NAND Gate					
n-channel floating	Poor for AB=01 and V_{FG0} > 4.7V	Poor for AB=11 and V_{FG0} < 0.6V	Not affected	Increase sig. for V_{FG0} < 1.0V	Raised for V_{FG0} > 0.6V and AB=01
p-channel floating	Poor for AB=01 and V_{FG0} > V_{DD}	Poor for AB=11 and V_{FG0} < 1.7V	Increase sig. for V_{FG0} > 4.5V	Not affected	Raised for all V_{FG0} and AB=11
Connected floating gates	Poor for AB=01 and V_{FG0} > 2.5V	Poor for AB=11 and V_{FG0} < 2.5V	Not affected for V_{FG0} < 2.5V	Not affected for V_{FG0} > 2.5V	Raised for all V_{FG0} and AB=01 or 11

Table 7.3 (a) Summary of floating gate fault effects on simple logic gates (Continued over the page)

Fault	V_{OH}	V_{OL}	T_R	T_F	I_{DDQ}
NOR Gate					
n-channel floating	Poor for AB=00 and $V_{FG0} > 2.5V$	Poor for AB=10 and $V_{FG0} < 0.4V$	Not affected	Increase sig. for $V_{FG0} < 0.5V$	Raised for all V_{FG0} and AB=00
p-channel floating	Poor for AB=00 and $V_{FG0} > V_{DD}$	Not affected	Increase sig. for $V_{FG0} > 4.5V$	Not affected	Raised for all V_{FG0} and AB=10
Connected floating gates	Poor for AB=00 and $V_{FG0} > 2.5V$	Poor for AB=10 and $V_{FG0} < 2.5V$	Not affected for $V_{FG0} < 2.5V$	Not affected for $V_{FG0} > 2.5V$	Raised for all V_{FG0} and AB=00 or 10

Table 7.3 (b) Summary of floating gate fault effects on simple logic gates

7.3 Transmission Gates and other Circuits

The transmission gate is an important element in many CMOS circuits. It is commonly used in bistables and circuits where tristate outputs are needed for bus drivers or data selector applications. Floating gate faults may occur on either device in the transmission gate, but the connected floating gate fault cannot occur because the gates are driven by separate control signals.

We will first consider an *n*-channel floating gate fault on the transmission gate. This fault may affect the operation of the gate when it is open or closed. When the gate is closed, the fault free *p*-channel device will be operating normally and for $V_{IN} > |V_{Tp}|$, the output voltage will follow the input. The output high voltage will therefore be V_{DD}. For $V_{IN} < |V_{Tp}|$ a potential problem arises due to the floating gate fault. In this range, the *p*-channel device is cut-off and if the *n*-channel floating gate voltage is sufficiently low, the output will be $|V_{Tp}|$. This is the normal voltage drop associated with a pass transistor which is overcome in the transmission gate by the parallel *n*-channel device. However, if the gate voltage is too low to switch the *n*-channel device on, this does not occur and the output low voltage is $|V_{Tp}|$. This may be quite high as a source to substrate bias exists for the *p*-channel device so that $|V_{Tp}| > |V_{Tp0}|$. For $V_{FG0} > V_{Tn}$ the output low voltage will be 0V. The raised V_{OL} value is unlikely to cause a logical error since $|V_{Tp}|$ will always be less than the transition voltage for the next gate. While the supply current for the next gate will be raised, the output will be logically correct.

The propagation delays associated with the transmission gate are not significantly affected by floating gate faults even for V_{FG0} = 0V. This is because the *p*-channel device is on for most of the transition. Furthermore, the coupling of the drain voltage to the gate ensures that the floating gate device is on for part of the transition for most values of V_{FG0}.

The floating gate also introduces a fault effect when the transmission gate is switched off. Clearly for $V_{FG0} > V_{Tn}$, the transmission gate will remain closed and the output will follow the input changes, within certain constraints. Consider the following example. The output is set to logic 1 when the *p*-channel device is switched off to open the transmission gate. If the input remains at logic 1, the output will also stay at this level for all values of V_{FG0}. If the input should drop to 0V however, the output will discharge to 0V for all values of V_{FG0} greater than approximately 0.5V. For $V_{FG0} < 0.5$V, the *n*-channel device will cut-off and the output would remain at a level slightly above 0V.

For the case of the input changing from logic 0 to logic 1 after the transmission gate is opened, the output will charge up to a value determined by V_{FG0}. The *p*-channel device is off and so the maximum output voltage is V_{FG} - V_{Tn}. Clearly for small values of V_{FG0}, the maximum output voltage may be very low.

These effects are illustrated in figure 7.11 which shows the simulated response of a transmission gate output for the sequences described above. A value of V_{FG0} = 2.0V was chosen as an example.

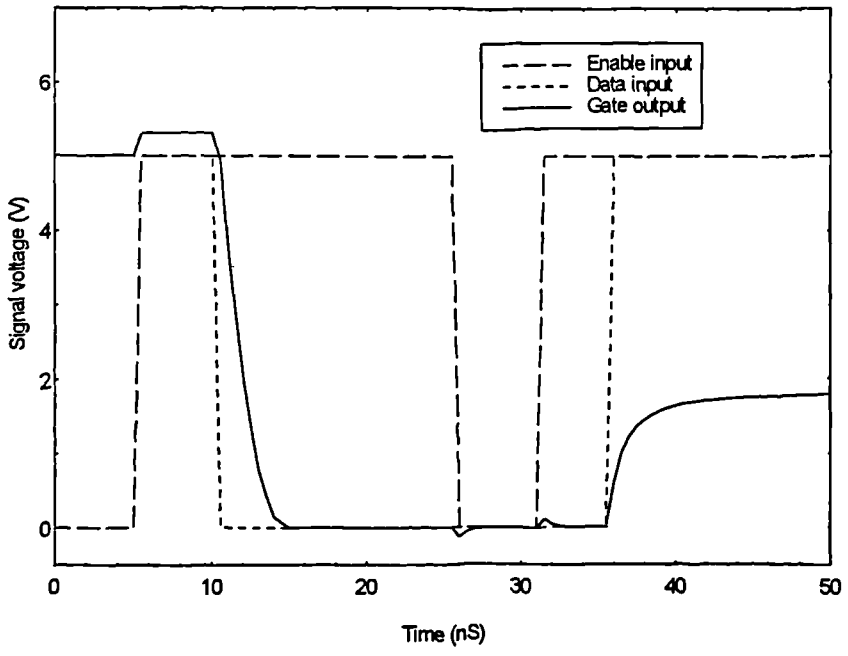


Figure 7.11 Output response of a transmission gate with *n*-channel floating gate fault, V_{FG0} = 2.0V

The operation of a transmission gate with a p-channel floating gate fault is essentially the same as that for the n-channel fault with the appropriate changes for logic states and voltages.

The consequence of the fault effects depends on the function of the transmission gate in its circuits. The high V_{OL} and low V_{OH} which occur when the gate is closed do not cause logical errors in any form of circuit. They will raise the supply current in subsequent gates but, with this exception, the fault will not be apparent.

A more serious effect occurs when the transmission gate cannot be switched off due to the presence of the fault. The result of this depends on whether the circuit is static or dynamic. In static circuits, the transmission gate is essentially used as a data selector. Typical applications include multiplexers, tri-state bus drivers and bistables. In all of these circuits, two or more transmission gate outputs are connected together and one is enabled to pass data to the common node. The transmission gate output is not generally set to a high impedance state for a significant time. As an example, we will consider the simple 2:1 multiplexer circuit shown in figure 7.12, in which a floating gate fault has been introduced in one of the n-channel transistors.

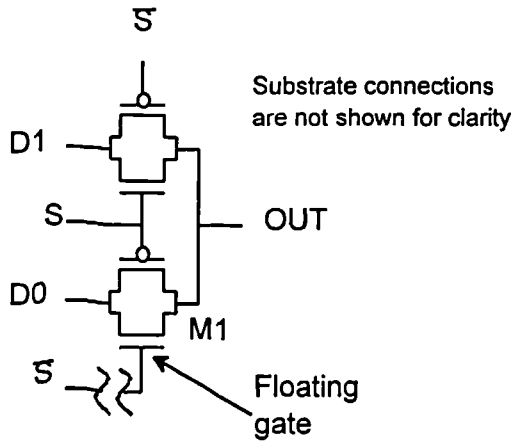


Figure 7.12 A multiplexer with floating gate fault introduced

For most values of V_{FG0} the lower transmission gate cannot be completely closed. Consequently, when the upper gate is enabled ($S = 1$), both transmission gates will attempt to drive the output node. A conflict will occur since the data inputs are different, but the fault free gate will always assert its logic value as the output impedance of this gate is lower than that of the faulty gate. A logical error will not occur and the fault will not be apparent. The supply current of the gates driving the data inputs will be raised when contention occurs.

Transmission gates are also used in dynamic circuits to store a logic value for a short period. Floating gate faults in such circuits may cause logical errors. This is because the transmission gate output cannot be isolated. Unlike the static circuits, there is no alternative

source of charge to maintain the required output value and so when the gate input changes, the output is charged or discharged and an error occurs. The error may not arise for a stored logic 0 and low V_{FG0} since the output can only achieve a maximum voltage of V_{FG} - V_{Tn}, which will probably be a logic 0 to following gates. However, in many cases, the fault would cause a clear logical failure.

In many cases, floating gate faults on transmission gates do not have a logical effect on circuit operation. Static circuits are particularly immune to this fault. Failures are however, quite likely to occur in dynamic circuits in which transmission gates are used to store logic values.

7.3.1 Dynamic Circuits and Floating Gate Faults

Dynamic circuits are generally more susceptible to logical errors arising from floating gate faults. A typical dynamic logic gate is shown in figure 7.13. The precharge transistor is used to charge the output to logic 1 whilst the control signal, *CK*, is at logic 0. When *CK* rises to logic 1, the precharge is switched off and the evaluate transistor is switched on. If a path exists through the evaluation network then the output is discharged to logic 0.

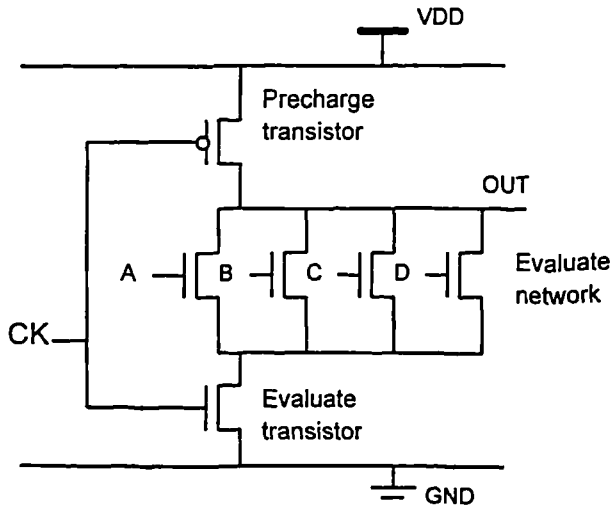


Figure 7.13 A dynamic four input NOR gate

If we introduce a floating gate fault into one of the *n*-channel devices in the evaluation network, the output may be discharged erroneously by the faulty device. This will only happen when all devices in series with the floating gate transistor are switched on. For low values of V_{FG0}, the output will discharge relatively slowly and so if the circuit is operating at high speed, the fault may not cause a failure. However, for V_{FG0} greater than approximately 0.5V, the fault would be apparent in most circuits. The resistance of the pull-down ratio will also affect the discharge speed. The effects of floating gate faults in the precharge and evaluate transistors will be similar to those for a NAND gate. Weak output voltages would

be produced and propagation delays increased for certain values of the initial floating gate potential.

7.4 Testing for Floating Gate Faults

The analysis discussed in the previous sections has provided a thorough understanding of the effect of floating gate faults on CMOS logic circuits. One of the aims of the work presented in this thesis is to identify techniques for the thorough testing of CMOS ICs. We have established that the floating gate fault is significant with wide ranging effects on different circuits. In Chapter 2 we discussed the fault models used to generate tests for CMOS circuits and outlined the application of these to different fault conditions. It is now necessary to correlate the effect of floating gate faults with these fault models in order to determine the effectiveness of tests based on these models. It should be possible to conclude the investigation by identifying the most reliable testing technique to be used for thorough coverage of floating gate faults.

7.4.1 The Stuck-At Fault Model

The stuck-at fault model assumes that faults cause a gate output or input to be stuck at logic 0 or 1. Tests generated using this model therefore attempt to assert logic 1 and 0 on each node in the circuit. If any node cannot be set to either state, then a fault exists. Certain floating gate faults do produce stuck-at fault behaviour. A connected floating gate fault on an inverter causes the output voltage to be stuck at a voltage between 0V and V_{DD}. The voltage will be interpreted as either logic 0 or 1 by the next level of gating. The fault would therefore be detected by a stuck-at 0 or stuck-at 1 test on the inverter output. The same test would also reveal that the input is stuck at the opposite logic value.

Connected floating gate faults would also be detected in multiple input gates when the faulty input is tested for stuck-at fault behaviour. The output of such a gate can change state via the fault free input, and so stuck-at 0 or 1 tests on the output would not necessarily reveal the fault.

Single floating gate faults are less likely to cause stuck-at fault behaviour. We have seen from test circuits and from simulation, that gate outputs may be set to either logic state by the faulty or fault free inputs. The *n*-channel floating gate fault in an inverter will cause the output to be stuck-at 1 for $V_{FG0} < 0.1$V. The output will be stuck-at 0 for values of V_{FG0} above approximately 9V. Both of these conditions are very unlikely to occur so that the *n*-channel floating gate fault is unlikely to be detected by stuck-at testing.

The *p*-channel floating gate fault also fails to prevent the output from adopting logic 1 or 0 for a wide range of initial floating gate potentials. However, for the observed range of

values for V_{FG0}, i.e. above V_{DD}, the fault may cause the output to be stuck-at 0. The speed of the test is a significant factor in the detection of such faults. At $V_{FG0} = V_{DD}$ (5V), the output rising delay is 17ns. Therefore, a test which samples the gate output less than 17ns after the application of the test vector will detect the fault. However, this would require a testing frequency of over 60MHz. Stuck-at testing is often performed at quite slow speeds which may be considerably slower than the design speed for a circuit. Tester speeds of 1MHz are not uncommon for the latest VLSI circuits [7.5]. Increases of the order of 17ns in the propagation delays of gates will not be detected by such tests. The *p*-channel floating gate fault is therefore only likely to be detected for V_{FG0} greater than approximately 6V for the samples in this study. At this voltage, the gate exhibits clear stuck-at fault behaviour rather than an increased propagation delay.

Single floating gate faults in multiple input gates are also unlikely to be detected by stuck-at testing. The ranges of V_{FG0} for which the tests would be successful will change with the pull-up to pull-down resistance ratio of gates but this will not have a significant effect on the stuck-at fault behaviour of a gate. Furthermore, the detection of floating gate faults by these tests may be invalidated by the previous state of the output.

Consider, for example, an *n*-channel floating gate fault in a two input NOR gate in which V_{FG0} is 0. This should be detectable by a test for the input stuck-at 0 because, when the output falls from logic 1, it can only reach approximately 3V before cut-off occurs in the *n*-channel device, leaving the output apparently stuck-at 1. This implies that the input is stuck-at 0 and the gate will fail the test. However, if the previous test vector had already set the output to logic 0, this would not change when the input stuck-at 0 test vector was applied and the gate would pass the test. This phenomenon has been recognised for some time for stuck-open faults and has resulted in the sequential test method described in Chapter 2. We now see that this is also a potential problem for floating gate faults.

In practice, the stuck-at fault model is not always rigorously applied. Integrated circuit manufacturers who use this model, do not normally require stuck-at tests for gate inputs *and* outputs. It is considered to be adequate to prove that a circuit node is not stuck. Since, in general, there are many ways to test each node in a circuit, this may leave many faults undetected. The stuck-at fault model can certainly detect a subset of floating gate faults but is not a sufficient model to guarantee to detect al such faults.

7.4.2 The Stuck-Open Fault Model

With stuck-open testing, the aim is to show that each pull-up and pull-down path for a gate can charge or discharge the output to logic 1 or 0. It is necessary, therefore, to ensure that the gate output is at logic 1 before testing for stuck-open faults in the pull-down path, and at logic 0 before testing the pull-up paths. An alternative name for stuck-open testing is

toggle testing. The name derives from the requirement of the test, that each gate output must be toggled from 0 to 1 and back to 0. It is necessary to do this for each gate input to prove each path in the networks, and to ensure that all parallel paths are disabled for each test.

Most floating gate faults actually result in the faulty device being closed (although not necessarily *stuck* closed) rather than open circuit. The only cases where stuck-open faults may occur are when the floating gate voltage is very low for *n*-channel devices (i.e. < 0V), and very high for *p*-channel (i.e. > V_{DD}). In these cases, stuck-open testing should reveal the fault. Consequently, very few single *n*-channel faults would be detected by stuck-open testing.

If we assume that V_{FG0} is above V_{DD} for *p*-channel faults as observed, then a significant proportion would be revealed by high speed testing. We should note that stuck-open testing cannot guarantee to detect all *p*-channel floating gate faults. The chance of detecting these faults, depends on the initial floating gate voltage and on the speed of testing. High speed testing is more likely to reveal the fault than slow testing at, for example, 1MHz. Stuck-open testing would reveal connected floating gate faults as it will not be possible to toggle the gate output via the faulty input.

Stuck open testing achieves slightly higher coverage of floating gate faults than stuck-at testing since it guarantees to detect cases in which the floating gate device is open circuit. However, the fraction of floating gate faults that will definitely be detected by stuck-open testing is still small.

7.4.3 The Stuck-Closed Fault Model

Stuck-closed testing generally assumes that devices are stuck in a "fully conducting" state, i.e. the gate voltage is logic 1 or 0 for *n*- and *p*-channel devices respectively. The result of this fault is an intermediate voltage on the gate output in the fault condition. This may represent a logical error but it is more probable that the fault will not produce a logical failure. If a failure does occur, then it will be detected by thorough stuck-at testing.

In general the stuck-closed fault produces a weak logic level which may be interpreted correctly by subsequent logic gates. In such cases the fault is not detected by any form of logic testing. In addition to the weak output voltage level, the stuck-closed fault causes the quiescent supply current to be raised significantly above the few nanoamps expected of a fault free gate. This will only occur for certain input conditions but may be used to detect the floating gate fault. This method is generally called I_{DDQ} (quiescent supply current) testing and was described in Chapter 2.

Thorough I_{DDQ} testing requires that a set of test vectors must be applied to reveal all failures. It is not adequate to simply power up the circuit and measure the supply current. If,

for example, the output of a gate with an n-channel floating gate fault is at logic 0, the fault would not be revealed. Ideally, a comprehensive test set is required to reveal all possible closed paths in each gate. Test generation is simplified in comparison with stuck-open testing, as it is not necessary to propagate the gate response to the output.

As already stated, most floating gate faults result in the faulty device being in a conducting state. We may therefore assume that the quiescent supply current will be raised in these gates and monitoring this will reveal the fault. However, the conductance of floating gate devices can vary significantly both during the circuit operation and with the initial floating gate potential. Consequently, the supply current may not be as high as expected.

Simulation of various gates with floating gate faults have been used to give an indication of the level of the supply current for various fault conditions. As an example, the graph in figure 7.14 shows the variation of the quiescent supply current with V_{FG0} for a minimum size inverter with an n-channel floating gate fault. The current is raised when the output is at logic 1 and is negligible when the output is at logic 0. The current varies from 0.8 μA to 320μA as V_{FG0} varies from 0 to 5V. For the measured range of V_{FG0}, the supply current varies from 10 to 65μA.

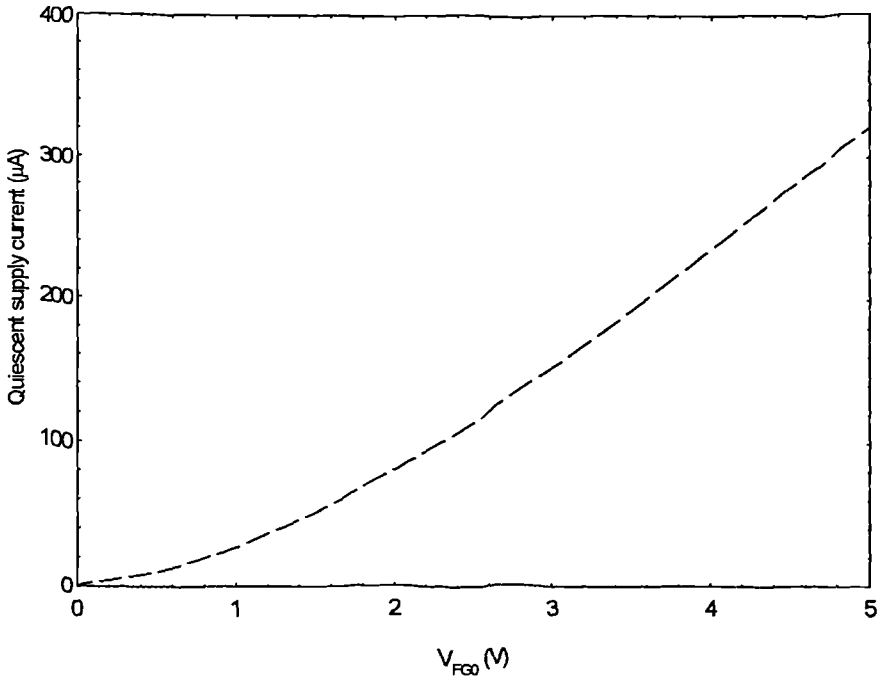


Figure 7.14 Variation of I_{DDQ} with V_{FG0} for an inverter with n-channel floating gate fault

The supply current for other circuits varies in a similar way. As expected, the current increases for wider devices but decreases for multiple input gates. It is important to note the wide variation in the current of over two orders of magnitude from V_{FG0} = 0 to 5V. If the floating gate device is assumed to be stuck-closed with a gate voltage of V_{DD}, the observed

current may be considerably lower than expected and cause faults to be missed. Measurements reported by Hawkins and Soden [2.56] also show a large variation in the current of inverters with connected floating gate faults.

The resolution of supply current monitors generally depends on the speed of testing. It is quite straightforward to measure quiescent currents of 10μA at reasonable speeds. Hawkins and Soden [7.6] suggest that this is possible at a test vector rate of 50kHz with external instrumentation. Resolution and speed can be improved by building current sensing circuits (BIC's) onto the chip. A BIC circuit reported by Welbers et al and cited in [7.6] can achieve a resolution of 2μA at a test speed of 2MHz. This is sufficient resolution to detect all floating gate faults that have been simulated except those which result in stuck-open behaviour.

In conclusion it would seem that power supply monitoring can provide a reliable method for detecting most floating gate faults. The only faults not detected by this method are those in which the residual gate charge and feedback coupling are insufficient to switch the faulty device on. Such faults would be detected by stuck-open testing and so both techniques are required to ensure detection of all floating gate faults. It is also important that the logical testing is performed at the rated operating speed of the circuit concerned. Reduced testing speeds may allow faulty circuits to pass undetected.

7.5 The Implications of Inadequate Testing

We have seen in the previous section that detection of all floating gate faults can only be guaranteed if both stuck-open and I_{DDQ} testing are used to verify a chip. There are many circumstances in which simple stuck-at or stuck-open testing would not detect floating gate faults. This is even true if tests are performed at high speed, for many circuits which include floating gate faults have a negligible increase in propagation delay and no logical error. Clearly if this is the case, we must consider whether it is important to detect such "faults". If they have no detectable effect on the logical operation of a circuit, why should they be a problem?

To answer this question we must consider the reliability of components. The users of integrated circuits require that chips should operate within specification for the lifetime of the system in which they are used. It is not, of course, adequate for a manufacturer to state that a circuit was working when it was tested but then to have no idea how long the chip will continue to work. The user must have confidence that the circuit will operate reliably for the required period of time.

For a manufacturer to be able to predict reliable performance, he must have knowledge of the failures that are likely to occur during the lifetime. Many of these are

instrinsic failures such as oxide breakdown or electromigration. The effects of these can be minimised by appropriate design rules, but they cannot be eliminated entirely. A further source of potential failure during the lifetime of a component is undetected production faults. These may result from inadequate testing requirements or from inadequate fault models. If, for example, a manufacturer requires 95% stuck-at fault coverage for a circuit, he must accept that a significant number of faulty chips will be supplied to the user and that consequently, failures will occur during the operation of circuits.

The problem may not be eliminated by simply improving the fault coverage of a test set for a circuit. A manufacturer who guarantees 100% stuck-at fault coverage cannot guarantee detection of all possible faults. The stuck-at model will miss many stuck-open, bridging, stuck-closed and floating gate faults. Once again, faulty chips may pass the "thorough" production testing and cause lifetime failures in the field.

Unfortunately the problem is not necessarily solved by improving the fault model. Ultimately, the model is only a *representation* of the real effect of faults and so is limited in its ability to reveal physical defects. Even with the most sophisticated models, faults will occur in circuits that are not detectable by the tests used. Such faults have the potential to cause a failure in the circuit but the testing technique was inadequate to reveal the fault. Another form of undetected fault is one which simply does not cause a failure at the time of testing. These might be called *latent* faults. Certain floating gate faults come into this category, allowing fault free logical operation at design speed. Should we be concerned about such faults?

Latent faults are not a problem provided that their characteristics do not change with time. Consider the following scenario. A CMOS circuit is tested at design speed with vectors derived using a thorough and reliable technique based on the stuck-open model. This circuit may contain single *n*-channel floating gate faults. Such a fault with an initial floating gate potential of around 1.5V will not be detected by the tests described and the circuit will operate correctly at full design speed. (The supply current will be raised for some states of the chip, but we are not considering the use of I_{DDQ} monitoring).

During the investigation of floating gate potentials on test devices, a wide spread in the gate voltage was observed for identical devices. It was concluded that this was due to charge leakage from the gate during storage. It was shown that extremely small leakage currents are required to change the gate voltage by one or two volts. It is reasonable to assume that this charge leakage will continue during the operation of circuits. In fact, with the presence of electric fields in the devices, leakage is likely to be more significant. We can therefore expect the charge on the floating gate device to change with time and hence for the characteristics of the faulty device to alter.

There are two possible outcomes to this change. The floating gate charge may become more positive. In this case the circuit will continue to operate at design speed but the supply current will increase. Alternatively, the positive gate charge may decrease. This will result in an increased output falling propagation delay for the faulty gate which would ultimately cause a failure of the circuit. The presence of the undetected floating gate fault therefore results in an increased probability of failure for the circuit and hence a reduced reliability.

It is of course unrealistic to demand complete detection of all possible physical defects in an integrated circuit. This is particularly true as the scale of integration increases and we commonly see VLSI circuits with over 3 million transistors. Simply because of the complexity and size of the problem, it is not possible to derive test patterns that can guarantee 100% toggle coverage of such circuits, let alone 100% physical defect detection. In such circumstances it is necessary to ask the question "How much testing is enough?". This question can only be answered by the manufacturer and the user as the answer depends on many factors that cannot be generalised, such as the application, cost of repairs, cost of test and production, desired confidence in the product, and many more. It is not the aim of this thesis to answer such questions. We simply aim to provide the means of ensuring the highest possible fault detection that economics will allow.

We might reasonably ask how much improvement in fault coverage is achieved by the proposed test strategy for floating gate faults. We will make an estimate of this, although a large number of assumptions are necessary.

- We will assume that 15% of photolithography defects produce floating gate faults. This is the lower limit calculated from the fault distribution experiments reported in Chapter 3.
- We will also assume that photolithography defects are the dominant source of failures located in the production testing of packaged ICs. We are therefore assuming that parametric tests on the wafers, visual inspection and other tests have eliminated more general or widespread failures.
- From the distribution of fault types obtained, we can estimate that around 5% of photolithography defects will result in *n*-channel floating gate faults. This includes connected *n*-channel faults which arise when more than one *n*-channel gate is floating but all are connected together. This accounts for many of the connected floating gate faults in, for example, the Multiply cell.
- We will further assume that the distribution of the initial floating gate potential is as measured on the test samples. This will result in 32% of the defects being detected by high speed toggle testing. This means that 68% of the *n*-channel faults will remain undetected.

- We will also assume that all *p*-channel and *n*- and *p*-channel connected floating gate faults are detected.

Based on the above assumptions, we obtain a figure of approximately 3.5% for the number of undetected photolithography defects which result from floating gate faults. Implementation of I_{DDQ} and stuck-open testing as recommended would result in all of these defects being identifed at production test. It is felt that this figure is not an overestimate of the improvement that may be achieved. Indeed, it is likely to be higher than sugested, as some of the *p*-channel faults may not be detected. Furthermore, we have assumed that toggle testing by all gate inputs at full design speed is used. Not all production testing is this thorough.

7.6 Chapter Summary

In this chapter we have investigated the effects of floating gate faults on a range of typical CMOS logic circuits. The effects of the fault are seen to be quite subtle in many cases. The accurate model derived to represent the fault is necessary to obtain a realistic assessment of the consequences of floating gate faults. The simplistic assumptions made by some workers in the fields of fault modelling and test generation are clearly inadequate and result in rather optimistic prediction of the detection rate of the fault. In particular, the increase in propagation delays that might be expected are insignificant in many cases and do not cause logic failures.

Testing techniques to reveal the floating gate faults have been identified. Standard CMOS test methods such as stuck-at or toggle testing have been shown to be inadequate for many instances of floating gate faults. To ensure detection of the faults, stuck-open testing must be combined with power supply monitoring or I_{DDQ} testing. This form of testing can reveal most cases of floating gate faults but it is important to note that adequate test vector sets are required to ensure that these faults are revealed. Simple static measurement of the power supply for a single input condition, which is normally done for CMOS chips, does not constitute thorough I_{DDQ} testing. This field needs further refinement before it can be claimed to be the solution to this, and many other testing problems. An estimate of the improvement in fault coverege that may be achieved by adopting the recommended test strategy has been made. The improvement is seen to be quite significant.

Chapter 8
Summary and Conclusions

In this chapter we will review the analysis, results and conclusions that have been presented in this thesis. Based on this, we will consider the wider implications of the work presented and draw some general conclusions. As is the case for all research work, the solution of a problem, inevitably raises new questions to be answered. We will therefore suggest some future avenues of research in the area of faults and testing which have arisen from this work.

8.1 A Summary of the Thesis

The aims of the work presented in this thesis were to show that current testing techniques for CMOS circuits are inadequate for a significant proportion of possible faults. The consequence of this is that the number of lifetime failures of ICs is higher than is necessary. We aimed to suggest improvements in the testing techniques currently used and hence to improve the reliability of CMOS integrated circuits.

To uncover any inadequacies in current testing techniques, it is necessary to understand the causes of faults in CMOS circuits. A detailed analysis of failure mechanisms and fault models was therefore performed to identify the areas of IC test that may be inadequate. Production and lifetime defects were considered and many possible faults were found to be adequately covered by standard testing methods. This is not surprising, as there is a great deal of theoretical knowledge and practical experience in the literature and industry, which has resulted in the current state of testing methods.

However, the analysis did reveal a number of significant forms of physical defect that are not covered by standard tests. Gate oxide shorts and hot electron effects are examples of defects which are not necessarily covered by stuck-at or stuck-open testing. They tend to produce parametric, rather than logical, faults in a circuit. Examples of this kind of fault are timing errors and increased supply current. These faults have received considerable attention in the literature. Indeed, hot electrons are seen as one of the most serious failure mechanisms for future devices.

Another fault which tends to produce non-logical fault effects is the open-circuit signal line or floating gate fault. This has received very little attention in the literature and yet it is an important problem. Most authors who consider the fault make very simplified assumptions about its effect. These range from assuming the device to be stuck-open to stuck-closed. There has been no theoretical analysis of this type of failure published and the significance of the fault suggests that detailed analysis and understanding would help the aim of improved testing for CMOS circuit.

The floating gate fault is investigated in detail in Chapter 3. Causes of the fault were identified in the production environment and for operation in the field. To attempt to quantify the significance of the fault, we have presented theoretical analysis of fault distributions in various CMOS circuits. This is based on scattering photolithographic defects on cell layouts and observing the resulting effects in terms of types of fault. The analysis showed that floating gate faults have a high probability of occurring during a photolithographic process, using negative resist. Typically, between 15 and 20% of such defects result in a floating gate fault. The sources of floating gates which might arise during the operation of circuits were also investigated. A theory is presented to show that asymmetric timing characteristics in CMOS circuits can lead to electromigration in signal tracks. This mechanism, along with contact and stress migration, can produce floating gate faults during the operation of ICs.

A simple model is used to demonstrate the effect of the fault in CMOS circuits. This is shown to be dependent on the charge on the floating gate. Sources of charge are identified and we see that the potential of the floating gate is difficult to predict. This is also the case for floating gates which occur during operation. The unpredictability of the floating gate voltage leads to considerable variation in the effect produced by the fault. To assist in the understanding of the operation of floating gate devices and circuits, a set of test circuits was designed and fabricated for experimental analysis.

Attempts were made to obtain direct information on the floating gate device by use of a scanning electron microscope, as described in Chapter 4. The voltage contrast technique was investigated as a means of obtaining a direct measure of the floating gate potential. The technique was not found to be suitable with the equipment available. Measurements of the floating gate voltage could not be made because of the limited resolution of the SEM under the operating conditions necessary for these devices. However, it was demonstrated that intensity analysis can be used quite satisfactorily for quantitative analysis of ICs where high resolution measurements are not required.

Analysis of the test devices and circuits, described in Chapter 5, enabled the characteristics of floating gate devices to be obtained. Some data was obtained on the initial floating gate voltage. This implies that the gate charge is positive for both *n*- and *p*-channel devices for the process used in this work. A considerable variation in this parameter is

observed for otherwise identical devices. This may result from non-uniform charge distribution during processing (i.e. when the gate charge may be deposited), or from variable charge leakage rates in devices. The consequence is that it is difficult to predict device performance. Characteristics such as the variation of drain-source current with applied voltage can vary by two or three orders of magnitude, depending on the value of the gate charge.

Another significant parameter is the coupling ratio, α. This determines the fraction of the drain and source potentials which is added to the initial gate voltage. The coupling ratio is seen to vary with initial gate voltage and with channel length and width. There is also some variation during the operation of the devices. The initial floating gate voltage and coupling ratio determine the operating characteristics of floating gate transistors, and their variability makes prediction of device performance difficult.

Experimental investigation of circuit performance with floating gate faults concentrated on the static characteristics of inverter circuits. This revealed that in most cases the fault would not be detected by static logic testing of the gates. The supply current for these circuits was increased, as expected, but it was not as high as simple models for this fault would suggest. Analysis of multiple input gates was hampered by a subtle design fault which lead to an interesting discovery relating to a different fault condition.

It is generally assumed that open circuit faults in power supply tracks will result in simple stuck-at faults in the affected circuits. This is seen not to be the case if substrate contacts still remain connected to the faulty gates. These will provide a path for the supply current for the gates, allowing them to operate quite satisfactorily in many cases. It is quite probable that the gates will be susceptible to latch-up failures due to the large transient substrate currents that may flow, but this will not always occur. The supply stuck-open fault must therefore be treated with caution. Stuck-at or stuck-open testing cannot be guaranteed to detect this form of failure. This discovery also has implications for circuit design in general. It is not inconceivable that switches may be inserted into supply lines as a form of power management in future ICs. We must ensure that such circuits account for the phenomenon described to prevent hazardous operation of circuits.

The results of the analysis of floating gate faults in multiple input gates confirmed the measurement made on inverters. An interesting consequence of this relates to the value of the initial floating gate potential. The samples which contained the multiple input gates and those which contained the inverters were obtained from different runs of the fabrication process used. Despite this, the observed range for the floating gate potentials were very similar. This may suggest that the measured values can be extrapolated to other devices, i.e. that they represent a fairly general result. However, this conclusion would require further investigation before it could be stated with much confidence.

The electrical analysis provided useful data but indicated the need for a detailed understanding of the operation of the floating gate device. To facilitate this, theoretical analysis of an MOS transistor with a fixed gate charge was performed. The aims of this were: a) to provide greater understanding of the device operation and of the experimental results obtained; and b) to obtain a relationship between the floating gate and the drain and source potentials. The latter aim should allow a model to be created for the prediction of circuit performance with floating gate faults, one of the overall aims of this work.

The theoretical analysis is described in Chapter 6. By considering a simple capacitance model for the device, a basic relationship between V_{FG} and V_{DS} was derived. This expression was then refined by analysis based on the charge distributions in the device. This provides a useful understanding of the physics of the device operation. The resulting equations cannot be solved analytically and a computer program was written to provide numerical solutions. The program can calculate the floating gate potential for a given drain potential and initial floating gate voltage. The device parameters can be varied to allow investigation of a range of devices and processes.

The model predicts considerable variation of the floating gate voltage with the drain or source voltages. This may cause devices which are initially switched off to become strongly conducting as the drain potential is increased. It is clear from this result that the floating gate device cannot be considered to be stuck in any particular conducting state. The model also predicts a variation in the coupling ratio, α with the initial gate voltage and with V_{DS}. The predictions of the model are compared with the measured characteristics and reasonable agreement is found for a range of device parameters.

The theoretical and experimental analysis described provides a basis for the investigation of the effects of floating gate faults on CMOS circuit operation. Through this, it has been possible to demonstrate that standard CMOS test techniques are insufficient to reveal some floating gate faults. It has also been possible to identify satisfactory tests for these failures. Investigation of the effects of floating gate faults on the operation of CMOS circuits is described in Chapter 7. In addition to simple analysis, SPICE simulations are used to help to quantify the effects. A novel SPICE model of the floating gate fault is developed and verified by comparison with previous experimental and theoretical analysis. This was used to make detailed investigation of the effect of the fault on the static and dynamic behaviour of CMOS circuits.

The results of this investigation support the hypothesis that current testing strategies are inadequate for floating gate faults. In many instances, the characteristics of logic gates are not significantly affected by floating gate faults. The output voltage of a gate is generally degraded for one logic state, but, in most cases, this is insufficient to cause a logical error. The fault also degrades the dynamic behaviour of gates, causing propagation delays to

increase. However, this is only significant for a small proportion of the possible floating gate faults that may arise. In general, standard CMOS tests cannot be guaranteed to detect floating gate faults, even when tested at full operating speed.

One characteristic that is common to most circuits with floating gate faults is that they have a high quiescent supply current in certain logic states. This fact can be exploited for the detection of the fault and it is recommended that I_{DDQ} testing is used to screen for floating gate faults. This is sufficient to cover most cases, but stuck-open testing is needed to provide full cover of all floating gate faults. The latter covers the fault when the floating gate charge and the device coupling are insufficient to switch on the transistor. The combination of I_{DDQ} and stuck-open testing is therefore recommended as a means of improving coverage of faults in CMOS circuits.

8.2 Conclusions of the Thesis

In this thesis we aimed to show that current testing techniques used for CMOS ICs are inadequate for detecting a significant proportion of physical defects. This was demonstrated convincingly and the floating gate fault was identified as an important class of fault that must be considered for thorough testing and reliable circuits. We have shown quantitatively that the fault has a relatively high probability of occurring as the result of production defects. We have also identified sources of this fault which may result in lifetime failures of circuits.

The effect of the fault on circuit operation has been shown, both experimentally and theoretically, to depend on the charge isolated on the floating gate. In many cases the fault does not significantly affect the logical operation of circuits. These two conclusions imply that existing tests which are based on stuck-at and stuck-open testing will fail to detect a significant number of production defects. The unpredictable nature of the fault makes it difficult to quantify the fraction of defects that ~~not~~ cannot be detected by conventional testing. However, a simple estimate, based on realistic assumptions, suggests that between three and four percent of photolithography faults will result in undetected *n*-channel floating gate faults if stuck-at and stuck-open testing is used.

A testing strategy has been identified to improve fault coverage. This requires quiescent power supply monitoring with a resolution of approximately 1μA, and stuck-open testing to guarantee detection of all faults.

In addition to satisfying the main aims of this research, several other achievements have been reported in this thesis.

- The residual charge on a floating gate transistor, which results from a production defect, was positive for all devices measured. Furthermore, this was found to increase approximately linearly with the length of the gate periphery. There is some evidence that

the floating gate charge decays to a value corresponding to an initial gate voltage of less than two volts for all devices.

- A theoretical model of the operation of the MOS transistor with fixed gate charge has been developed.
- It has been demonstrated that quantitative voltage contrast can be performed with simple intensity analysis in the scanning electron microscope. The limitations of this method have been found.
- A mechanism for electromigration in signal lines has been postulated and justified. This failure can result in floating gate faults which occur during the operation of ICs.
- It has been demonstrated that the stuck-open supply fault may not result in a stuck-at failure. Affected circuits can continue to operate correctly although they may be susceptible to latch-up problems.

8.3 Suggestions for Further Work

The work presented in this thesis has answered many questions concerning faults in CMOS circuits and, in particular, the floating gate fault. Inevitably, many other questions have been raised, and some of these would form a useful basis for further research in this field.

The initial floating gate voltage, which is determined by the residual charge on the gate, was found to be an important parameter in determining the characteristics of the floating gate device. It would be most useful to identify, experimentally, the source of this charge. The variation of the charge with time is also an important factor, as it has implications for the reliability of ICs in which floating gate faults remain undetected, and this should be investigated. Experimental verification of the theory concerning electromigration in signal lines would also help in determining the effect of floating gate faults on IC reliability.

Quiescent current supply monitoring has been identified as a useful technique for the detection of many parametric defects. More work is needed to find built-in current monitor circuits that can be used to detect field failures. In particular, the circuits must have sufficient resolution and be compact to be useful in built-in-self-test schemes. Furthermore, they should be able to detect the location of a defect, to allow reconfiguration schemes to operate efficiently.

More generally, it has become apparent that other forms of defect are not detected by standard test techniques. Failures caused by bridging faults and hot electron injection appear to be the next most significant form of fault to be considered and reliable techniques for the detection of these failures still have to be identified.

8.4 Some General Conclusions

For practical and economic reasons, the testing of integrated circuits cannot be perfect. Whatever level of refinement is used for test generation models, some physical defects will have a significant probability of passing all tests. Clearly the manufacturer must ask the question: "How much test is enough?". As indicated earlier, the answer to this question depends on many factors outside of failure physics and reliability calculations and the test engineer must always consider the high cost of both test generation and application.

However, it is important that testing is as thorough as possible. If improvements can be made with little extra cost (or even a reduction in cost), they are worth pursuing. The work presented in this thesis shows that the addition of thorough I_{DDQ} testing to the testing methods used by manufacturers, should improve the detection rate of faulty circuits which will improve component reliability and hence reduce costs in the long term. The cost of implementing this test may not be too high and it has the benefit of identifying other failures in addition to floating gate faults.

There is an increasing use of BIST structures in circuits as the level of integration increases. Consequently, built-in-current monitor circuits must be developed that are compact and have the required performance. They must also be reliable - testing the tester is always a problem. We can only hope to move ever closer to the asymptote of 100% testing of ICs. I hope that this thesis has moved the state of the art somewhat closer.

Appendix A

Program Listing of the FLGMOD Program

```
{ ************************************************************************

  Program to model the floating gate MOS transistor.
  Written by S. Johnson, University of Durham

  This program implements the model of a floating gate MOS transistor
  developed in the Ph.D. thesis, "Modelling and Analysis of Failures
  in CMOS Integrated Circuits", by S. Johnson.

  The program is written in Turbo Pascal and is compatible with
  Versions 4 through to 7 of this language. The program should run on
  any PC runing DOS 3.1 or higher.

  ************************************************************************ }

program flgmodel16;

uses crt, graph, plotgrph, drivers;

const   Esi = 1.0443E-10;                          { Fundamental constants }
        Eox = 3.4515E-11;
        q = 1.6022E-19;
        k = 1.38E-23;
        ni = 1.45E16;
        T = 300;

        xgrids = 51;                               { Program constants }
        Q_tol = 0.001;                             { Convergence tolerance constants: }
        Vx_tol = 1e-4;
        alpha_tol = 1e-4;
        err_tol = 0.01;
        debug = false;                             { Debug flag }

var     Cox, Cgd, Cgs, Cdep0,                      { Capacitances in the MOST }
        Phif, Phis, alpha, L, W, Lp, gamma,        { Device variables }
        Vt0, Vt, Vgs, Vds, Vg, Vgst, Vds_sat, Ids, { Device variables }
        vgs_min, vgs_max, vgs_inc, vds_min, vds_max, vds_inc, { Loop variables }
        Qtotal, Qgate, Qgate0, Qstotal, Qbtotal,   { Device charges }
        Na, Vfb, mu, Tox, Length, dL, Width, dW, VFG, { Device parameters }
        last_qdiff, last_alpha, last_Vxt, last_err, { Iteration variables }
        temp, a_frac
                                          : real;
                                                   { Program flags }
    cannot_converge, converged, Vx_converged,
      Vx_cannot_converge, non_convergence, linear,
      inv, newcalc, newplot, save_data     : boolean;
                                                   { Loop variables }
    i,j,d,sign, num_vds : integer;
    pfile : text;                                  { Parameter file variables }
    pf_name : string[12];
    npf_name : string[8];
                                                   { Arrays for plotting data }
    Vx, Qg, Qb, Qs, Cx, Cdeps, Ctotal, xdata, ydata, idata,
      qsdata, qbdata, qgdata, adata, vgdata, vdata : datapoints;
```

```
procedure waitforkeypress;

{ Wait until key has been pressed and clear keyboard buffer }

var dummy : char;
begin
  repeat until keypressed;
  dummy := readkey;
end;

procedure graphics;

{ Put screen into detected graphics mode, ega or vga expected }

var gd, gm : integer;
begin
  gd := detect;
  initgraph(gd,gm,'');
end;

function rt3_2(x : real) : real;

{ Returns square root of parameter cubed }

begin
  rt3_2 := sqrt(x*x*x);
end;

function find_Qgate : real;

{ Returns the charge on the gate over the channel region by
  subtracting the charge on the overlap regions from the total
  gate charge }

var  Qgate : real;
begin
  Qgate := Qgate0 - Vgs*Cgs - (Vgs - Vds)*Cgd;
  find_Qgate := Qgate;
  if debug then
    writeln('Qg,Qg0,QCd,QCs: ',Qgate:10,Qgate0:10,Vgs*Cgs:10,(Vgs-Vds)*Cgd:10);
end;

function find_Phis(v : real):real;

{ Returns the surface potential for the gate potential passed as v }

var eta : real;
begin
  if v < Vt0 then
    begin
      eta := 2*Esi*q*Na*Tox*Tox/(Eox*Eox);
      find_Phis := v-Vfb - eta*(sqrt(1+2*((v-Vfb)-2*k*T/q)/eta)-1);
    end
  else
    find_Phis := 2 * Phif;
end;
```

```
procedure calc_consts;

{ Calculates device variables for parameters from the setup screen.
  These will be constant for this particular caluculation run }

begin
  L := Length - 2 * dL;
  W := Width - 2 * dW;
  Cox := Eox / Tox;
  Cgd := dL * W * Eox / Tox;
  Cgs := dL * W * Eox / Tox;
  Phif := (k * T / q) * ln(Na / ni);
  gamma := sqrt(2 * Esi * q * Na) / Cox;
  Vt0 := 2*Phif + gamma * sqrt(2*Phif) + Vfb;
  Phis := find_Phis(VFG);
  Cdep0 := W*L/2*sqrt(2*Esi*q/Phis*Na);
  Qgate0 := VFG*Cgs + VFG*Cgd + W*L*Cox*(Vfg-PhiS);
end;

procedure find_Vx(Vds, Vg : real);

{ Calculates the potential along the channel using interval halving }

var Vxt, x, err, a, b, c, e, Vd : real;
   i : integer;

begin
  linear := true;
  Lp := 0;
  Vx[0] := 0;
  xdata[0] := 0;
  Qtotal := 0;
  qbtotal := 0;
  qstotal := 0;
  Vd := Vds;
                                  { Find the threshold voltage for the
                                    current gate voltage and test for
                                    saturation }

  Vt := Phis-Vfb-0.5*gamma*gamma*(1-sqrt(1+4*(Vg-Vfb)/(gamma*gamma)));
  Vds_sat := Vg - Vt;
  if (Vds >= Vds_sat) then          { If saturated, find the length of the
                                    pinch-off region using equation 6.40 }
   begin
    linear := false;
    Lp := 1/(1/(sqrt(2*Esi/(q*Na)*(Vds-Vds_sat)))+(Eox/(Esi*Tox)*
            ((0.2*(Vds-Vgs)+0.6*(Vgs-Vds_sat))/(Vds-Vds_sat))));
    if Vds_sat < 0 then
       Vd := 0
    else
      Vd := Vds_sat;
   end;
  for i := 0 to xgrids do           { Find Vx for each segment along the channel }
  begin
   x := L * i / xgrids;
   xdata[i] := x;                   { Check to see if x is within the pinch-off region }
   if (x > (L - Lp)) and (not linear) then
    begin                           { If x is in the pinch-off region, use
                                      linear interpolation to find Vx, and set
```

```
                                        surface charge to 0 }
    Vx[i] := Vds_sat + (x-(Length-2*dL-Lp))*(Vds-Vds_sat)/Lp;
    Qs[i] := 0;
   end
  else
  begin                                 { If x is in inverted region, use interval halving }
  last_Vxt := Vd;
  last_err := 0;
  Vxt := 0;                             { Variable for trial }
  Vx_converged := false;
  Vx_cannot_converge := false;
  sign := 1;                            { Flag for increment or decrement of interval }
  if Vds >= Vds_sat then
    e := x/(2*(L-Lp))*Vds_sat*Vds_sat   { If device is saturated, use
                                          simplified form of equation 6.33 }
  else
    e := x/(L-Lp)*(Vd*(Vg-Phis) - Vd*Vd/2 -
                  2/3*gamma*(rt3_2(Phis+Vd)-rt3_2(Phis)));
  if Vgs >= Vt then                     { Check to see if channel exists }
   repeat
    a := Vxt*Vxt/2;
    b := 2/3*gamma*(rt3_2(Phis+Vxt)-rt3_2(Phis));
    c := Vxt*(Vg-Phis);
    err := a + b - c + e;               { Calculate equation 6.33 for the trial values
                                          of Vgs and Vx, for the current value of x }
    if debug then writeln('err= ',i,err:8);
                                        { If the error is less than the convergence
                                          tolerance set in program, then this
                                          iteration is complete }
    if abs(err) < err_tol then Vx_converged := true
      else
      begin                             { If error is too great, find new estimate for Vx }
       temp := Vxt;
       if last_err * err < 0 then sign := -sign;
       Vxt := Vxt + sign*abs(Vxt-last_Vxt)/2;
       Vx_cannot_converge := abs(Vxt-last_Vxt) < Vx_tol;
       last_Vxt := temp;
       last_err := err;
      end;
   until Vx_converged or Vx_cannot_converge
                                        { Repeat iteration until solution found
                                          or increment is less than tolerance
                                          - convergence failure }
  else
    Vxt := 0;                           { If Vgs < Vt then set Vx to 0 }
  if cannot_converge then               { Notify user of convergence failure }
    begin
      non_convergence := true;
      if true then
      writeln('No convergence for Vx at: Vds=',Vds:5:2,' VFG=',VFG:5:2,
              ' Vgs=',Vg:6:4,' x=',x:8,' alpha=',alpha:5:2);
      Vxt := 0;
    end;
  Vx[i] := Vxt;
                                        { Find the charge in the surface region
                                          for the current segment and sum it }
  Qs[i] := W*L/(xgrids+1)*Cox*(Vg-(Phis+Vx[i])-gamma*sqrt(Phis+Vx[i]));
  qstotal := qstotal + Qs[i];
  end;
```

```
    if false then
      writeln('Vx,Phis,qs,qb,Qg',Vx[i]:10,Phis:10,qstotal:10, qbtotal:10,Qtotal:10);
                                          { Find the depletion (bulk) and gate charges
                                            for the current segment and sum }
    Qb[i] := W*L/(xgrids+1)*Cox*gamma*sqrt(Phis+Vx[i]);
    qbtotal := qbtotal + Qb[i];
    Qg[i] := W*L/(xgrids+1)*Cox*(Vg-(Phis+Vx[i]));
    Qtotal := Qtotal + Qg[i];
  end;                                    { End of loop for current segment.
                                            Repeat above calculations for each
                                            channel segment }
  if debug then
     writeln('Phis,qs,qb,Qg',Phis:10,qstotal:10, qbtotal:10,Qtotal:10);
end;

procedure find_Vgs;

{ Calculates the floating gate voltage for the current drain voltage }

begin
    converged := false;
    cannot_converge := false;
    alpha := 0.5;                         { Initial estimate for alpha }
    last_alpha := alpha * 2;
    last_qdiff := 0;
    a_frac := -0.5;                       { Interval halving constant }
    repeat
      Vgs := VFG + alpha * Vds;           { Find current estimate for gate voltage }
      Qgate := find_Qgate;                { Find the charge on the gate over the channel }
      Phis := find_Phis(Vgs);             { Find the surface potential for the current
                                            estimate of Vgs }
      find_Vx(Vds,Vgs);                   { Find the channel potential and charge
                                            distributions }
      if debug then
        writeln('Vgs,Vds,Vt,alpha,Qtotal,Qgate: ',Vgs:6:3, Vds:4:1, Vt:5:2, alpha:6:3, Qtotal:9, Qgate:9);
                                          { Compare the gate charge over the channel
                                            with the total charge in the surface
                                            and bulk depletion regions. If less
                                            than the specified tolerance, calculation
                                            is complete. If not find new estimate
                                            for alpha and repeat calculation }
      if abs((Qtotal - Qgate)/Qtotal) <= Q_tol then converged := true
       else
        begin
        temp := alpha;
        if (Qtotal-Qgate) * (last_qdiff) < 0 then a_frac := -a_frac;
        alpha := alpha + a_frac*abs(alpha-last_alpha);
        cannot_converge := abs(alpha-last_alpha) < alpha_tol;
        last_alpha := temp;
        if debug then
           writeln(qtotal-qgate:11, last_qdiff:11, a_frac:4);
        last_qdiff := (Qtotal-Qgate);
        end;
    until converged or cannot_converge;
    if converged then                     { If converged, collect data for
                                            current value of Vds }
      begin
        vgdata[i] := vgs;
        adata[i] := alpha;
```

```
        qbdata[i] := qbtotal;
        qsdata[i] := qstotal;
        qgdata[i] := Qtotal;
      end
    else
      begin                                  { If no convergence, notify user }
        if true then
        writeln('No convergence for Vgs at: Vds=',Vds:5:2,' VFG=',VFG:5:2,
              ' Vgs=',Vg:6:4,' alpha=',alpha:5:2);
        non_convergence := true;
        vgdata[i] := 0;
        adata[i] := 0;
        qbdata[i] := 0;
        qsdata[i] := 0;
        qgdata[i] := 0;
      end;
end;

procedure calc_Ids;

{ Calculate the drain-source current for current values of Vgd and Vds.
  Uses second order equations with improved pinch-off model due to
  Frohman-Bentchkowsky and Grove }

var vd,vt : real;
begin
  Lp := 0;
  Vd := Vds;
  Vt := -Phis-Vfb+0.5*gamma*gamma*(1-sqrt(1+4*(Vgs-Vfb)/(gamma*gamma)));
  Vds_sat := Vgs + Vt;
  if debug then writeln(Vgs, Phis, gamma, Vds_sat);
  if Vgs >= Vt then
  begin
    if Vds >= Vds_sat then
     begin
      Lp := 1/(1/(sqrt(2*Esi/(q*Na)*(Vds-Vds_sat)))+(Eox/(Esi*Tox)*
            ((0.2*(Vds-Vgs)+0.6*(Vgs-Vds_sat))/(Vds-Vds_sat))));
      Ids := mu*Cox*W/(2*(L-Lp))*(Vds_sat*Vds_sat);
     end
    else
      Ids := mu*Cox*W/(L-Lp) * ((Vgs-Vfb-Vt0-Vd/2)*Vd - 2/3*gamma*
              (rt3_2(Vd+Phis)-rt3_2(Phis)));
  end
  else
    Ids := 0;
  Idata[i] := Ids;
  vdata[i] := Vds;
end;

procedure make_data_files;

{ Creates text data files of alpha, Vgs and Ids for range of Vds
  File names are <current parameters file name>[a/v/i].dat }

var     afile, vgsfile, idsfile : text;
        i : integer;
begin
    assign(afile,concat(npf_name,'a.dat'));
    assign(vgsfile,concat(npf_name,'v.dat'));
```

```
    assign(idsfile,concat(npf_name,'i.dat'));
    rewrite(afile); rewrite(vgsfile); rewrite(idsfile);
    writeln(afile,'Vds        Alpha');
    writeln(vgsfile,'Vds        Vgs');
    writeln(idsfile,'Vds        Ids (uA)');
    for i := 0 to num_vds do
      begin
        writeln(afile,vdata[i]:6:2, adata[i]:7:3);
        writeln(vgsfile,vdata[i]:6:2, vgdata[i]:7:3);
        writeln(idsfile,vdata[i]:6:2, idata[i]*1e6:8:3);
      end;
    close(afile); close(vgsfile); close(idsfile);
end;

procedure plot_graphs;

{ Plot graphs of Vx, Qb(x), Qg(x), Qs(x) against x for final Vds value,
  and Ids, Vgs, alpha, Qb, Qg, Qs against Vds }

begin
  graphics;
  set_range(0,0,0,0);
  plot_graph(0,0,319,239,xdata,Vx,xgrids+1,'x','Vx','Channel Voltage',true);
  set_range(0,0,0,20e-16);
  plot_graph(320,0,319,239,xdata,Qb,xgrids+1,'x','Qb','Depletion charge',true);
  plot_graph(320,240,319,239,xdata,Qg,xgrids+1,'x','Qg','Gate charge',true);
  plot_graph(0,240,319,239,xdata,Qs,xgrids+1,'x','Qs','Channel charge',true);
  waitforkeypress;
  cleardevice;
  set_range(0,0,0,0);
  plot_graph(0,0,319,239,vdata,idata,num_vds+1,'Vds','Ids','Ids v. Vds',true);
  set_range(0,0,0,0);
  plot_graph(320,0,319,239,vdata,vgdata,num_vds+1,'Vds','Vgs','Vgs v. Vds',true);
  set_range(0,0,0,0.5);
  plot_graph(0,240,319,239,vdata,adata,num_vds+1,'Vds','Alpha','Alpha v. Vds',true);
  set_range(0,0,0,10e-14);
  plot_graph(320,240,319,239,vdata,qgdata,num_vds+1,'Vds','Charge (Qg-g,Qb-b,Qs-r)','Total charges',true);
  plot_graph(320,240,319,239,vdata,qbdata,num_vds+1,'Vds','Qstotal','Channel charge',false);
  plot_graph(320,240,319,239,vdata,qsdata,num_vds+1,'Vds','Qstotal','Channel charge',false);
  waitforkeypress;
  restorecrtmode;
end;

procedure read_pfile;

{ Read a previously stored parameter file into program }

begin
  readln(pfile,vds_min);
  readln(pfile,vds_max);
  readln(pfile,vds_inc);
  readln(pfile,VFG);
  readln(pfile,Na);
  readln(pfile,Vfb);
  readln(pfile,mu);
  readln(pfile,Tox);
  readln(pfile,Length);
  readln(pfile,dL);
  readln(pfile,width);
```

```
  readln(pfile,dW);
  close(pfile);
end;


procedure read_parameters;

{ Get name of parameter file to be read and call reading procedure }

begin
  write('Enter parameter filename: ');
  readln(npf_name);
  assign(pfile,concat(npf_name,'.par'));
  {$I-}reset(pfile);{$I+}
  if IOResult = 0 then
    begin
      read_pfile;
      pf_name := npf_name;
    end
  else
    begin
      writeln('Cannot find file ',concat(npf_name,'.par'));
      delay(1000);
    end;
end;

procedure save_parameters;

{ Save the current device parameters in a file }

begin
  write('Enter parameter filename: ');
  readln(npf_name);
  assign(pfile,concat(npf_name,'.par'));
  rewrite(pfile);
  writeln(pfile,vds_min);
  writeln(pfile,vds_max);
  writeln(pfile,vds_inc);
  writeln(pfile,VFG);
  writeln(pfile,Na);
  writeln(pfile,Vfb);
  writeln(pfile,mu);
  writeln(pfile,Tox);
  writeln(pfile,Length);
  writeln(pfile,dL);
  writeln(pfile,width);
  writeln(pfile,dW);
  close(pfile);
  pf_name := npf_name;
end;


procedure setup;

{ Provides the user inteface to the program.  A screen containing all
  variable parameters and some calculated contants is displayed.  Device
  variable and various program actions can be initiated using single
  key presses. }
```

```
var legal_key, settings_ok : boolean;
   dummy  : char;
begin
  non_convergence := false;
  textbackground(1);
  textcolor(10);
                                          { Display the program menu }
  clrscr;
  window(10,1,80,25);
  writeln('*************************************************************');
  writeln('*                                                           *');
  writeln('*        Floating Gate MOST model - V16.0                   *');
  writeln('*                                                           *');
  writeln('*************************************************************');
  repeat
   calc_consts;
   window(14,7,80,9);
   clrscr;
   writeln('The settings for the calculation are as follows.');
   settings_ok := true;
   window(4,9,40,25);
   clrscr;
   writeln('   Calculation parameters.');
   writeln('1.  Vds :',Vds_min:5:2,' to ',Vds_max:5:2,' :',
           Vds_inc:5:2);
   writeln('2.  VFG : ',VFG:5:2);
   writeln('3.  Na : ',Na:10);
   writeln('4.  Vfb : ',Vfb:5:2);
   writeln('5.  Mobility : ',mu:6:4);
   writeln('6.  Tox : ',Tox:10);
   writeln('7.  Channel length    : ',Length*1e6:5:2);
   writeln('8.  Channel shortening : ',dL*1e6:5:2);
   writeln('9.  Channel width     : ',Width*1e6:5:2);
   writeln('A.  Channel narrowing  : ',dW*1e6:5:2);
   writeln;
   writeln('Q.  Quit the program.');
   window(41,9,80,25);
   clrscr;
   writeln('    Program constants.');
   writeln('Cox : ',W*L*Cox:10);
   writeln('Cgd (Cgs) : ',Cgd:10);
   writeln('Cdep0 : ',Cdep0:10);
   writeln('Phis : ',Phis:10);
   writeln('Gamma : ',gamma:10);
   writeln('Vt0 : ',Vt0:10:2);
   writeln('Qgate0 : ',Qgate0:10);
   writeln('    File functions');
   writeln('Parameter file name: ',pf_name);
   writeln('R.  Read parameter file');
   writeln('S.  Save parameter file');
   writeln('D.  Save solution data: ', save_data);
   window(20,23,80,25);
   clrscr;
   writeln('Type the number of parameter to be changed');
   write('or <Enter> to run the calculation.');
                                    { Interpret user requests }
   repeat
   legal_key := true;
   dummy := readkey;
```

```
    writeln(dummy);
    if dummy<>chr(13) then
    begin
     case dummy of
     '1': begin write('Minimum Vds : ');readln(Vds_min);
            write('Maximum Vds : ');readln(Vds_max);
            write('Vds increment : ');readln(Vds_inc);
          end;
     '2': begin write('Floating gate voltage: ');readln(VFG);end;
     '3': begin write('Subsrate doping: ');readln(Na);end;
     '4': begin write('Flat band voltage: ');readln(Vfb);end;
     '5': begin write('Electron mobility: ');readln(mu);end;
     '6': begin write('Gate oxide thickness: ');readln(Tox);end;
     '7': begin write('Channel length: ');readln(Length);Length := Length*1e-6;end;
     '8': begin write('Channel shortening: ');readln(dL);dL:=dL*1e-6;end;
     '9': begin write('Channel width: ');readln(Width);Width := Width*1e-6;end;
     'A': begin write('Channel narrowing: ');readln(dW);dW := dW*1e-6;end;
     'R','r': read_parameters;
     'S','s': save_parameters;
     'D','d': save_data := not save_data;
     'Q','q': begin
            closegraph;
            halt;
            end;
     else legal_key := false;
     end; {case}
     settings_ok := false;
    end;
    until legal_key;
    clrscr;
  until settings_ok;
  window(1,1,80,25);
  clrscr;
end;

procedure set_def_values;

{ Set default values for device parameters }

begin
     vds_min := 0.01;                    { Set default device and process parameters}
     vds_max := 5;
     vds_inc := 0.5;
     VFG := 2.0;
     Na := 1.88E22;                      { Device and process }
     Vfb := -0.7;
     mu := 0.0558;
     Tox := 50E-9;
     Length := 3.0E-6;
     dL := 0.65E-6;
     Width := 10.0E-6;
     dW := 0.51E-6;
end;

procedure get_parameters;

{ Read device parameters specifed from the command line }

begin
```

```
    pf_name := '';
    if paramcount < 1 then
      set_def_values
    else
     begin                                  {readin device and process parameters}
      pf_name := concat(paramstr(1),'.par');
      assign(pfile,pf_name);
      {$I-}
      reset(pfile);
      {$I+}
      if IOResult <> 0 then
       begin
        writeln('Cannot find parameter file ',pf_name,' - using defaults.');
        set_def_values;
       end
      else
       read_pfile;
     end;
  end;

                                          { The main body of the program }
  begin
   get_parameters;
   repeat
    setup;                                { Setup the user interface }
    alpha := 0.5;
    Vds := vds_min;
    num_vds := round((Vds_max-Vds_min)/Vds_inc);
    for i := 0 to num_vds do              { Loop for each value of Vds }
     begin
      find_Vgs;                           { Find Vgs for current Vds value }
      calc_Ids;                           { Find Ids for current values of Vds and Vgs
                                            and display the results }
      write('Vds=',Vds:6:2,' Vgs=',Vgs:7:3,' alpha=',alpha:7:3,' Ids=',Ids:8);
      if not linear then writeln(' Vds_sat=',Vds_sat:6:2,' Lp=',Lp*1e6:5:3)
                    else writeln;
      Vds := Vds + vds_inc;
     end;
    if non_convergence then
       writeln('Convergence failure occured during calculation.');
    writeln;
    writeln('Distributions calculated, press key to continue...');
    waitforkeypress;
    plot_graphs;                          { Display the device characteristics for
                                            this run and save the results if necessary }
    if save_data then make_data_files;
   until false;                           { Loop until the user quits }
  end.
```

References

1.1 Crook, D.L., "Evolution of VLSI relibility engineering", Proc. European Symposium on Reliability of Electron Devices, 1991, pp. 293-312

1.2 Woods, M.H., "MOS VLSI reliability and yield trends", Proc. IEEE, Vol. 74, No. 12, Dec 1986, pp. 1715-1729

1.3 Gelsinger, P.P, "Design and test of the 80368", IEEE Design and Test of Comp., June 1987, pp. 42-50

1.4 Moore, W., "A review of fault-tolerant techniques for the enhancement of integrated circuit yield", GEC J. of Res. Vol. 2, 1984, pp.1-15

1.5 Smith, G.L., "Model for delay faults based upon paths", Proc. Int. Test Conf., 1985, pp. 342-349

1.6 Fergusson, F.J., Taylor, M., Larrabee, T., "Testing for parametric faults in static CMOS circuits, Proc. Int. Test Conf., 1990, pp. 436-443

2.1 Fantini, F., Morandi, C., "Failure modes and mechanisms for VLSI IC's - a review.", IEE Proc. Vol. 132, Pt. G, No. 3, June 1985, pp. 74-81

2.2 Burgess, N., Damper, R.I., Totton, K.A., Shaw, S.J., "Physical faults in MOS circuit and their coverage by different fault models.", IEE Proc., Vol. 135, Pt. E, No. 1, Jan 1988, pp. 1-9

2.3 Burgess, N., Damper, R. I., Shaw, S.J., Wilkins, D.R.J., "Fault and fault effects in NMOS circuit - impact on design for testability.", IEE Proc., Vol 132, Pt. G, No. 3, June 1985, pp. 82-89

2.4 Kemp K.G., Poole, K.F., Frost, D.F., "The effect of defects on the early failure of metal interconnects.", IEEE Trans. Rel. Vol. 39, No. 1, Apr. 1990, pp. 26-29

2.5 Al-Arian, S.A., Agrawal, D. P., "Physical failures and fault models of CMOS circuits", IEE Trans. Circ. Sys., Vol. CAS-34, No. 3, Mar. 1987, pp. 269-279

2.6 Galiay, J., Crouzet, Y., Vergniault, M., "Physical vs. logical fault models for MOS LSI circuits: impact on their testability", IEEE Trans. 1980, TC-29 pp. 527-531

2.7 Banerjee, P., Abraham, J., "Characterisation and testing of physical failures in MOS logic circuits", IEEE Design and Test of Computers, Aug. 1984, pp. 76-86

2.8 Courtois, B., "Failure mechanisms, fault hypotheses and analytical testing of MOS LSI circuits", in Gray J.P. (Ed), "VLSI '81", Academic Press, 1981, pp. 341-345

2.9 Courtois, B, "Failure mechanism, fault hypotheses and analytical testing of LSI-NMOS (HMOS) circuit", Proc. VLSI'81, 1981, pp. 341-350

2.10 Abstracting Board of the International Council of Scientific Unions, "International Classification for Physics"

2.11 Woods, M.H., "MOS VLSI reliability and yield trends", Proc. IEEE, Vol. 74, No. 12, Dec 1986, pp. 1715-1729

2.12 Wang, C.T., (Ed.), "Hot carrier design considerations in MOS devices and circuits", Van Nostrand Reinhold, 1992

2.13 Boulaire, J.Y., Bauduin, B., Lelievre, A., "Field components reliability analysis for French telecommunications equipment", Proc. 3rd European Symposium on Reliability of Electron Devices, 1992, pp. 81-91

2.14 Crisenza, G., Ghidini, G., Tosi, M., "Floating gate memory reliability", Proc. 2nd European Symposium on Reliability of Electron Devices, 1991, pp. 505-516

2.15 Hawkins, C.F., Soden, J.M., "Reliability and Electrical Properties of Gate Oxide Shorts in CMOS ICs", Proc. International Test Conference, 1986, pp. 443-451

2.16 Sinnadurai, N., Kuppuswamy, T.S., Chandramouli, R., Rao, B.K.N., "Environmental testing and component reliability observations in telecommunications equipment operated in Indian climatic conditions", Proc. 2nd European Symposium on Reliability of Electron Devices, Oct 1991, pp. 55-63

2.17 Strack, H., Mayer, K.R., Kolbesen, B.O., "The detrimental influence of stacking faults on the refresh time of MOS memories", Solid State Electronics, Vol. 22, pp. 135-140

2.18 Syrzycki, M., "Modelling of gate oxide shorts in MOS transistors", IEEE Trans. on CAD, Vol. 8, No. 3, Mar. 1989, pp. 193-202

2.19 Hawkins, C.F., Soden, J.M., "Reliability and electrical properties of gate oxide shorts in CMOS ICs", Proc. Int. Test Conf. 1986, pp. 443-451

2.20 Shih, D. Y., Ficalora, P.J., "The reduction of Au-Al intermetallic formation and electromigration in hydrogen environments.", IEEE Trans. Elec. Dev. , Vol. ED-26. 1979, pp. 27-34

2.21 Black, J.R., "Electromigration - A brief survey and some recent results", IEEE Trans. Elec. Dev., Vol. 16 , 1969, pp. 338-347

2.22 Richards, B.P., Footner, P.K., "Failure analysis of semiconductor devices - Rationale, methodology and practise", GEC Journal of Research, Vol 1., 1983, pp. 74-91

2.23 Henderson, C.L., Soden, J.M., Hawkins, C.F., "The behaviour and testing implications of CMOS IC logic gate open circuits", Proc. Int. Test Conf. 1991, pp. 302-310

2.24 Anolick, E.S., Nelso, G.R., "Low-field time-dependent dielectric integrity", IEEE Trans. on Rel. Vol. R-29, 1980, pp. 217-221

2.25 Chen I, Holland, S.E., Hu, C., "Electrical breakdown in thin gate and tunnelling oxides", IEEE Trans. on Elec. Dev. Vol. ED-32, 1985 pp. 413-422

2.26 Yamabe, K., Taniguchi, K., "Time-dependent dielectric breakdown of thin thermally grown SiO_2 films", IEEE Trans. on Elec. Dev. Vol. ED-32, 1985 pp. 423-428

2.27 Hess, K., Shah, C.T., "Hot carriers in silicon inversion layers.", J. Appl. Phys.. Vol. 45, 1974, pp. 1254-1257

2.28 Ning, T., Osburn, C.M., Yu, H.N., "Emission probability of hot electrons from silicon into silicon dioxide.", J. Appl. Phys., Vol 48, 1977, pp. 286-293

2.29 Hu, C., "Lucky-electron model for channel hot electron emission", Int. Elec. Dev. Meeting Tech. Dig., 1979, p. 22

2.30 Eitan, B., Frohman-Bentchkowsky, D., Shappir, J., "Impact ionisation at very low voltages in silicon.", J. Appl. Phys., Vol 53, 1982, pp. 1244-1247

2.31 Tam, S., Hsu, F.-C., Hu, C., Muller, R.S., Ko, P.K., "Hot electron currents in very short channel MOSFETs.", IEEE Trans. Elec. Dev. Lett., Vol. EDL-4, 1983, pp. 249-251

2.32 Hsu, F.C., Tam, S., "Relationship between MOSFET degradation and hot electron induced interface state generation.", IEEE Trans. Elec. Dev. Lett., Vol. EDL-5, 1984, pp. 50-52

2.33 Takeda, E., Kume, H., Toyabe, T., Asai, S., "Submicrometer MOSFET structure for minimising hot carrier generation.", IEEE Trans. Elec. Dev. , Vol. ED-29, 1982, pp. 611-618

2.34 Katto, H., Okuyama, K., Meeguro, R., Nagai, R., Ikeda, S., "Hot carrier degradation modes and optimisation of LDD MOSFET's.", Int. Elec. Dev. Meeting Tech. Dig., 1984, p. 774

2.35 Aur, S., Duvvury, C., Mcadams, H., Perrin, C., "Identification of DRAM Sense-amplifier imbalance using hot carrier evaluation.", IEEE J. Solid State Circuits, Vol. 27, 1992, p. 451

2.36 Pimbley, J.M., "A measurement method for the increase of switching time due to hot electron stress.", IEEE Trans. Elec. Dev. Lett., Vol. EDL-6, 1985, pp. 366-368

2.37 Wadsack R.L. "Fault modelling and logic simulation of CMOS and MOS integrated circuits", Bell Sys. Tech. J., Vol 57, 1978, pp. 1449-1474

2.38 Abraham, J.A., Fuchs, W.K., "Fault and error models for VLSI", Proc. IEEE, Vol. 74, 1986, pp. 639-654

2.39 Hayes, J.P., "Fault modeling", IEEE Design and Test of Comp., Apr. 1985, pp. 88-95

2.40 Shen J.P., Maly, W., Ferguson F.J., "Inductive fault analysis of MOS integrated circuits", IEEE Design and Test of Comp. Dec. 1985, pp. 13-26

2.41 Maly, W., "Modeling of lithography related yield losses for CAD of VLSI circuit", IEEE Trans. on Comp.-Aided Design, Vol. CAD-4, 1985, pp. 166-177

2.42 Stapper, C.H., "Modeling of integrated circuit defect sensitivities.", IBM J. Res. Dev., Vol. 27, 1983, pp. 549-557

2.43 Stapper, C.H., Armstrong, F.M., Saji, K., "Integrated circuit yield statistics", Proc. IEEE, Vol. 71, No.4, 1983, pp. 453-470

2.44 Fergusson, F.J., Taylor, M., Larrabee, T., "Testing for parametric faults in static CMOS circuits", Proc. IEEE Int. Test Conf. 1990, pp. 436-443

2.45 Burgess, N., Damper, R.I., Totton, K.A., Shaw, S.J., "Physical faults in MOS circuits and their coverage by different fault models", IEE Proc., Vol 135, Pt. E, 1988, pp. 1-9

2.46 Bate J.A., Miller D.M., "Exhaustive testing of stuck-open faults in CMOS combinational circuits", IEE Proc. Vol. 135, Pt E, 1988, pp. 10-16

2.47 Bensouiah, A., Johnson, S., Morant, M.J., "On the elimination of data compression in response analysis for BIST", Proc. 5th Int. Conf. on Quality in Electronic Components, 1991, pp. 595-597

2.48 Reddy, M.K., Reddy, S.M., "Detecting FET stuck-open faults in CMOS latches and flip-flops", IEEE Design and Test of Computers, Vol. 3, No. 5, 1986, pp. 17-26

2.49 Liu, D.L., McCluskey, E.J., "Design of CMOS VLSI circuits for testability", J. of Semicustom ICs, 1987, pp. 5-10

2.50 Soden J.M, Treece, R.K., Taylor, M.R., Hawkins, C.F., "CMOS IC stuck-open fault electrical effects and design considerations", Proc. Int. Test Conf., 1989, pp. 423-430

2.51 Moritz, P.S., Thorsen, L.M., "CMOS circuit testability", IEEE J. Solid State Circuits, Vol. SC-21, 1986, pp. 306-309

2.52 Banerjee, P., Abraham, J.A., "A multivalued algebra for modelling physical failures in MOS VLSI circuits", IEEE Trans. Comp. Aided Design, Vol. CAD-4, 1985, pp. 312-321

2.53 Liu, D.L., McCluskey, E.J., "Designing CMOS circuits for switch level testability", IEEE Design and Test of Computers, 1987, pp. 42-49

2.54 Nelson, G.F., Boggs, W.F., "Parametric tests meet the challenge of high density ICs", Electronics, 11 Dec.1975, pp. 108-111

2.55 Levi, M.W., "CMOS is most testable", Proc. Int. Test Conf., 1981, pp. 217-220

2.56 Hawkins, C.F., Soden, J.M., "I_{DD_Q} testing for CMOS IC open circuits", Private communication.

2.57 Rodriguez-Montanes, R., Segura, J.A., Champac, J.F., Rubio, J.A., "Current vs. logic testing of gate-oxide shorts, floating gate and bridging failures in CMOS circuits", Proc. Int. Test Conf., 1991

2.58 Richardson, A.M., Dorey, A.P., "Reliability indicators", 3rd European Symp. on Rel. Elect. Devices, 1992, pp. 277-283

2.59 Waicukauski, J.A., Lindbloom, E., Rosen, B.K., Iyengar, V.J., "Transition fault simulation", IEEE Design and Test of Computers, 1987, pp. 32-38

2.60 Smith, G.L., "Model for delay faults based upon paths", Proc. Int. Test Conf., 1985, pp. 342-349

2.61 Renovell, M., Cambon, G., "Topology dependence of floating gate faults in MOS integrated circuits", Elec. Lett. Vol. 22, 1986, pp. 152-153

2.62 Maly, W., Nag, P.K., Nigh, P., "Testing oriented analysis of CMOS ICs with opens", Proc. Int. Conf. Comp. Aided Des., 1988, pp. 344-347

3.1 Bauer, C.L., "Stress and current induced degradation in thin-film conductors", Proc. 2nd European Symposium on Reliability of Electron Devices, Oct. 1991, pp. 161-170

3.2 Black, J.R., "Electromigration - a brief survey and some recent results", IEEE Trans. Elec. Dev., Vol. ED-16, 1969, pp. 338-347

3.3 Ferguson, F.J., Shen, J.P., "Extraction and simulation of realistic CMOS Faults using inductive fault analysis", Proc. Int. test Conf., 1988, pp. 475-484

3.4 "International Workshop on Designing for Yield", Oxford, UK, 1987

3.5 Stapper, C.H., "Modelling of integrated circuit defect sensitivities.", IBM J. Res. Dev., Vol. 27, 1983, pp. 549-557

3.6 Gandemer, S., Tremintin, B., Charlot, J.J., "A method for determining critical areas and critical levels for IC yield estimation", Proc. Int. Workshop on Designing for Yield, Oxford, 1987

3.7 Polcarin, N.R., Lloyd, J.R., Cvikevich, S., "Electromigration failure in heavily doped polycrystalline silicon", Proc. Int. Rel. Phys. Symp, 1980, pp. 178-185

4.1 Everhart, T.E., Wells, O.C., Matta, R.K., "Evaluation of passivated integrated circuits usig the scanning electron microscope", J. Electrochemical Soc., Vol. 111, 1964, pp. 929-936

4.2 Hachenberg, O., Brauer, W., in Advances in Electronics and Electron Physics, Vol. XI, L. Marton, Ed., Academic Press, 1959, pp. 413-499

4.3 Wolfgang, E., "Electron beam testing: problems in practise", Scanning, Vol. 5, 1983, pp. 71-83

4.4 Chung, M.S., Everhart, T.E., "Simple calculation of energy distribution of low energy secondary electrons emitted from metals under electron bombardment", J. Appl. Phys., Vol. 45, No. 2, 1974, pp. 707-709

4.5 Dinnis, A., "Static and dynamic voltage measurements using an electron probe", J. Phys. E:Sci. Instrum., Vol. 21, 1988, pp. 522-533

4.6 Nakamae, K., Fujioka, H. Ura, K., "Local filed effects on voltage contrast in the scanning electron microscope", J. Phys. D: Appl. Phys., Vol. 14, 1981, pp. 1939-1960

4.7 Wells, O.C., Bremer, C.G., "Voltage measurement in the scanning electron microscope", J. Sci. Instrum., (J. Phys. E), Vol. 1, 1968, pp. 902-906

4.8 Feuerbaum, H.P., "VLSI testing using the electron beam probe", Scanning Electron Microscopy, Vol. 1, 1979, pp. 285-296

4.9 Gopinath, A., "Estimate of the minimum measurable voltage in the SEM", J. Phys. E: Sci. Instrum., Vol.10, 1977, pp.911-913

4.10 Pensak, L., "Conductivity induced by electron bombardment in thin insulating films", Phys. Rev., Vol. 75, 1949, pp. 472-478

4.11 Richardson, N., "E-beam probing for VLSI circuit debug", VLSI Systems Design, 1987, pp. 24-29/58

4.12 Brettell, D.J. "Image Processing", Internal Report, University of Durham, School of Engineering and Computer Science, 1987

4.13 Ranasingh, D.W., "State of the art electron beam testing of IC's", at Electron Beam Testing of Integrated Circuits, The Institute of Physics, Jan. 1989

6.1 Kahng, D., and Sze, S.M., "A floating gate and its application to memory devices", Bell Systems Tech. J., Vol. 46, 1967, pp. 1288-1295

6.2 Frohman-Bentchkowsky, D., "Memory behaviour in a floating-gate avalanche-injection MOS (FAMOS) structure), Appl. Phys. lett., Vol. 18, 1971, pp. 332-334

6.3 Frohman-Bentchkowsky, D., "FAMOS - A new semiconductor charge storage device", Solid State Electronics, Vol. 17, 1974, pp. 517-529

6.4 Card, H.C., Elmasary, M.I., "Functional modelling of non-volatile memory devices", Solid State Electronics, Vol. 19, 1976, pp. 863-870

6.5 Card, H.C., Heasell, E.L., "Modelling of channel enhancement effects on the write characteristics of FAMOS devices", Solid State Electronics, Vol. 19, 1976, pp. 965-968

6.6 Renovell, M., Cambon, G., "Topology dependence of floating gate faults in MOS integrated circuits", Elec. Lett. Vol. 22, 1986, pp. 152-153

6.7 Renovell, M., Faure, P., Cambon, G., "Mask dependent models of faults in MOS integrated circuits", Proc. Int. Circ. Technology Conf., 1986, pp. 125-134

6.8 Grove, A.S., "Physics and Technology of Semiconductor Devices", 1967, Wiley

6.9 Sze, S.M., "Physics of Semiconductor devices", 2nd Ed.,1981, Wiley

6.10 Brews, J.R.,"A charge sheet model of the MOSFET", Solid State Electronics, Vol. 21, 1978, pp. 345-355

6.11 Frohman-Bentchkowsky, D., Grove, A.S., "Conduction of MOS transistors in saturation", IEEE Trans. Elec. Devices, Vol. ED-16, 1969, pp. 108-113

6.12 Reddi, V.G.K, Sah, C.T., "Source to drain resistance beyond pinch-off in MOS transistors", IEEE Trans. Elec. Dev., Vol. ED-12, 1965, pp. 139-141

7.1 Shichman, H., Hodges, D.A., "Modelling and simulation of insulated gate field effect transistor switching circuits", IEEE J. Solid-State Circuits, Vol. SC-3, No. 3, 1968, pp. 285-289

7.2 Profumo, E., "The MOS transistor", in Antognetti, P., Massobrio, G. (Ed.), "Semiconductor Device Modelling with SPICE", 1988, McGraw-Hill Book Company

7.3 Meyer, J.E., "MOS models and circuit simulation", RCA Review, Vol. 32, 1971

7.4 Ward, D.E., Dutton, R.W., "A charge oriented model for MOS transistor capacitances", IEEE J. Solid State Circ., Vol. SC-13, 1978, pp. 703-707

7.5 "SOLO 1400 Reference Manual", European Silicon Structures, 1991

7.6 Soden, J.M., Hawkins, C.F., "Power supply current monitoring as a CMOS reliability indicator", European Symposium on Reliability of Electron Devices Tutorial Notes, 1992